

## INTRODUCTION TO THE SERIES

The aim of the *Handbooks in Economics* series is to produce Handbooks for various branches of economics, each of which is a definitive source, reference, and teaching supplement for use by professional researchers and advanced graduate students. Each Handbook provides self-contained surveys of the current state of a branch of economics in the form of chapters prepared by leading specialists on various aspects of this branch of economics. These surveys summarize not only received results but also newer developments, from recent journal articles and discussion papers. Some original material is also included, but the main goal is to provide comprehensive and accessible surveys. The Handbooks are intended to provide not only useful reference volumes for professional collections but also possible supplementary readings for advanced courses for graduate students in economics.

KENNETH J. ARROW and MICHAEL D. INTRILIGATOR

## INTRODUCTION AND OVERVIEW

DAVID THROSBY\*

*Macquarie University, Sydney, Australia*

### Contents

1. Introduction	4
1.1. Development and scope of the field	4
1.2. Basic questions	6
2. Behavior of individuals	8
3. Behavior of firms and industries	11
4. Behavior of markets	13
4.1. Historical aspects	13
4.2. Present-day art markets and prices	14
4.3. Legal regulation affecting markets	15
5. Arts and culture in the economy	16
5.1. Culture in economic performance	17
5.2. Cultural policy	17
5.3. Some specific policy areas	19
6. Conclusions	21
References	21

\* I am grateful to my co-editor, Victor Ginsburgh, for allowing me to indulge my desire to write an introductory chapter to this volume, and I express my gratitude to him for his contributions to and comments on an earlier draft, though I alone take responsibility for views expressed.

*Handbook of the Economics of Art and Culture, Volume 1*

*Edited by Victor A. Ginsburgh and David Throsby*

*Copyright © 2006 Elsevier B.V. All rights reserved*

*DOI: 10.1016/S1574-0676(06)01001-5*

## 1. Introduction

### 1.1. Development and scope of the field

Over the last 30 or 40 years a substantial literature has grown up in which the tools of economic theory and analysis have been applied to problems in the arts and culture. Economists who have surveyed the field generally locate the origins of contemporary cultural economics (as the discipline is loosely known) as being in 1966, the year of publication of the first major work in modern times dedicated specifically to the economics of the arts [Baumol and Bowen (1966)]. Not only did the Baumol and Bowen book demonstrate that straightforward economic analysis could illuminate the supply of and demand for artistic services and the role of the arts sector in the economy, the work is also notable in retrospect for having put forward one of the most enduring theoretical propositions in cultural economics, namely the productivity lag or “cost disease” phenomenon which afflicts the live performing arts.

Nevertheless, although 1966 is generally agreed to be the birth year of the economics of the arts as a recognizable field of study, this is not to say that there was no pre-history. In fact, as Craufurd Goodwin (Chapter 2) shows, a number of economists of the past have been interested in the arts in one way or another, going back to well before Adam Smith. Goodwin notes that Hume, Turgot and later Smith speculated in some detail on the place of the arts in society, although their lead was not followed by the major political economists of the nineteenth century who (apart from Ruskin) saw little of particular interest in this area, apparently following Bentham with his famous remark about pushpin and poetry. Similarly, except for Veblen’s writings, not much was said about the arts by the American Institutionalists of the late nineteenth and early twentieth centuries, although across the Atlantic in England the Bloomsbury group, of which Keynes was a leading light, wrote extensively about the importance of the arts to individual and social well-being, and proposed ways in which governments could help through enlightened policy.

This brief summary of the pre-history brings us up to the point of departure of the modern discipline as noted above. Following the appearance of Baumol and Bowen (1966), interest in the economics of the arts initially expanded slowly. The first collection of papers in the field appeared in the mid-1970s [Blaug (1976)]; in his introductory essay to this volume, Blaug noted that there was as yet insufficient substance in this area of economic inquiry to justify a textbook, but that an assembly of articles would be useful. The works gathered together were focused mainly on public policy issues, in particular the rationale for public subsidy and the evaluation of public expenditure, issues that have remained prominent on the agenda of the economics of the arts ever since.

In 1977 the *Journal of Cultural Economics* was founded by William Hendon and his colleagues at the University of Akron and in 1979 the first of what was to become a regular series of biennial international conferences was held (in Edinburgh). By the early 1990s, a sufficiently large body of work had accumulated to warrant an overview and

assessment of the field in the *Journal of Economic Literature* [Throsby (1994)]. This survey reflected on the expanding variety of issues in the arts that had engaged the attention of economists up to that time. It began by pointing to the origins of demand for the arts as lying in the peculiar processes of taste formation and cumulative consumption that make the arts an “experience good”. Markets for art objects and for artistic labor were also covered, as well as issues in the performing arts and in public policy. Throsby’s overall evaluation of the state of cultural economics at that time foresaw a further expansion of the domain of interest in the field in the future.

In fact such an expansion was mirrored to some extent in the choice of articles reprinted in the major collection assembled in Towse (1997). In her Introduction to this two-volume collection, Towse pointed out that “culture” is a term with a wider embrace than “the arts” and hence that cultural economics, for the purpose of her project, could be taken to include the media and heritage as well as the arts. Again, however, the central themes of defining cultural goods, the cost-disease phenomenon, and questions relating to public subsidy of the arts were brought forward as the major defining issues that had characterized the development of the discipline over the period covered by her anthology.

The classification of topics in cultural economics that Towse used for her 1997 collection was taken up four years later by Blaug in his survey that asked “Where are we now in cultural economics?” [Blaug (2001)]. The list of topics that he discussed included demand, supply, industrial organization, the art market, history, labor markets, firm behavior (including the cost disease) and public subsidies. It seemed that by now, at the turn of the century and after three or four decades of research and writing by an ever-increasing number of economists, the list had become more or less settled, though Blaug noted in the conclusion to his survey that there was much further expansion of the field in store.

In another overview published in the same year, Ginsburgh (2001) observed that cultural economics is not well-defined because it is located at the crossroads of several disciplines: art history, art philosophy, sociology, law, management and economics. Nevertheless it was from economics that the literature he surveyed was drawn, covering again the principal topics noted above. Ginsburgh gave prominence to the policy questions surrounding public support for the arts, with a comprehensive listing of the various normative rationales that have been put forward, and a discussion of the means for intervention once a rationale has been accepted. He also saw markets for visual arts as providing particularly fertile ground for economists to work in, especially since this is an area where good data are available that open up possibilities for econometric investigation. Similar ground is traversed in the entry on the economics of art in the second edition of *The New Palgrave Dictionary of Economics* [Throsby (2007)]. This article points out that amongst the multitude of applications of economic analysis to the arts, the field of industrial organization remains under-represented, offering fruitful opportunities for future researchers.

The specific schools of economic thought and analysis that have been applied to problems in the arts and culture are usefully summarized by Towse (2003) in the final survey



of the field that we consider here. The basic analytical methods of neoclassical price theory and welfare economics comprise the main approaches that microeconomists have adopted in looking at issues of production, consumption and exchange in the arts and cultural industries; at the macroeconomic level relationships between the cultural and other sectors of the economy are implicated, and the familiar tools of macroeconomic policy analysis have been applied to policy-making affecting the cultural sector. Similarly public choice theory and the field of contemporary political economy have a bearing on policy formation relating to the arts, while institutional economics offers an alternative means for analyzing organizational structures and their influence on behavior. Towse also notes the relevance of the law-and-economics tradition deriving from Coase and others as the basis for analysis of transactions costs and property rights questions (especially copyright) relating to artistic goods and services.

What does the above survey of the surveys tell us about the scope and achievements of the economics of the arts and culture? At one level it suggests that a sufficiently cohesive body of work has accumulated for us now to be able to identify a field labeled “cultural economics” or “the economics of the arts and culture” with some confidence, even though the outer boundaries of the field remain somewhat fuzzy. It also suggests that the time is right for assembling a purposeful collection of essays commissioned from researchers working at the theoretical and empirical frontiers of the field, such as is contained in the present volume. At a deeper level our survey reveals that a basic issue that gives the field its particular appeal is that of defining whether and how cultural goods and services differ from other goods and services in the economy. Whilst it is the mysterious and possibly unfathomable nature of art in all its forms that ultimately must underlie much of the behavior of individuals, firms and markets in the arts, economics requires a more systematic basis on which its analyses can rest. We turn to these fundamental definitional questions in the next section.

### *1.2. Basic questions*

Several chapters in this volume begin with a discussion of what the words “art” and “culture” mean. Of course these are foundational questions in disciplines such as art history, art theory, philosophy, aesthetics, sociology and anthropology. By and large contributors to this volume accept in the first instance a broad-ranging view of culture as comprising or being defined by a set of attitudes, beliefs or values common to a group that somehow identifies and binds the group together. Thus it is possible to speak of a national culture, a religious culture, a corporate culture and so on. From the noun “culture” comes the adjective “cultural”, meaning reflective of the particular shared values. Culture is also defined in this book in a more functional sense to indicate the practices and products of cultural activity, including especially the arts. Within this meaning the terms “high culture” and “low culture” are frequently used to refer to the “serious” and “popular” arts, respectively. The arts as a whole are sometimes divided into the performing arts (acting, dancing, singing, playing a musical instrument, etc.) and the initial creative arts (visual art, sculpture, craft, creative writing, musical composition, etc.) though there

are obvious overlaps between these categories. According to these definitions the arts comprise a subset of culture more broadly defined.

This leads us in turn to considering the definition of artistic and cultural goods. The distinguishing characteristics of such goods (and services) that have been identified in the literature of cultural economics include the following, though by no means every characteristic will necessarily be present in every such good:

- cultural goods are experience goods, the taste for which grows as they are consumed in greater quantities; they are therefore subject to rational addiction;
- cultural goods have some public-good properties; in aggregate they yield positive externalities or diffused benefits that may be demanded in their own right;
- cultural goods result from production processes in which human creativity is an important input;
- cultural goods are the vehicles for symbolic messages to those who consume them, i.e. they are more than simply utilitarian but serve in addition some larger communicative purpose;
- cultural goods contain, at least potentially, some intellectual property that is attributable to the individual or group producing the good; and
- cultural goods embody or give rise to forms of value that are not fully expressible in monetary terms and that may not be revealed in either real or contingent markets.

This set of characteristics of cultural goods is illuminating in capturing most of the distinguishing features that such goods display in varying degrees, but it is problematical insofar as it does not indicate whether any one or any combination of the characteristics could be regarded as a *sufficient* condition for defining a cultural good, or, if this is not possible, which or how many of the characteristics would be regarded as *necessary* conditions. Nevertheless the list does at least represent a checklist if not a basis for a watertight definition.

A further problem lies in the question of whether, within such a set of characteristics, there is a difference between an artistic and a cultural good. Using an aesthetic definition of art and a broadly anthropological definition of culture, Roger McCain ([Chapter 5](#)) goes so far as to argue that artistic and cultural goods are non-overlapping sets; he sees the distinguishing characteristic of the former as lying in their aesthetic properties and in the creativity that has gone into their making, whereas objects reflecting a group's culture do not necessarily possess these qualities. McCain's approach stands in contrast to that indicated above. However if, as argued earlier, art is a subset of culture, the adjective "artistic" would presumably become a subset of the adjective "cultural", and hence artistic goods would be seen as a subset of the wider category of cultural goods, not in a separate category of their own.

Questions of aesthetics relate not just to definitional issues surrounding artistic goods, but go to deeper matters to do with the value of art. There has been little interaction between the disciplines of economics and philosophy in this area, yet philosophers have had much to say that is relevant to the ways in which economists might approach questions of value and valuation in the arts. Michael Hutter and Richard Shusterman ([Chapter 6](#)) trace the evolution of ideas about aesthetic value in philosophy from Plato

and Aristotle onwards, comparing and contrasting these ideas with those of economists from the seventeenth century to the present day. In doing so they raise the intriguing issue of the relationship between cultural and economic value, an area of emerging cross-disciplinary interest at the present time.<sup>1</sup>

In the remainder of this introductory chapter I provide a brief overview of the book. Instead of proceeding chapter by chapter, I adopt here a more wide-ranging classification system than that used in the volume itself, in order to give a broader integrated overview of the field. In Sections 2, 3 and 4 of the present chapter I consider the behavior of individuals, firms/industries, and markets, respectively. Section 5 draws together chapters that treat the role of the arts and culture in the economy, and Section 6 presents some conclusions.

## 2. Behavior of individuals

The origins of artistic consumption, production and exchange lie in the behavior of individuals, whether they are consumers who demand cultural goods and services in the marketplace, or producers who supply them. If we were to adopt a broad definition of cultural goods that embraced popular cultural forms such as sport, television programs, magazines, etc., we would open up an enormous field of interest in mass cultural consumption and production, a field that has been extensively explored within sociology, contemporary cultural studies, and media economics. Most of this work lies beyond our immediate scope; for present purposes we focus specifically on individual behavior in regard to the core creative arts.

Looking first at demand, we have noted already the processes of taste formation that are peculiar to cultural goods and that drive the evolution of individual preferences. In common with demand analysis elsewhere in economics, much of the empirical work on demand for the arts has taken tastes as given, and has concentrated on estimating the relative influences of the sorts of explanatory variables normally included in demand functions – own price, price of substitutes, income, sociodemographic characteristics of consumers and so on. Apart from studies of demand for art works (discussed further below), the majority of research on demand for the arts has focused on the performing arts – opera, music, theater and dance. As Bruce Seaman ([Chapter 14](#)) shows in his comprehensive appraisal of this field, results have varied surprisingly widely, and even apparently self-evident propositions such as that the arts are a luxury good are by no means universally confirmed. Nevertheless the empirical evidence suggests, amongst other things, that education is a stronger determinant of demand than income (a result consistent with the art-as-an-acquired-taste hypothesis) and that output quality plays a significant role in determining consumption behavior. Seaman goes on to observe that non-standard socioeconomic factors such as “life-style” characteristics may exert considerable influence on demand, and deserve to be better understood.

<sup>1</sup> For a collection of papers by economists and others on this topic, see [Hutter and Throsby \(2006\)](#).

Turning to the behavior of individuals in the production of art brings us of course to a consideration of the economic role of individual artists as producers of art works or as suppliers of artistic goods and services. An economic analysis of these roles can make the standard distinction between production and supply functions. But although this distinction can be drawn straightforwardly in conventional terms, application of traditional production models or labor supply functions to artists needs to account for the particular conditions under which artistic effort is deployed. Let us look first at production. Efforts to represent the relationship between the output of art and the inputs of factors of production such as labor and capital must immediately confront a fundamental question. As any artist will attest, the writing of a poem, the painting of a picture, or the composition of a flute sonata is not a mechanical process where criteria of technical efficiency and predictable input–output relationships can be applied. Rather, achievement in the arts is conditioned by ephemeral and possibly unquantifiable influences like luck, inspiration and imagination. In short, artistic output is critically dependent on the input of creativity, a characteristic of artists (as well as of some other occupations such as scientific researchers or inventors) that is as important as it is elusive. Nevertheless, as Tony Bryant and David Throsby (Chapter 16) show, some progress can be made in modeling creativity in economic terms. They propose a model in which the artist decides on an optimum level of creativity to apply in the production of “pure” and “commercial” art; the artist’s choice is determined in part by the levels of economic and cultural reward yielded by the output produced. Although the measurement of creativity is fraught with difficulty, Bryant and Throsby present some empirical results suggesting that modeling the creative input in artistic production processes is by no means impossible.

We turn now to the matter of labor supply. Markets for artistic labor, including the labor both of performing artists such as actors and musicians and of “initial creative artists” such as writers and painters, have been widely studied in theoretical and empirical terms in an effort to compare and contrast artists with other occupations in their labor market behavior. The picture is complicated by three features of artistic labor that combine to set artists somewhat apart from other workers. These features are

- financial rewards to professional artistic practice are generally lower than in other occupations with otherwise similar characteristics (education and training requirements, etc.); thus artists’ labor market profiles typically exhibit multiple job-holding;
- the level of variability of artistic earnings is generally higher than in comparable occupations, making an individual artist’s attitudes to risk an important determinant of his or her labor market participation;
- non-pecuniary motives are important in the allocation of an artist’s time between alternative labor markets, i.e. the “inner drive” to create art may dominate financial incentives or at least mediate their influence.

A considerable volume of empirical evidence has accumulated on all three of these aspects of artistic labor markets, as Neil Alper and Gregory Wassall (Chapter 23) show in their extensive overview of the field. They bring these three aspects together in the context of labor-market entry and exit decisions of artists, i.e. in the analysis of career

patterns in the arts. This is an area where, as Alper and Wassall demonstrate, availability of data is critical, whether obtained from official statistical sources such as censuses or from purpose-built surveys.

It is well known that in many countries at the present time labor markets across the board are changing, with greater casualization and increased occupational mobility in the workforce. Labor markets in the arts have been particularly susceptible to change, as Pierre-Michel Menger ([Chapter 22](#)) argues; long-term employment has been replaced by a project-based system of production relying on short-term hiring, large parts of business risk are transferred downwards onto the workforce, and artists learn to manage risk and to stay alive through multiple job-holding, occupational versatility, diversification of job portfolios and occasional income transfers from social security or other sources. Despite manifold deterrents to an artistic career, an excess supply of artists persists in many countries, attributable in part to the non-pecuniary attraction of work as an artist as mentioned above. The excess supply may also be influenced by the entry of large numbers of artists hoping to become a superstar, i.e. someone who breaks away from the pack, becomes a celebrity and enjoys earnings considerably in excess of those that marginal productivity theory would predict. A number of contributors to this volume make reference to this phenomenon, all of them noting the origins of superstar theory in [Sherwin Rosen's \(1981\)](#) classic paper. [Moshe Adler \(Chapter 25\)](#) provides an appraisal of this field, drawing attention to questions of efficiency in terms of the effects of superstardom on consumers' prices and on artists' psychic income. Empirical testing of various superstar hypotheses is relatively sparse and, as Adler suggests, offers prospects for some interesting further research, provided suitable data can be found.

A final question in our discussion of artists and their labor markets brings us back to first principles and provides a link between all of the chapters discussed above. The question is: to what extent does basic human capital theory help to explain all of the phenomena we have been discussing – the role of creativity and talent, the levels and variability of artists' earnings, multiple job-holding, career choice, and the superstar phenomenon? [Ruth Towse \(Chapter 24\)](#) examines this issue, arguing that human capital theory provides at best only a weak basis for explaining artists' decisions relating to investment in training and their choice of an artistic career. Notwithstanding the possible effects of education, training and on-the-job experience, it is ultimately the artist's innate talent that is likely to be the main determinant of artistic success. Moreover, some of the traditional features of the human capital model may be being undermined by technological developments; for instance, Towse argues that the reproducibility of artworks, together with the expansion of the reach of copyright law, overturns the view that human capital cannot be separated from labor. While this latter point may not be of great significance in itself, it does indicate a need for further research on the relationship between copyright and artists' labor supply decisions.

### 3. Behavior of firms and industries

Relatively few cultural goods and services pass from an individual supplier such as an artist directly to the final consumer; most are produced and/or marketed by firms. Across many fields in the arts and culture the predominant firm types in numerical terms are small or medium-sized enterprises constituted on either a for-profit or not-for-profit basis. Microeconomic theory offers straightforward means for representing the production and cost conditions of such firms. Much attention in the economics of the arts has been devoted to not-for-profit firms because of their prevalence in the “serious” arts where commercial for-profit production may not be economically viable. In examining the economic structure and operation of non-profit firms in the performing arts, Arthur Brooks ([Chapter 15](#)) looks particularly at product quality issues, cross-subsidization possibilities and Baumol’s “cost disease” as influences on the firms’ decision-making. The achievement of artistic excellence and the pursuit of innovation are often significant goals of these firms, objectives likely to be shared by potential donors both public and private; these considerations will influence the financial strategies pursued by non-profit firms in the arts, including their ticket-pricing behavior and their fundraising activities.

Profit-making firms in the arts and culture, on the other hand, present a different picture. While corporate enterprises producing movies, books, Broadway musicals, rock concerts, etc. share many of the economic features of commercial firms in other sectors, they are still distinctive on account of the artistic or cultural nature of the product they produce. Although profit-maximization may remain a dominant goal of such firms, it is likely to be mediated, sometimes quite profoundly, by the desire of the firm’s owners and managers to meet certain artistic or cultural standards or to strive for artistic success. While such behavior might be able to be rationalized simply in terms of the firm’s profit-seeking objectives if “better” standards are likely to be more profitable, there remains a sense in which some corporate satisfaction may derive, at least to some extent, from purely artistic or cultural achievement.

Be that as it may, an increasing amount of attention has been paid in recent years to the structure and operation of the so-called cultural or creative industries. The boundaries around these industries remain somewhat fuzzy – for example, do they include fashion, design, software, advertising? Nevertheless if the broad definition of cultural goods discussed earlier is accepted, the cultural industries can be defined simply as those industries producing such goods. In practice the cultural industries have been taken to include the arts, heritage, film, television and radio, the press, publishing, video and computer games, some aspects of tourism, etc. The organization of these industries depends on many types of contracts, the basis for Richard Caves’ ([Chapter 17](#)) analysis of their structure, conduct and performance. Caves distinguishes between simple cultural activities in which a single artist deals with one agent or firm (the writer of a book or movie script) and more complex situations in which several creative inputs are needed and combined by the firm (the production of a movie). The peculiar characteristics of the contracts involved derive from what he describes as the “bedrock properties” of creative work and creative products.

Four chapters in the present volume examine the features of a particular cultural industry in detail, dealing in turn with the media, the movies, popular music, and book publishing. All four industries are characterized by the fact that they produce under high fixed costs but negligible marginal costs, and under conditions of considerable uncertainty: the rate of success is small, but successes command large revenues. First, Simon Anderson and Jean Gabszewicz (Chapter 18) analyze the media industries as perfect examples of two-sided markets: in the case of television, for example, on the one side viewers watch (and pay for) programs, on the other side firms use the market to advertise their products. The two sides are coordinated by one or several platforms; in the television case broadcasters have to balance income from advertising by choosing the right mix of programs (which consumers want) and advertising (to which consumers are generally averse). The same idea applies to newspapers, magazines, the Internet and radio broadcasting. The authors' theoretical analysis is backed by data that show that the media industry is large not only in terms of its importance in GNP, but also because of its influence in shaping public opinion; in 2003, for example, the average American family spent almost eight hours per day watching television.

Second, another important cultural industry to have grown up in the twentieth century is movies. A particular economic characteristic of the movie industry is the fact that it is subject to "wild uncertainty", a phenomenon discussed in detail by Arthur De Vany (Chapter 19). He shows that the statistical laws that motion pictures obey are not normal, but stable Paretian. Each movie is unique in itself, very few are successful, although when they are they can be very profitable. Nothing is predictable, in other words "nobody knows", a property that Caves suggests applies to all cultural industries.

Music, of course, is a much older industry, but during the twentieth century its market has changed dramatically, going from the stage of small manufacturers acting in small venues to a large and powerful industry, in which "classical" music is declining and popular music is on the rise; in 1994, for example, rock and roll and pop music led with 45 percent of unit sales of CDs, while classical music reached less than 4 percent (down from 25 percent in 1950). Popular music is discussed by Marie Connolly and Alan Krueger (Chapter 20), who cover mainly the economics of live performances. Live appearances on stage provide the bread and butter for many performers, giving rise to the Bowie Theory; David Bowie is quoted as saying that performers should "be prepared to do a lot of touring because that's really the unique situation that's going to be left". Connolly and Krueger consider trends in the industry, including the difficult problem of measuring prices over time, and also make incursions into the world of broadcasting and the intellectual property issues that have appeared with the advent of file sharing via the Internet.

Finally, book publishing has also become an industry, in which authors, publishers, retailers and readers do not always agree, and have often called for government intervention, at least in Europe. Marcel Canoy, Jan van Ours and Frederick van der Ploeg (Chapter 21) describe these aspects, basing their analysis on insights from the theory of industrial organization. They conclude that there is in fact little need for public interven-

tion, except perhaps to promote reading. The chapter also provides a host of statistics on various aspects of the industry.

The four chapters devoted to the cultural industries outlined above make use of the theory of industrial organization, but they also draw attention to the special conditions under which art markets operate that lead to new insights and new models. Economic analysis of the cultural industries is obviously a field that draws upon, and can also contribute to, both theoretical and applied economics. In particular, as is made clear in Caves' contribution, there exists a host of contracts that look unusual, suggesting that this is a field that will repay further exploration using contract theory.

## 4. Behavior of markets

### 4.1. Historical aspects

Markets have always existed in the arts, and it is instructive for anyone interested in present-day markets in the arts and cultural sectors to look back to how markets for artistic goods and services have evolved over time. The first art markets in our meaning of the word flourished in the Renaissance in Italy (Florence) and in Flanders (Bruges and Ghent). The Dutch art market emerged somewhat later. Subsequently secondary or resale markets began to grow, leading to the development of auctions, mainly in Northern Europe and the United Kingdom. Several of the features of early auction processes survive today, as Neil De Marchi and Hans Van Miegroet (Chapter 3) demonstrate in their account of the evolution of art markets from the 14th to the 18th centuries. They point out that 17th century auctioneers produced printed sales catalogs containing a number of rules, including:

- successful bidders were required to pay a part of the price before leaving the sale (thus minimizing the risk of irresponsible bidding), a practice represented by the numbered paddle which identifies the buyer in contemporary salerooms;
- to accelerate the sale, some auctioneers suggested a minimum increment between bids – the contemporary equivalent of the “tick”;
- commissions, that is bids by the auctioneer himself or on behalf of absentees, were not allowed – this is not the case nowadays;
- in some cases, auctioneers committed to the disclosing of reserve prices – a rule that is partly followed in contemporary auctions, when a relation between the printed lower pre-sale estimate and the reserve price is announced.

De Marchi and Van Miegroet, as well as the late Michael Montias with his important work on Dutch inventories [Loughman and Montias (2000)], show how archives can be used to provide information on market structures which took about 250 years to develop; using such archives it is possible to demonstrate that by 1750 most of the ingredients of the art market that we know today were already present.

Similar remarks might be made about the evolution of markets for artistic labor. For example, employment of composers has always depended on the existence of a sufficiently lucrative market for the works they produce. F.M. Scherer (Chapter 4) points out



that employment possibilities for composers in the 17th century were mainly provided by courts and the church, though after 1750 the increase in private wealth led to more support from individual patrons, at the same time as publishing and concert tours began to provide more opportunities. The role of education in influencing musical tastes is illustrated in Scherer's analysis by the prevalence of music in Austria and Czechoslovakia from the mid-17th to the mid-19th centuries, a phenomenon he attributes to the schooling system in which schoolmasters, even in small villages, were expected to teach singing and violin to their pupils.

#### 4.2. *Present-day art markets and prices*

The two chapters discussed in the previous section illustrate that markets in the arts have existed for many centuries, but that their forms have evolved, quite dramatically in some cases. In the case of markets for art objects such as paintings, sculpture, jewelry, antiques, etc., most of the dealing today is still done by dealers and galleries as in the past. Auctions, however, whether occurring in salerooms or over the Internet, have become more important. The leading role played by auctions derives from the fact that this is where important works are sold. Auctions serve as a reference to the market in general, since they constitute the almost unique way to observe how the market and prices evolve over time.

Anderson (1974) and Stein (1977) were among the first to look into the financial performance of art, using some econometrics. They were followed by Baumol's (1986) very influential paper based on prices obtained at auction over a 300-year period (1650–1960). Baumol found that the average real rate of return was equal to 0.55 percent per year, some 2 percent lower than the return on bonds; the difference, according to Baumol, is attributable to the return provided by aesthetic pleasure. Baumol's paper was followed by a very large number of studies that almost invariably used prices observed at auctions. This led economists who were interested in art markets to take a closer look at how the art auction system works. In their wide-ranging review of this work, Orley Ashenfelter and Kathryn Graddy (Chapter 26) deal with two important issues. They first show how auction prices have been used to derive conclusions on: returns to art as an investment; portfolio diversification; the "masterpiece" effect (masterpieces are claimed to reap higher returns, which does not seem to be true); and the "burning" effect (the effect on the future value of a work if it is unsold). Second, they describe the various inefficiencies to which art markets and/or auctions are prone: prices of homogeneous lots sold in sequence decrease over time (the declining price anomaly); the law of one price does not hold, since systematic price differences across salerooms and/or countries are observed and may persist; and pre-sale estimates made by salerooms do not reflect all the information that is available to experts.

The difficulty in measuring returns and constructing price indices is due to the heterogeneous character of artworks. In order to homogenize, researchers have based their computations on resales of identical works only. This entails important limitations:

- the number of resales is often small and represents a small proportion of total sales;

- resales may embody a sample selection bias; and
- unobserved resales may also take place between observed sales at auctions and if one takes these into account, returns may turn out to be very different from those usually obtained.

The first two of these limitations are partly circumvented if all sales are accounted for and hedonic regression is used instead of repeat-sales regression.<sup>2</sup> Two of the estimation techniques developed to deal with the computation of returns (hedonic and repeat-sales regression) are surveyed by Victor Ginsburgh, Jianping Mei and Michael Moses (Chapter 27), who discuss some technical details that may help researchers to decide which technique to use given the data at hand, or how data should be collected. Some preliminary simulations reported in this chapter point to the fact that hedonic regression is as good as repeat-sales regression when the number of observations is large enough, but that hedonic regression gives better results when this number is small – this is usually the case when the time span is small, or when it is necessary to disaggregate and deal with finer markets or with individual artists.

The special case of copies of artworks presents a number of puzzling issues. While copies of artworks were sold at half the price of originals during the 17th century and before [De Marchi and Van Miegroet (1996)], the situation changed towards the end of the 19th century, at which time the relative price of copies dropped very sharply. Nevertheless copies continue to be produced and sold, raising a number of issues for economists, lawyers, philosophers, art historians and curators. In discussing these issues, Françoise Benhamou and Victor Ginsburgh (Chapter 8) point out that the development of a market for copies is part of a wider contemporary questioning of the boundaries between originality and copy. They analyze whether and how the various participants in the art market contribute to valuing (or de-valuing) copies, alongside markets for original works of art.

#### 4.3. Legal regulation affecting markets

There are several ways in which the interactions between law and economics find application in the arts. The most obvious revolves around property rights and transactions costs; following Ronald Coase it can be proposed, for example, that beneficial or costly alternatives, of which there are numerous examples in the arts, can be resolved through market processes if property rights can be defined and enforced and if transactions costs are negligible. Most of the theoretical and applied work of relevance to the arts in this expanding field has been concerned with the rights of artists, especially copyright. Copyright obviously puts restrictions on markets but, as William Landes (Chapter 7) points out, it can be rationalized as a mechanism for protecting the creative expressions of authors and artists and as a means of promoting economic efficiency in markets for

<sup>2</sup> Goetzmann (1996) points to an additional problem that is encountered when using repeat-sales regression, the risk of obsolescence: many artworks are never resold because they disappear for natural reasons (fire, physical decay, etc.) or they are dumped because they are valueless or at least thought to be such.

creative work. As well as discussing the economic rationale for copyright, Landes also considers other rights of artists, including moral rights, which protect authors against alterations and destruction of their works. The chapter also considers disputes of ownership, authenticity issues and resale rights, illustrating these various problems through cases.

New communications technologies present new challenges to the assertion of artists' rights. Patrick Legros ([Chapter 9](#)) argues that the so-called curse of the Internet, and of new technologies which allow for infinite copying, may in fact be a blessing rather than a threat to artists, since these developments foster the emergence of new artworks. Legros compares this situation to that which prevailed when photography was born. Artists can show their work on the Internet, bypassing any form of intermediation, and can still obtain revenues from their creative ideas, as long as interpretation is needed. He shows that strict copyright rules may have little to do with social efficiency, and that the preferences of artists for strict or weak copyright are related to their creativity. Good and bad artists may bless the curse, and opt for weak copyright, while others may feel protected by strong copyright.

Restrictions of another kind are imposed by censorship on the creation of art. The practice of censorship is inherited from ancient times. As Tun-Jen Chiang and Richard Posner ([Chapter 10](#)) point out, offensiveness, whether religious, political or scientific, is a concept that exists in all countries, is not objective, and changes over time. It affects markets directly, and may cause the emergence of informal or underground transactions as a means of circumventing the censorship restrictions. The most obvious cases of censorship are seen in literature and the visual arts, although theater and even music have been affected at various times.

## **5. Arts and culture in the economy**

We turn now to the role of arts and culture in the economy at large. At the outset we can observe the structural transformations that have affected economies in the industrialized world in recent times, loosely described as a transition from a manufacturing economy to a services economy and onwards to an information or knowledge-based economy. The so-called "new economy" which is emerging as this transformation process continues depends on increasingly rapid technological change in the communications sector. In the new business environment that has evolved as a result of these developments, creativity is seen as a key resource through which firms can maintain a competitive edge. This in turn throws the spotlight on the creative industries as discussed above, which are increasingly being promoted as primary sources of creative ideas to sustain the expansion of the new economy. The revolutionary growth of the new economy has profoundly affected the arts, as William Baumol ([Chapter 11](#)) argues. He points to the wide-ranging opportunities that new communications technologies open up for artistic creation, dissemination and preservation. In turn these developments raise pricing

problems surrounding the trade-off between the encouragement of creativity and the facilitation of dissemination.

Despite these changes, however, the interest of economists in the role of art and culture in the economy has focused on more traditional concerns – how does culture affect economic performance in developed and developing countries, and what are the many ramifications of the intersections between cultural and economic policy? We consider these issues in the remainder of this section.

### *5.1. Culture in economic performance*

If culture is defined as shared values and beliefs that identify and bind a group together, its influence on economic performance in developed economies can be interpreted at various levels corresponding to the extent of the group concerned, ranging from the individual firm or organization up to the level of the whole economy. The shared values might influence the articulation of the group's objectives and/or they might affect the group's economic efficiency in various ways – for example, does a shared Protestant work ethic increase labor productivity in the firm or in the economy? Mark Casson (Chapter 12) sees the culture of a group as being a particular type of public good created by leaders and shared by their followers. This may set up a basis for competition between groups. He defines four key dimensions of culture (individualism vs. collectivism, pragmatism vs. proceduralism, low-trust vs. high-trust, and high-tension vs. low-tension) and shows how they influence economic performance, leading to the rise and decline of organizations and of whole societies.

Somewhat similar remarks could be made about developing countries, where it might be expected that shared values, beliefs and cultural practices of various sorts would be closely related to economic performance. However, traditional development economics has largely neglected culture as an element in economic development, most writers either seeing culture as an obstacle to development or ignoring it altogether. Paul Streeten (Chapter 13) argues that this situation has changed recently through a shift in thinking from a goods-centered to a human-centered view of the development process. Such a shift is consistent with a more holistic view of the developing economy, in which economic, social, cultural and environmental systems are seen as interrelated; in these circumstances economic and cultural development can be seen to advance hand in hand. Streeten goes on to consider the role of conflict in bringing about social change, the impacts of globalization on the cultures of developing countries, and the economic significance of tourism in the Third World.

### *5.2. Cultural policy*

Cultural policy is emerging as an increasingly significant component of government policy formation at both national and international levels. There has been considerable scholarly interest in cultural policy studies in recent years in several countries, but much

of it derives from disciplines outside of economics.<sup>3</sup> Yet the intersections between cultural and economic policy are extensive, including in the following areas:

- the prospects for the cultural industries as dynamic sources of innovation, growth and structural change in the new economy, as discussed above;
- the appropriate levels for government support for the arts and culture, and the balance between direct (grant-related) and indirect (tax-related) means of support;
- trade policy in respect of cultural goods and services;
- the possibilities for public/private partnerships in the preservation of cultural heritage;
- the role of the arts and culture in employment creation and income generation in towns and cities, especially those affected by industrial decline; and
- legal and economic questions concerning the regulation of intellectual property in cultural goods and services, as discussed earlier.

At an international level, a newly-adopted UN convention on cultural diversity will, if eventually ratified, have important implications for cultural policy in both the developed and the developing world.

Having outlined the broad reach of cultural policy in general terms, we can go on to point out that its coverage and its level of importance in national policy agendas vary considerably between countries. In Europe, for example, there are substantial differences between countries in the volume of public resources they devote to the arts and culture and in the ways those resources are deployed. Administrative structures also vary, as Frederick van der Ploeg (Chapter 34) points out, ranging from arms-length arts councils to full-scale Ministries of Culture. There are also sub-regional and supra-regional interests in cultural policy in Europe; on the one hand sub-national regions with distinctive cultural identities promote their own artistic and cultural endeavors, on the other hand the European Union and the Council of Europe have interests in developing pan-European cultural policies in various fields.

In the United States, by contrast, there has never been a great deal of interest in the formulation of specific cultural policies at the central government level; rather the major players in the formation and execution of direct public expenditure programs in the arts and culture are located at sub-national levels of government. Nevertheless, as Dick Netzer (Chapter 35) points out, the largest share of government support for the arts and culture is provided indirectly, via tax concessions for gifts to not-for-profit arts organizations and also via tax concessions to private owners of cultural heritage buildings and sites. The role of tax incentives in stimulating cultural donations and the significance of philanthropy as a characteristic of American life are discussed by Mark Schuster (Chapter 36) and Stanley Katz (Chapter 37), respectively. Both authors refer to the peculiarly American tradition of individual and corporate giving that has sustained the growth of the arts in the United States for more than a century. Despite

<sup>3</sup> See, for example, the collection of readings in Lewis and Miller (2003), which assembles articles drawn from cultural studies journals and texts that focus on the political and social role of culture.

the emergence of what many see as a distinctive and successful “American model” for cultural support, Schuster warns that it may be difficult to replicate it in other countries where institutional structures, income levels and tax-price responsiveness may differ markedly from the US situation.

### 5.3. *Some specific policy areas*

In this section we take up four specific areas where cultural and economic policy intersect: support for the arts; trade in cultural goods; heritage; and urban and regional development issues.

First, as we noted in Section 1 above, the rationale for state support for the arts has been one of the longest-running issues to be discussed in the literature of cultural economics over the last 40 years. Despite the normative appeal of market-failure arguments, the empirical evidence is by no means extensive, and positive theories of assistance may be more useful in explaining levels and patterns of support actually observed in various democratic countries around the world. One of the earliest contributions to the debate was Alan Peacock’s (1969) classic paper on the welfare economics of public support for the arts, and he returns to this question in the present volume (Chapter 32). Peacock asks why, at a time when the trend is towards privatization in many sectors of the economy, the provision of government grants, tax relief and regulatory protection continues in the arts and heritage industries. The answer, he suggests, lies in governments’ skepticism that consumers know what is best for them; in these circumstances producer interests can exert a strong influence on government policy, with consequent effects on both allocative and productive efficiency.

The second specific policy area of interest is cultural trade. Acheson and Maule (Chapter 33) point out that there have always been international dimensions to cultural production, consumption and exchange, with longstanding traditions of movements of artists between countries, and of trade in cultural products. Nevertheless, the interests of contemporary economists in this field have focused on a particular issue of growing concern – whether cultural goods and services such as movies, television programs, magazines, etc. should be regarded differently from other merchandise in multilateral and bilateral trade negotiations. On the one side are those who argue that free-trade principles should apply universally, on the other are countries such as Canada and France who believe their own cultural identities are threatened by allowing open access to cultural products from other countries, especially from the US. Acheson and Maule suggest that arguments for protection versus openness for cultural activities are more complex and nuanced than for other economic sectors because of a wide range of views on how international cultural policy affects individuals and the national culture.

Thirdly, we turn to cultural heritage. Policy in this area has until quite recently been determined largely by cultural experts – archaeologists, anthropologists, museum curators, architectural historians, and so on. When economists first started asking questions about the opportunity costs of resources used in heritage preservation, or how the benefits of heritage were to be evaluated, their intrusion onto sacred ground was resented.

Now, however, there appears to be a stronger sense that economists have methods for analyzing heritage that can usefully inform policy-making processes, and that they are not insensitive to the cultural values that heritage yields. In particular the concepts of cultural capital and sustainability as discussed by Ilde Rizzo and David Throsby (Chapter 28), hold promise for representing and analyzing heritage conservation decisions in ways that link the economic and cultural dimensions of the problem. Valuation methods for non-market benefits that have been developed in the field of environmental economics have found ready application to cultural heritage assets, emphasizing the formal parallels between cultural and natural capital. The economics of cultural heritage is also particularly interested in policy-making processes in different jurisdictions, for example, in analyzing the optimal level of decentralization of decision-making in multi-jurisdictional systems.

An important means by which heritage is conveyed to the public is via museums, whether these are art museums displaying moveable heritage items such as paintings, sculptures and artifacts, or museums built around particular heritage buildings or sites. The number of museums has increased dramatically in recent years, and so have numbers of visitors. But the objectives of museums are also becoming more diversified and complex, as is stressed in the contribution by Bruno Frey and Stephan Meier (Chapter 29). The question of “what should be maximized” is becoming more crucial as organizations become more complex, and depends on whether a museum is private or public. Increasing the number of visitors is certainly a legitimate objective, but involves new challenges: rooms may be congested; services aimed at specific audiences have to be developed, etc. The resulting choices have direct implications for entrance fee policies, for the amount and diversity of funding (whether private or public), for acquisitions, for the presentation of permanent collections and for special exhibitions. Museums are generally wealthy, given the value of the assets they hold; at the same time, they are poor, since their assets are frozen and operating costs are growing. Frey and Meier discuss a number of issues related to museum economics including de-accessioning (selling works from the collection), the role of superstar museums, and the growing importance of blockbuster exhibitions.

The final cultural policy area to be considered here relates to the geographical location of arts and cultural facilities and of cultural industries. Given the relatively labor-intensive character of artistic production, and given also the role of the arts in fostering social cohesion, it is not surprising that the establishment of arts facilities, etc. has been seen as a means of urban revitalization and employment creation, especially in depressed areas. The attention of economists in this field has been directed towards quantifying the economic effects – both direct and indirect, tangible and intangible – of such processes. Studies have looked at the short-run employment and income-generating impacts of cultural facilities and events, and at the longer-term contribution of the arts and culture to sustained regional growth and development. In reviewing this field, Trine Bille and Günther Schulze (Chapter 30) look at the culture/development nexus from two points of view: they ask on the one hand how the arts and culture can generate urban development, and on the other how regional economic growth can in-

fluence the cultural sector. They point to agglomeration processes in cultural industry location, a matter taken up by Walter Santagata ([Chapter 31](#)) in his exposition of the phenomenon of “cultural districts”. In Santagata’s analysis, these are districts where firms producing particular cultural goods tend to cluster, opening up the opportunity for them to establish a collective intellectual property-right over their marketed output. Experience in developed countries in the establishment of such rights may be taken to indicate some potential for encouraging the formation of such cultural districts in the developing world.

## 6. Conclusions

The overview of the economics of art and culture that I have presented in this chapter has been intended to provide both an interpretation of the principal concerns of the field of cultural economics, as well as an introduction to the various chapters in this volume. Readers are now invited to turn to the chapters themselves, where they will find a much fuller discussion of the issues touched upon only fleetingly above.

There are a number of topics that my co-editor and I decided not to treat in this volume and that could, under a liberal interpretation of culture, have been included. An obvious one is the economics of religion, an area that has been gaining some prominence in recent years and that is pigeon-holed alongside the economics of art and culture under the “Z” classification in the *Journal of Economic Literature*. The economics of language is another area deserving of consideration, given the ubiquitous presence of language in many types of economic interactions. More broadly, an extension of the concept of culture further into popular culture than is attempted in this volume would open up the economics of sport and a wider range of media issues, taking us into areas that are so substantial they may well warrant separate volumes of their own.

Finally, it is worth noting that recent years have witnessed a steady expansion in the number of economists taking an interest in the economics of art and culture. Most chapters in this volume make suggestions for promising lines of future research; if these suggestions are taken up we are likely to see a continuation of this expansion in the future.

## References

- Anderson, R.C. (1974). “Paintings as an investment”. *Economic Inquiry* 12, 13–25.
- Baumol, W.J. (1986). “Unnatural value or art as a floating crap game”. *American Economic Review* 76, 10–14.
- Baumol, W.J., Bowen, W.G. (1966). *Performing Arts: The Economic Dilemma*. Twentieth Century Fund, New York.
- Blaug, M. (Ed.) (1976). *The Economics of the Arts: Selected Readings*. Martin Robertson, London.
- Blaug, M. (2001). “Where are we now on cultural economics?” *Journal of Economic Surveys* 15 (2), 123–141.



- De Marchi, N., Van Miegroet, H. (1996). "Pricing invention: Originals, copies and their relative value in seventeenth century Netherlandish art markets". In: Ginsburgh, V., Menger, P.-M. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier, Amsterdam, pp. 27–70.
- Ginsburgh, V. (2001). "Economics of art and culture". In: Smelser, N., Baltes, P. (Eds.), *International Encyclopaedia of the Social and Behavioural Sciences*. Elsevier, Amsterdam.
- Goetzmann, W.N. (1996). "How costly is the fall from fashion: Survivorship bias in the painting market". In: Ginsburgh, V., Menger, P.-M. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier, Amsterdam, pp. 71–84.
- Hutter, M., Throsby, D. (Eds.) (2006). *Beyond Price: Value in Culture, Economics and the Arts*. Cambridge University Press, New York.
- Lewis, J., Miller, T. (Eds.) (2003). *Critical Cultural Policy Studies: A Reader*. Blackwell, Oxford.
- Loughman, J., Montias, J.M. (2000). *Public and Private Spaces: Works of Art in Seventeenth-century Dutch Houses*. Waanders, Zwolle.
- Peacock, A. (1969). "Welfare economics and public subsidies to the arts". *Manchester School of Economic and Social Studies* 37 (4), 323–335.
- Rosen, S. (1981). "The economics of superstars". *American Economic Review* 71, 845–858.
- Stein, J.P. (1977). "The monetary appreciation of paintings". *Journal of Political Economy* 85, 1021–1035.
- Throsby, D. (1994). "The production and consumption of the arts: A view of cultural economics". *Journal of Economic Literature* 32, 1–29.
- Throsby, D. (2007). "Art, economics of". In: Blume, L., Durlauf, S. (Eds.), *The New Palgrave Dictionary of Economics*, second ed. Palgrave Macmillan, London. In press.
- Towse, R. (Ed.) (1997). *Cultural Economics: The Arts, the Heritage and the Media Industries*, 2 vols. Edward Elgar, Cheltenham.
- Towse, R. (Ed.) (2003). *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham.

## ART AND CULTURE IN THE HISTORY OF ECONOMIC THOUGHT\*

CRAUFURD GOODWIN

*Duke University, USA*

### Contents

Abstract	26
Keywords	27
1. Introduction	28
2. The earliest years	29
2.1. Art and culture as luxury, waste and dissipation	29
2.2. Early reflections on price, patronage and externalities	31
3. The eighteenth century: Imitation and imagination in the Enlightenment	33
3.1. David Hume	33
3.2. Anne Robert Jacques Turgot	34
3.3. Adam Smith	37
4. Classical economics: The shadow of Bentham	41
5. Humanist critics	48
5.1. Matthew Arnold	49
5.2. John Ruskin	50
5.3. William Morris	52
6. The marginal revolution	53
6.1. The marginalists' treatment of arts and culture	53
6.2. Alfred Marshall	54
6.3. William Stanley Jevons	54
6.4. Lionel Robbins	58
7. The arts in American economics	59
8. Keynes and the Bloomsbury Group	61
8.1. The arts in human life	61
8.2. The character of the artistic experience	62
8.3. The arts in society	63

\* I am grateful to the History of Economic Thought Workshop at Duke University, to Harro Maas, and to Sir Alan Peacock for helpful comments and especially to Jonathan Greene for research assistance. Research support was generously provided by the Henry Luce Foundation.

8.4. Art markets	63
8.5. Policy reform	64
9. Conclusion	66
References	66

## Abstract

Attention to art and culture goes far back in the history of economic thought. In the seventeenth century those activities were viewed suspiciously as likely to be either wasteful extravagances of the aristocracy, or dangerous distractions for the working classes. Eighteenth century economic thinkers offered more positive and thoughtful speculations. Mandeville and Galiani observed that the prices of art works were determined almost entirely on the demand side of the market, often by fashion and the search for distinction. The Enlightenment economic thinkers were intrigued by various aspects of art markets. Hume and Turgot perceived positive social benefits emerging from the arts, and they attempted to understand of what these consisted. Smith picked up some of the hints that were dropped and looked at art markets in a depth that had not been undertaken before. Like some other Enlightenment thinkers, Smith pictured the arts as being mainly about the imitation of perfection.

Jeremy Bentham, with his emphasis on utility as a tool by which both to understand and judge market performance, insisted that the arts should not be distinguished from other forms of entertainment: pushpin, he asserted, equals poetry. Other political economists followed Bentham's lead and steered away from exploration of the economics of the arts. To some extent the void thus created was filled by humanistic writers, novelists, and essayists, notably Arnold, Ruskin, Dickens, and Morris, who were highly critical of the industrialization of the period and the emerging discipline of political economy that they perceived to go with it.

In the "marginal revolution" of the 1870s the Benthamite injunction against special treatment for the arts was largely observed. At the same time, several of the new economists, notably William Stanley Jevons, became "closet esthetes", enjoying their guilty pleasures but not often subjecting the arts to economic analysis.

Disappointingly little concerning the arts and culture can be found in the distinctive American economics of the late nineteenth and early twentieth century. There was almost a reversion to the seventeenth century view of the arts as the corrupt playthings of the idle rich. However, something like a return to the rich speculation of the eighteenth century Enlightenment occurred in the Bloomsbury Group that included the economist John Maynard Keynes. They rejected "Benthamism" and distinguished between the artistic experience and human consumption, and between the "imaginative life" of the mind and the biological activity of humans and other creatures. They discerned complex effects of the arts throughout society and placed arts policy high on the policy agenda.

**Keywords**

arts policy, cultural policy, art and economics, economics of art

*JEL classification:* B1, Z1

## 1. Introduction

What may modern cultural economists learn from this chapter which reviews what earlier, mainly English-speaking, authors have written about the economics of art and culture? They may discover that their subject has roots that are far deeper than they ever imagined. And some of these roots have vigorous life still. “Cultural economics” came on the scene as a full-fledged sub-discipline of economics in the 1960s, marked soon by a journal, an association and canonical texts such as *Performing Arts: The Economic Dilemma* (1966) by W.J. Baumol and W.G. Bowen. In fact, however, many of the topics addressed by these new cultural economists had been under discussion already with greater or lesser sophistication for two centuries or more. The modern economist may find in this disparate literature insights and interpretations that have continuing relevance. Discovery and rehabilitation of some of these early texts may at a minimum reduce the need to reinvent the wheel. It is sobering to reflect on how long economists have struggled over some questions in cultural economics that now seem almost eternal: How to recognize and specify externalities arising from activities in the arts? Are there perhaps negative as well as positive externalities associated with the arts? Many in the eighteenth and nineteenth centuries thought so! Is there something about the arts that distinguishes their production and consumption from those of other goods that occur in society? Or could it be that the arts are simply a complex category of consumption dominated by conspicuous behavior and with its own collection of rent seekers? What forms of patronage work best to sustain and stimulate the arts, if they do indeed have special merit and deserve some sort of payment outside the market? What conditions strengthen the creativity of artists? Can the free market, even under the best of competitive conditions, be counted upon to sustain the arts and culture at levels approaching those that are socially optimal? If not, what is the proper role for the state as supporter, planner, and operator in this field? Is it possible that the best response of the state when faced with entreaties from the arts is to “just say No”? Definitive answers to these questions are not likely to emerge from an examination of the pre-history of cultural economics, but some light may be cast.

The terms art and culture came into common use with their modern meaning only in the late eighteenth century, and then as an evolution from earlier meanings. “Art” moved from signifying a particular skill to “a set body of activities of a certain kind”, “a particular group of skills, the ‘imaginative’ or ‘creative’ arts” [Williams (1958, p. xv)]. At the same time “art came to stand for a special kind of truth, imaginative truth”. “*The arts* – literature, music, painting, sculpture, theater – were grouped together, in this new phrase, as having something essentially in common which distinguished them from other human skills” (pp. xv, xvi). The word “culture” changed in rather the same way as “art” from meaning the “tending of natural growth” to several other definitions including “a general state or habit of the mind”, “the general state of intellectual development, in a society as a whole”, “the general body of the arts”, and “a whole way of life, material, intellectual and spiritual” (p. xvi). It must be remembered that at different times all these various meanings were intended by writers on economics as well as on

other topics over the several centuries spanned by this chapter. The chapter is broadly chronological, beginning in the late sixteenth century and culminating around the middle of the twentieth, i.e. just prior to the emergence of the “modern” sub-discipline of cultural economics referred to above.

## 2. The earliest years

### 2.1. *Art and culture as luxury, waste and dissipation*

Writers on economic topics before the middle of the eighteenth century were deeply conscious of the problem of scarcity and the need for careful allocation of limited resources. In a subsistence economy the price of waste could be starvation and death, and at a minimum waste meant sacrificing an important alternative goal. Sometimes the norm prescribed or implied for allocation of resources was achievement of the good life, sometimes it was national power, and sometimes it was the smooth running of a fast-developing market economy. For the most part economic writers appreciated that resources when wisely used could feed the working class, sustain the state in its various projects or adventures and permit the production of investment goods that would achieve growth in the future. Art and culture had little place in this picture, except in a few cases as means to some worthy end; they were at best an enigma and at worst an annoyance. Could they, critics asked, be any more than the extravagance of the monarchy, the aristocracy and the church, and therefore increasingly inconsistent with modernity? Were they not, as the modern economist might ask, in all respects noxious and demeritorious goods? Some early writers thought so – for various reasons. There was not organized controversy in this literature on this subject as much as uninhibited, undisciplined and somewhat random commentary.

Favorable attention was given from time to time to art and culture in the mercantilist literature when it could be shown that domestic production of artworks would contribute to a more favorable balance of trade, or when it was noted that domestic expenditures on culture sustained employment during a recession. But attention sometimes was also unfavorable. For example, in the 1730s there was a move to limit the number of playhouses in London as a way to reduce the urban vice that was said to be corrupting the nation. In the view of the pamphleteer Erasmus Jones, “the town” was fast becoming “a forge of vanity, a nursery of vice, a snare to the young, a curse to the old, and a perpetual spring of new temptations” [Jones (1736, p. 31)]. The arts were part of “our pretended diversion, viz. horse-races, cockpits, tennis-courts, balls, assemblies, and musick entertainments, &c . . . We are like so many bees in a garden, humming and roving from one flower to another, foolishly endeavouring to keep up our course of pleasure, by a continued succession, and circle of varieties” (p. 32). In this pamphlet commercial activity is presented as the only virtuous activity outside the home, while the notorious revels of the upper classes were the worst sort of vice. “Virtue only deserves the name of business, and none but they that practice it, can be truly said to be employed; for all

the world beside are idle” (p. 33). The arts necessarily suffered here by the division of all human activity into the vigorous and virtuous on one side and the idle and vicious on the other; they were placed unambiguously in the latter category.

Before the emergence of an economics discipline at the end of the eighteenth century, writers on economic topics often focused on the “passions” that supposedly motivated human action in markets as elsewhere. For the most part they perceived an inevitability about these passions that made it difficult to constrain, control, or balance them. However, to complicate the matter the passions – and the vices to which they sometimes gave rise – clearly helped to make the world go round or at least the economy to perform, and it might be dangerous to meddle with them. Bernard Mandeville’s recipe for economic growth depended heavily on arousing and harnessing the passions, among which envy, pride and covetousness loomed large. “Would you render a society of men strong and powerful, you must touch their passions. Divide the land, tho’ there be never so much to spare, and their possessions will make them covetous: rouse them, tho’ but in jest, from their idleness with praises, and pride will set them to work in earnest. Teach them trades and handicrafts, and you’ll bring envy and emulation among them . . .” [Mandeville ([1732] 1924, I, p. 184)].

Envy and emulation were seen by Mandeville as the forces that drove the art markets in particular: “the worldly-minded, voluptuous and ambitious man, notwithstanding he is void of merit, covets precedence every where, and desires to be dignify’d above his betters: He aims at spacious palaces, and delicious gardens; his chief delight is in excelling others in stately horses, magnificent coaches, a numerous attendance, and dear-bought furniture . . . His table he desires may be served with many courses, and each of them contain a choice variety of dainties not easily purchased, and ample evidences of elaborate and judicious cookery; while harmonious musick and well couched flattery entertain his hearing by turns” (p. 148). In Mandeville’s cynical eyes at least, the arts were not distinguishable in their role in the economy or greater in social importance than extravagant modes of transport, cuisine, and even flattery as means of responding to the demands for emulation.

Mandeville observed that exercise of the various passions through extravagant encouragement of the fine arts was happily less injurious to the health of the individual than the alternatives. Given the choice of “Wine, Women and Musick”, the well-advised wastrel following Mandeville’s suggestion chose the third. Identification of the fine arts with the luxurious vices rather than with virtues practiced in the market place, and hence seeing them as a waste of scarce resources, guaranteed them low status in the eyes of commentators on economic questions. John Bodin carried out a useful exercise in 1606 to spell out “in what order citizens are to be placed” in the “ceremonies of every city”, thereby establishing their social value. He recommended of course that in the front row should stand the monarch, followed by the clergy, senators, generals in the army, and so on. In the back there would be a motley crew, among whom would be the artists [Bodin ([1606] 1962, p. 402)].

## 2.2. *Early reflections on price, patronage and externalities*

Bodin, Mandeville, and Galiani were among the first to speculate about the determinants of the prices of art works. The general tenor of their comments was that all luxury goods, unlike subsistence goods which made up most of national product at the time and whose prices reflected their costs, were simply frivolous and their value was socially determined on the demand side. Bodin observed that luxuries rose in price when they were favored by the king because of his patronage and for no reason inherent in the things themselves: “The people conform always to the wish of the king, and consequently esteem and raise in price everything that the great lords like, though the things of themselves are not worth that valuation; for example, the Emperor Caracalla gave an inestimable worth to yellow amber, as history says, because it was like the color of the hair of his sweetheart” [Bodin ([1578] 1947, p. 34)]. Anticipating the comments of Thorstein Veblen on conspicuous consumption more than three hundred years later, Bodin went on to note that: “when the great lords see that their subjects have an abundance of things that they themselves like, they begin to despise them . . . This goes to show that the abundance of pearls has caused them to be disregarded, and from the indifference has come the low price” (p. 35). Improvidence among the rich was another reason for the relatively high price of luxuries, including works of art. In the case of precious textiles “The high price comes from the waste; for we are not content to dress the rascals and lackeys with it, but also we cut it up in such a manner that it cannot last nor serve but one master” (p. 36). Since the high prices of luxurious consumption goods were one of their attractive features for those who wished to demonstrate their wealth there were few pressures to constrain costs. “It is true that the excesses helped much to raise the price of food; for the rich, it is found, sometimes did not know how to spend their means; a thing that happened to Aesop, the tragedian, who to sharpen his appetite, had himself served a dish that cost fifteen thousand crowns, full of singing nightingales, starlings, blackbirds, and other birds, which had learned to talk, though such birds were unsavory and not tasty: however, the cost gave them a good taste” (p. 41).

Mandeville concluded that four factors were most powerful in determining the prices of art works: the fame of the artists, the reputation of their owners, their scarcity, and their faithfulness to nature. He wrote:

The value that is set on paintings depends not only on the name of the master and the time of his age he drew them in, but likewise in a great measure on the scarcity of his works, and what is still more unreasonable, the quality of the persons in whose possession they are as well as the length of time they have been in great families . . . Notwithstanding all this, I will readily own, that the judgment to be made of painting might become of universal certainty, or at least less alterable and precarious than almost anything else: the reason is plain; there is a standard to go by that always remains the same. Painting is an imitation of nature ([1732] 1924, I, p. 326).



And not only real nature, but nature with a positive spin! “It is not nature, but agreeable nature, *la belle nature*, that is to be represented; all things that are abject, low, pitiful and mean, are carefully to be avoided, and kept out of sight; because to men of the true taste they are as offensive as things that are shocking, and really nasty” (II, p. 33).

Galiani had two things to say about the determinants of price in the fine arts. The first concerned the force of fashion. “The power of fashion applies wholly to the beautiful, not at all: to the useful . . . The beautiful is of two classes: one is based upon certain ideas, which are engraved upon our mind from the beginning; the other, though it does not seem so, is merely an habituation of the senses, which makes a thing appear beautiful. It is only over this second class, which is considerably larger than the first, that the power of fashion extends” ([1751] 1930, p. 295). Galiani’s second observation about the prices of artworks concerned those that could be described as “unique, and monopolies, that is either things for which there is no substitute, such as the Venus de Medici, or those which become unique because there is only one seller” (p. 295). Galiani denied that the price of these goods was either “infinite” or “indefinite” “for I hold that every human thing possesses order and limits, and that the indefinite is no less foreign to them than the infinite. They have these limits, then: that their price always corresponds to the needs or desires of the buyer and the esteem of the seller combined, and forming a compound ratio. Hence it is that the value of a unique thing may sometimes be equal to nothing; and it is always governed by rule, though it may not be the same under all conditions” (pp. 295–296).

Mandeville argued that, at least among the fine arts, painting was concerned mainly with the magnificence and sensibilities of aristocratic patrons. “The painter has nothing to do with the truth of the history; his business is to express the dignity of the subject, and in compliment to his judges, never to forget the excellency of our species: All his art and good sense must be employ’d in raising that to the highest pitch: Great masters don’t paint for the common people, but for persons of refin’d understanding” ([1732] 1924, II, p. 35). The arts in general should represent moral principle more than an accurate reflection of nature or events. “There is a grandeur to be express’d in things that far surpasses the beauties of simple nature . . . What gentle touches, what slight and yet majestick motions are made use of to express the most boisterous passions. As the subject is always lofty, so no posture is to be chosen but what is serious and significant as well as comely and agreeable; should the actions there be represented as they are in common life, they would ruin the sublime, and at once rob you of all your pleasure” (II, pp. 36–37). Of special relevance to those concerned about the condition of social relations, the arts showed promise of modulating the human passions. In this interpretation by Mandeville the arts began to lose their reputation as always the companion of vice and dissipation and to emerge as, perhaps, a path to virtue. The opera, with the opportunity it presented for moral lessons set to music, was a prime case in point. “There is no place, where both sexes have such opportunities of imbibing exalted sentiments and raising themselves above the vulgar, as they have at the opera; and there is no other sort of diversion or assembly from the frequenting of which young persons of quality

can have equal hopes of forming their manners, and contracting a strong and lasting habit of virtue” (II, p. 40).

### **3. The eighteenth century: Imitation and imagination in the Enlightenment**

#### *3.1. David Hume*

A fresh approach to the fine arts appeared around the middle of the eighteenth century in the works of David Hume and Anne Robert Jacques Turgot. Hume rejected what he took to be the prior consensus that all “luxury”, including the fine arts, was necessarily wasteful and vicious. He argued that depending on what form it took luxury could be either virtuous or vicious, socially beneficial or pernicious. “Luxury is a word of an uncertain signification, and may be taken in a good as well as in a bad sense. In general it means great refinement in the gratification of the senses; and any degree of it may be innocent or blamable, according to the age, or country, or condition of the person. The bounds between the virtue and the vice cannot here be exactly fixed, more than in other moral subjects” [Hume ([1752] 1965, p. 48)]. In order to make the distinction between vicious and virtuous luxury Hume developed a line of argument that was to become influential in the humanities and the arts in the nineteenth and twentieth centuries, but in its attempt to establish the arts as exceptional would for the most part be rejected by the economics discipline. It was easy enough to demonstrate, Hume observed, that luxurious living carried to extremes was both wasteful and socially pernicious. But how about “innocent” luxury that included partaking of literature and the fine arts? From his historical studies Hume observed that “the ages of refinement are both the happiest and most virtuous” (p. 49). Moreover, achievements in the arts went along with accomplishments in other areas such as the economy and were not competitive with them. “The same age which produces great philosophers and politicians, renowned generals and poets, usually abounds with skilful weavers, and ship-carpenters . . . Thus industry, knowledge, and humanity, are linked together, by an indissoluble chain, and are found, from experience as well as reason, to be peculiar to the more polished, and what are commonly denominated, the more luxurious ages” (pp. 50–51). Indeed luxury seemed to provide some of the innocent incentives for economic development to proceed briskly: “where luxury nourishes commerce and industry, the peasants by a proper cultivation of the land, become rich and independent” (p. 56).

But it remained to be explained how luxury actually achieved these good results through time. Hume concluded that the answer lay in the psychological impact the innocent luxuries had on human behavior. “The mind acquires new vigor; enlarges its powers and faculties; and by an assiduity in honest industry, both satisfies its natural appetites, and prevents the growth of unnatural ones, which commonly spring up, when nourished by ease and idleness. Banish those arts from society, you deprive men both of action and of pleasure . . .” (p. 50). Hume found the taste for innocent luxuries, including the fine arts, to be closely linked to the level of contribution to private and public life.

But industry, knowledge, and humanity, are not advantageous in private life alone; they diffuse their beneficial influence on the *public* and render the government as great and flourishing as they make individuals happy and prosperous. The increase and consumption of all the commodities, which serve to the ornament and pleasure of life, are advantages to society; because at the same time that they multiply those innocent gratifications to individuals, they are a kind of *storehouse* of labor, which, in the exigencies of the state, may be turned to the public service. In a nation where there is no demand for such superfluities, men sink into indolence, lose all enjoyment of life, and are useless to the public, which cannot maintain or support its fleets and armies from the industry of such slothful members (pp. 51–52).

The arts, Hume concluded, meaning both the fine arts and all forms of human inquiry, had not only these positive economic externalities; they had political externalities as well and, indeed, were essential to a democracy. “The liberties of England, so far from decaying since the improvements in the arts, have never flourished so much as during that period . . . If we consider the matter in a proper light, we shall find, that a progress in the arts is rather favorable to liberty, and has a natural tendency to preserve, if not produce a free government” (p. 55). The arts and learning were critical to nurturing those psychological characteristics essential for civic virtue. “Here then is the chief triumph of art and philosophy: it sensibly refines the temper, and points out to us those dispositions which we should endeavour to attain, by a constant *bent* of mind, and by repeated *habit*” [Hume (1825, p. 168)]. Hume’s general conclusion seems to be that although the arts might be mildly vicious in their detraction from the production of more useful goods and services, they made important contributions by controlling other more damaging vices such as indolence and brutality.

### 3.2. Anne Robert Jacques Turgot

Like Hume, Turgot was intrigued by the differing rates of progress among nations and he speculated that the explanation might be in “the origin and growth of the arts and sciences and the revolutions which have taken place in them” [Turgot ([1750] 1973, p. 42)]. He concluded that success in the accumulation of what we would call today intellectual capital was the key to economic and political development. In the same way as Hume, Turgot drew his insight more from history than from models of economic and social behavior. In the case of economic growth he wrote: “nature, distributing her gifts unequally, has given to certain minds an abundance of talents which she has refused to others. Circumstances either develop these talents or allow them to become buried in obscurity; and it is from the infinite variety of these circumstances that there springs the inequality in the progress of nations” (p. 43). Turgot used the history of ancient Greece to demonstrate that artistic and intellectual progress lay at the heart of social development. Indeed, the arts usually came first before the humanities. “It was only after several centuries that philosophers appeared in Greece – or rather it was only then that the study of philosophy became the business of particular thinkers and appeared

sufficiently extensive in its scope to occupy them fully. Until then, the poets had been at the same time the only philosophers and the only historians. When men are ignorant it is easy to know everything” (p. 49). It was essential to the maintenance of Greek success that the fine arts retained their position of prominence. “Happy centuries, in which all the fine arts spread their light on every side, and in which the passion of a noble emulation was swiftly transmitted from one city to another! Painting, sculpture, architecture, poetry, and history grew up everywhere at the same time, as we see in the expanse of a forest a thousand different trees springing up, growing, and being crowned together” (p. 50).

No longer in Hume or Turgot is there any suggestion, as there had been in economic writers a hundred years before, that the fine arts are simply a particular kind of wasteful luxury; waste and corruption were a possibility in the fine arts, as they were in all human activities, but no more so. Turgot observed that through history “blind luxury, which, born of vanity, and judging works of art less as objects of taste than as symbols of opulence, is as opposed to their perfection as a civilised love of magnificence is favourable to it” (p. 52). Turgot drew an intriguing distinction between the sciences and the arts. Progress in the former, he suggested, was limited only by the cleverness of a few scientists acting as detectives to discover and proclaim the laws of nature. The arts on the other hand were constrained by the talent, imagination, and training of the entire artistic community: “Knowledge of nature and of truth is as infinite as they are: the arts, whose aim is to please us, are as limited as we are. Time constantly brings to light new discoveries in the sciences; but poetry, painting, and music have a fixed limit which the genius of languages, the imitation of nature, and the limited sensibility of our organs determine, which they attain by slow steps and which they cannot surpass. The great men of the Augustan age reached it, and are still our models” [Turgot ([1750] 1973, p. 52)]. It is not clear that Turgot’s distinction between the sciences and the arts can be sustained, but it does provide a case for education in and support for both these areas of human endeavor.

Turgot concluded from his historical studies that patronage of the arts was as critical to their development as was creativity within them. “Immortal names of the Medici, of Leo X, of Francis I, be consecrated for ever! May the patrons of the arts share the glory of those who cultivate them!” (p. 57). He saw troubled times ahead for the traditional patrons of the arts, but he was optimistic that out of turmoil new sources of demand would appear: “do the flowered stems of the fine arts grow when they are watered with blood? A day will come, and it is not far off, when they will beautify all the countries of Europe” (p. 57).

When he concluded that the cultural evolution of mankind helped to explain progress in all spheres including the economy, Turgot was led to explore an anthropological explanation of the emergence of the fine arts. He thought that they began with music, dance and poetry, which had “their source in the nature of man. Created to live in society, his joy manifests itself externally: he leaps and shouts. A common joy expressed itself in swaying movements, in leaps, and in simultaneous and confused shouts. Little by little people became accustomed to leap in a similar manner; the steps were marked

by sounds; and the latter were separated by regular intervals. The ear, with very little experience, and by following nature alone, learned to appreciate the primary relations between sounds. When it was desired to communicate the reason for one's joy in words, these were arranged according to the beat of the sounds. This was the origin of dance, of music, and of poetry, which was at first written in order to be sung" (p. 91). So also did philosophy emerge out of fiction. "The poverty of languages, and the necessity for metaphors which resulted from this poverty, led to the employment of allegories and fables to explain physical phenomena. They are the first steps of philosophy, as can still be seen in India" (p. 92). It was a similar story in the visual arts. "The arts of design, sculpture, and painting have many connections with poetry in the feelings which the artist experiences, and in those which he strives to communicate. They had a natural origin in the desire to preserve historical or mythological records; and genius in this sphere was heightened by that patriotic or religious zeal which sought to express with feeling, depth, and force the ideas and memories which these records were bound to recall" (p. 93).

Like Hume, Turgot suggested that progress in the arts, just as in other parts of the economy, required competition among as many aspirants as possible, a kind of evolutionary struggle that needed the presence of the unfit as well as the fit. For this there must be a sustained demand and "a market for pleasing objects and the employment of second-rate artists, among whom the great artists who shine out from them are formed" (p. 103). He contrasted the conditions for painters in several countries. "The Italians, the French, and the Flemish, and a very small number of Germans and Spaniards, have been the only ones to be successful in this art. The reason for this is that the English pay only for good pictures. By banning images from the churches, they deprived themselves of the means of supporting bad pictures, and even second-rate ones. And in all crafts where bad workers cannot gain a livelihood, and second-rate workers are not comfortably off, great men are not created" (p. 103n). The charge that the Reformation in England had paralyzed the art market and destroyed the industry by terminating the demand for run-of-the-mill pictures became a frequent subject for discussion in Britain thereafter.

Turgot found that conspicuous consumption was not a sound basis for patronage of the arts and was potentially destructive of artistic quality. When this element was dominant the arts were likely to be characterized by waves of fashion and technical virtuosity rather than by genuine creativity. Turgot's words on this point have a strikingly modern ring. "Extravagant luxury, where vanity causes ornaments to be accumulated because it regards them less as ornaments than as symbols of affluence, smothers taste. Men no longer seek for the pleasure which things afford to the senses and the mind; they no longer search their own hearts. They no longer listen to anything but fashion. The sure way to judge badly in any sphere is not to judge with one's eyes. When each individual judges, the multitude judges well, because its judgement is that of a large number of people; but when no one does anything but listen, the multitude judges badly. Another cause of bad taste was often the progress of technique in the arts. Men are always liable to mistake the difficult for the beautiful" (pp. 103-104).

Turgot was above all a social theorist rather than an empiricist and his writings are replete with speculations about the economics of the arts that might have been carried further by others and perhaps subjected to empirical test. Here is one more example. He suggested that the quality of the works of art that were produced at any time were determined by among other things the challenge presented by the materials available to the artist. For painters the challenge grew dramatically when their palette moved from black and white to polychrome materials. Great artists met this challenge successfully while lesser ones fell by the wayside. Similarly the quality of the literary arts depended upon the condition of the language in which writers worked. A complex language weeded out the weak but stimulated the strong. “The multiplicity of abstract ideas which our languages express, and which enter into our analogies, demand great dexterity in their employment. That is the disadvantage of perfected languages” (p. 112).

What economists nowadays think of as “demand” was often called “taste” in the eighteenth century, and Turgot reminded his readers that especially in the arts taste must be sustained consistently over an extended period. Great art was seldom a flash in the pan. “No art whatever can be cultivated during a long succession of centuries without passing through the hands of several inventive minds” (p. 116). A problem for the arts was that they were subject to great waves of fashion, and a variety of causes could lead to loss of demand. “Artistic taste can be lost as the result of a multitude of purely moral causes. The diffusion of a spirit of apathy and softness in a nation, pedantry, contempt for men of letters, eccentricity in the tastes of princes, tyranny, and anarchy can corrupt it” (p. 117).

### 3.3. Adam Smith

Adam Smith was exceptionally widely read and was thus able to draw from his predecessors and contemporaries when constructing his own analysis of the arts and culture. In the aesthetics that he set forth in *The Theory of Moral Sentiments* ([1759] 1976) he seemed to draw more from Mandeville than from his “never to be forgotten” teacher Hutcheson. But like Hutcheson (1729), Smith was particularly intrigued by the demand side of art markets. Why, he wondered, do people buy art? He thought that “notions of beauty and deformity” that presumably motivated buyers were affected especially by “custom and fashion, principles which extend their dominion over our judgments concerning beauty of every kind” [Smith ([1759] 1976, p. 194)]. These principles applied to all the arts. “Dress and furniture are allowed by all the world to be entirely under the dominion of custom and fashion. The influence of these principles, however, is by no means confined to so narrow a sphere, but extends itself to whatever is in any respect the object of taste, to music, to poetry, to architecture” (p. 195). Humans were naturally unwilling to concede that their responses to works of art were socially rather than individually determined, but it was true all the same. “Few men therefore are willing to allow, that custom or fashion have much influence upon their judgments concerning what is beautiful, or otherwise, in the productions of any of those arts; but imagine, that all the rules, which they think ought to be observed in each of them, are founded

upon reason and nature, not upon habit or prejudice. A very little attention, however, may convince them of the contrary . . ." (p. 195). The dominant place of fashion in the arts gave much importance to successful innovators. "An eminent artist will bring about a considerable change in the established modes of each of those arts, and introduce a new fashion of writing, music, or architecture" (p. 197). Smith cited the authority of "a learned Jesuit father Buffier" on the power of fashion in determining demand: "the beauty of every object consists in that form and colour, which is most usual among things of that particular sort to which it belongs" (p. 198).

Turning from the demand to the supply side of art markets, Smith observed that great artists seemed to develop a notion of "ideal perfection" in the arts and that they pursued and used this ideal as a goal and standard toward which they aspired even while knowing they would never achieve it.

In all the liberal and ingenious arts, in painting, in poetry, in music, in eloquence, in philosophy, the great artist feels always the real imperfection of his own best works, and is more sensible than any man how much they fall short of that ideal perfection of which he has formed some conception, which he imitates as well as he can, but which he despairs of ever equaling. It is the inferior artist only, who is ever perfectly satisfied with his own performances. He has little conception of this ideal perfection, about which he has little employed his thoughts; and it is chiefly to the works of other artists, of, perhaps, a still lower order, that he deigns to compare his own works (p. 248).

Notwithstanding the fact that great artists recognized imperfection in their work, Smith did observe a dangerous conceit among some of those who had achieved most success with the public. "Great success in the world, great authority over the sentiments and opinions of mankind, have very seldom been acquired without some degree of this excessive self-admiration" (p. 250). If not controlled, this self-admiration could lead to "a vanity that approached almost to insanity and folly" (p. 250).

In the *Wealth of Nations* Smith did not single out the arts and culture as a special topic for attention. However he did touch upon them quite often when dealing with other subjects. On the question of what determined the demand for luxury products of all kinds he was even more cynical than he had been in his discussion of the force of custom and fashion in *Theory of Moral Sentiments*. The desire for emulation was predominant. When he discussed prices for articles made from precious metals he observed that "The merit of their beauty is greatly enhanced by their scarcity" ([1776] 1976, p. 190). Then follows a passage that may have influenced Veblen in developing the notion of conspicuous consumption:

With the greater part of rich people, the chief enjoyment of riches consists in the parade of riches, which in their eyes is never so compleat as when they appear to possess those decisive marks of opulence which nobody can possess but themselves. In their eyes the merit of an object which is in any degree either useful or beautiful, is greatly enhanced by its scarcity, or by the great labour which it requires to collect any considerable quantity of it, a labour which nobody can afford

to pay but themselves. Such objects they are willing to purchase at a higher price than things much more beautiful and useful, but more common. These qualities of utility, beauty, and scarcity, are the original foundation of the high price of those metals, or of the great quantity of other goods for which they can every where be exchanged (pp. 190–191).

When he discussed the determination of the supply price for labor, Smith noted the need for compensating differentials in the wages of various occupations to cover such costs as education. He gave artists as examples as well as the learned professions. “Education in the ingenious arts and in the liberal professions, is still more tedious and expensive. The pecuniary recompence, therefore, of painters and sculptors, of lawyers and physicians, ought to be much more liberal: and it is so accordingly” (p. 119). Special compensation for trust was also due some artists. “The wages of goldsmiths and jewelers are every-where superior to those of many other workmen, not only of equal, but of much greater ingenuity; on account of the precious materials with which they are intrusted” (p. 122). On the other side of the ledger “public admiration” was for some artists a substantial part of the compensation they required to carry on their arts: “in poetry and philosophy it makes almost the whole” (p. 123). Yet the theater, opera, and dance also provided Smith examples that were the opposite of poetry and philosophy, where compensation for public obloquy was required to attract artists to these occupations. “The exorbitant rewards of players, opera-singers, opera-dancers, &c. are founded upon these two principles: the rarity and beauty of the talents, and the discredit of pursuing them in this manner” (p. 124).

On the question of productive versus unproductive labor, Smith observed that in general “The labour of some of the most respectable orders in the society is, like that of menial servants, unproductive of any value, and does not fix or realize itself in any permanent subject, or vendible commodity, which endures after that labour is past, and which for an equal quantity of labour could afterwards be procured” (p. 330). These orders included public servants and the military, and “In the same class must be ranked, some of the gravest and most important, and some of the most frivolous professions: churchmen, lawyers, physicians, men of letters of all kinds; players, buffoons, musicians, opera-singers, opera-dancers, &c. The labour of the meanest of these has a certain value, regulated by the very same principles which regulate that of every other sort of labour; and that of the noblest and most useful, produces nothing which could afterwards purchase or procure an equal quantity of labour. Like the declamation of the actor, the harangue of the orator, or the tune of the musician, the work of all of them perishes in the very instant of its production” (p. 331). In this distinction Smith emphasized that the results of unproductive labor did have legitimate value; they were simply intangible and could not lead directly to capital accumulation and growth. At the same time his use of words like “frivolous” and “meanest” to designate the arts harked back to the denigration of the Mercantilist pamphleteers.

In his later life Smith turned more seriously to the study of the fine or “imitative” arts as he called them. An essay by him on the subject was published posthumously



in his *Essays on Philosophical Subjects*. The relationship between the arts and imitation, a subject discussed by Aristotle in his *Poetics*, remained a puzzle for 18th century thinkers. Smith's essay reads like the musings of someone who has not himself had many aesthetic experiences but knows others who have had and wishes to understand them. He extends Hutcheson's (1729) notions that the arts are mainly about imitation, but like Burke ([1757] 1869) he stresses that they are not simply about copying; the arts involve the search for resemblance in a wider sense. "In Painting, the imitation frequently pleases, though the original object be indifferent, or even offensive" ([1795] 1980, p. 179). The interpretation that involves imitation often upgrades the original subject: "art cannot, without degrading itself, stoop to represent any thing that is offensive, or mean, or even indifferent" (p. 179). For imitation to be pleasing it is necessary for there to be sufficient disparity between "the imitation and the imitated object". This is why artistic success is harder to achieve in sculpture than in painting where all the colors of the palette are available to introduce variation (p. 180).

In this essay Smith returned to a point he had made in *Wealth of Nations*, that the arts could be addressed to various audiences, on the one hand to "the prudent and the wise", and on the other hand "to the rich and the great, to the proud and the vain" (pp. 182–183). With the latter "we ought not to wonder if the appearance of great expence, of being what few people can purchase, of being one of the surest characteristics of great fortune, should often stand in the place of exquisite beauty, and contribute equally to recommend their productions. As the idea of expense seems often to embellish, so that of cheapness seems as frequently to tarnish the lustre even of very agreeable objects" (p. 183). Smith used the case of topiary in different countries to demonstrate how relative costs and prices affected demands for works of art by the "proud and vain". In France low agricultural wages had made topiary widely available and therefore no longer able to signal great wealth. In England by contrast "Such ornaments, not having in that country been degraded by their vulgarity, have not yet been excluded from the gardens of princes and great lords" (p. 184).

In the second part of his *Essay on the Imitative Arts* Smith made what would become an heretical assertion for most later economists, that the products of the imitative arts are a distinctive feature of human civilization. Moreover the pattern of artistic production changes with the progress of mankind. Like Turgot he offered an anthropological hypothesis. First came music and dance: "In the progress of art and improvement they are, perhaps, the first and earliest pleasures of his own invention" ([1795] 1980, p. 187). But among the products of civilization these were inferior goods whose relative demand decreased with progress; they were too time-consuming for advanced societies: "In civilized nations, the inferior ranks of people have very little leisure, and the superior ranks have many other amusements; neither the one nor the other, therefore, can spend much of their time in Music and Dancing. Among savage nations, the great body of the people have frequently great intervals of leisure, and they have scarce any other amusement; they naturally, therefore, spend a great part of their time in almost the only one they have" (p. 187). The third of "those three sister Arts" that appeared in civilized societies after music and dance was poetry, an art form that demanded more of the artist and

the listener. “Poetry, however, is capable of expressing many things fully and distinctly, which Dancing either cannot represent at all, or can represent but obscurely and imperfectly; such as the reasonings and judgments of the understanding; the ideas, fancies, and suspicions of the imagination; the sentiments, emotions, and passions of the heart. In the power of expressing a meaning with clearness and distinctness, Dancing is superior to Music, and Poetry to Dancing” (p. 189). Smith proposed here a stage theory of cultural development and a qualitative hierarchy, somewhat parallel to his well-known stage theory of economic growth.

Smith explored in some depth the particular imitative powers of the various arts so as to discover why people patronize one art form over another. He found music limited in some respects but still more evocative than painting. “This power of exciting and varying the different moods and dispositions of the mind, which instrumental Music really possesses to a very considerable degree, has been the principal source of its reputation for those great imitative powers which have been ascribed to it” (p. 198). Music simply did not have the dimensions to imitate in any depth. “Instrumental Music, however, without violating too much its own melody and harmony, can imitate but imperfectly the sounds of natural objects, of which the greater part have neither melody nor harmony” (p. 200). Indeed, Smith seemed to wonder whether the metaphor of artistic creation as a process of imitation broke down in the case of music. “That music seldom means to tell any particular story, or to imitate any particular event, or in general to suggest any particular object, distinct from that combination of sounds of which itself is composed . . . The subject of a composition of instrumental Music is a part of that composition: the subject of a poem or picture is no part of either” (p. 205).

Smith’s observations about the fine arts were less systematic and confident than his pioneering work on the general principles of market behavior. But he was thoroughly intrigued by them, and like Hume and Turgot he was certain that they were a legitimate subject for social inquiry.

#### **4. Classical economics: The shadow of Bentham**

Considering the rich and suggestive menu of topics in the economics of art and culture left by Turgot and Hume and especially by Adam Smith, it is surprising how little was done with the subject by political economists in the nineteenth century. J.R. McCulloch’s comprehensive review of *The Literature of Political Economy* in 1845 mentioned the arts and culture not at all [McCulloch ([1845] 1938)]. The explanation for this silence may lie at two levels: first, the context of social relations and policy controversies of the time and second, the methodological issues that were arising in the emerging discipline of political economy.

In the nineteenth century many areas of inquiry evolved from being the undisciplined playground of amateurs, aristocrats, clerics and others to become rigorous sciences with accepted rules of procedure, authorities, professional journals and societies, and a

growing presence in university faculties and curricula. This was true of geology, anthropology, and the biological sciences, to mention only a few. Political economy also joined in the search for scientific status, a search that suggested amongst other requirements the need for simple and generalizable models of explanation and prediction. Exceptions had to be kept to a minimum to keep a paradigm viable, and when exceptions were proposed every effort had to be made to keep them inside the disciplinary tent. In the moral or social sciences, of which political economy was perhaps the strongest exemplar, the search for universal “laws” of behavior for the units under study led naturally to the postulation of optimizing and self-interested human beings. Any suggested exceptions to this postulate were in danger of threatening the scientific claims of the discipline. The speculations about the unique nature of aesthetic experience by the Enlightenment thinkers were exactly the sort that if accepted might throw into question the scientific legitimacy of political economy.

The threat perceived by the political economists to the scientific claims of their discipline from the Enlightenment approach to culture and the arts can be seen clearly in the writings of Jeremy Bentham, and especially in his *The Rationale of Reward*, published first in French, and in English in 1825. From his first reading of the *Wealth of Nations* Bentham argued for limiting the exceptions to Smith’s powerful market model. He claimed that except for a few very limited cases like patents and copyrights the state should avoid arranging for anyone to receive more than what the free market provided; the correct rationale for reward in almost all circumstances should be that free markets provided it automatically. The world was full of false claimants who argued for special treatment from government over and above what the market would provide. They should be resisted, argued Bentham, even in those special circumstances recommended by Smith such as the colonies and the capital markets. He observed that the arts contained an especially vocal collection of special pleaders who argued that they were distinctly different from other claimants and deserved favoritism. Bentham reported that for the most part he was not persuaded by these claims. He began by making a charge that was repeated often in the twentieth century, namely that public expenditures on the arts were usually regressive in their effects on the distribution of income and wealth. The rich despoiled the poor so that they could indulge their extravagant tastes. He condemned “Expenditure, of money, on articles, for the accommodation or amusement of the comparatively *opulent few*, at the expense of all, including, in prodigiously greater number, the *unopulent many*, who are incapable of participating in the benefit: productions of the *fine arts*, for instance . . .” [Bentham (1962, II, p. 251)].

Poetry was a frequent and symbolic target for Bentham because a case was regularly made for its public support and because it appeared to him to be a quintessential case of an art form that should be able to stimulate its own adequate reward from the joy it was supposed to give to the poet and the patron. “There are some countries in which the relish for literature is confined to such small numbers, that it may, upon the whole, be beneficial to encourage it by factitious rewards. But if we consider how intense are the enjoyments of the man born with poetic talents, the sudden reputation which they produce, and the ample profit they often yield, especially in the dramatic line, it will

be found that the natural rewards attached to them are far from being inconsiderable” (II, pp. 212–213). Bentham juxtaposed the fanciful structure of poetry to the hard facts and essential usefulness of science. “Happiness depends upon the correctness of the facts with which our mind is furnished, and the rectitude of our judgment; but poetry has no very direct tendency to produce either correctness of knowledge or rectitude of judgment. For one instance in which it [poetry] has been employed to combat mischievous prejudices, a thousand might be cited in which they have been fostered and propagated by it” (II, p. 213). He was incensed that the various forms of creative literature, but especially poetry, had appropriated the “imagination” as their point of origin, as if other intellectual endeavors did not have imaginative roots as well. “These, too, are all together placed under the head of *imagination*; as if, in the first place, to the exercise of all *these* branches of art, the exercise of the imaginative faculty were necessary; and as if, in the next place, it were not so to any of the *others*” [Bentham (1962, VIII, p. 76)].

Bentham took pains to point out often that the fine arts were simply “amusement”, by contrast with “the arts and sciences of curiosity” that were serious intellectual endeavor. “By arts and sciences of amusement, I mean those which are ordinarily called the *fine arts*; such as music, poetry, painting, sculpture, architecture, ornamental gardening, &c. &c . . .” (II, p. 253). Both amusement and the satisfaction of curiosity did give pleasure and therefore yielded utility to those who engaged in them, and their worth could be assessed from the additions to utility they contributed in relation to the utility that might be received from all other alternatives. In perhaps the most oft-cited and influential words in the history of cultural economics, Bentham expressed surprise that humans were, in fact, inclined so often to prefer the titillation of the fine arts to the truth that emerged from science and the innocent distractions provided by idle amusement. Certainly he could see no case for encouraging one over the other.

The utility of all these arts and sciences – I speak both of those of amusement and curiosity – the value which they possess, is exactly in proportion to the pleasure they yield. Every other species of pre-eminence which may be attempted to be established along them is altogether fanciful. Prejudice apart, the game of push-pin is of equal value with the arts and sciences of music and poetry. If the game of push-pin furnish more pleasure, it is more valuable than either. Everybody can play at push-pin: poetry and music are relished only by a few. The game of push-pin is always innocent: it were well could the same be always asserted of poetry. Indeed, between poetry and truth there is a natural opposition: false morals, fictitious nature. The poet always stands in need of something false. When he pretends to lay his foundations in truth, the ornaments of his superstructure are fictions; his business consists in stimulating our passions, and exciting our prejudices. Truth, exactitude of every kind, is fatal to poetry. The poet must see everything through coloured media, and strive to make everyone else to do the same. It is true, there have been noble spirits, to whom poetry and philosophy have been equally indebted; but these exceptions do not counteract the mischiefs which have resulted from this magic art. If poetry and music deserve to be preferred before a game of

push-pin, it must be because they are calculated to gratify those individuals who are most difficult to be pleased (II, pp. 253–254).

Bentham conceded that sometimes the arts had some indirect moral utility (what we might today call a positive externality) when they occupied what otherwise could be idle hands. But this was faint praise indeed. “All the arts and sciences, without exception, inasmuch as they constitute innocent employments, at least of time, possess a species of moral utility, neither the less real or important because it is frequently unobserved. They compete with, and occupy the place of those mischievous and dangerous passions and employments, to which want of occupation and ennui give birth. They are excellent substitutes for drunkenness, slander, and the love of gaming” (II, p. 254). He went on to suggest that it might even be claimed that credit should be given to the fine arts for an increase of pacifism. “It is to the cultivation of the arts and sciences, that we must in great measure ascribe the existence of that party which is now opposed to war: it has received its birth amid the occupations and pleasures furnished by the fine arts. These arts, so to speak, have enrolled under their peaceful banners that army of idlers which would have otherwise possessed no amusement but in the hazardous and bloody game of war” (II, p. 254). But surely this too was no justification for serious attention to the arts. Bentham had thrown down the gauntlet to those interested in the arts and culture to demonstrate that they were more than trivial and that they deserved the scarce time of those concerned with the analysis of public policy.

Bentham emphasized that as far as he could see “quality” in the arts was purely a matter of individual taste. Since the fine arts were all about amusement, how could anyone say – let alone dictate – what would amuse someone else and what would not. No one could state with any authority what is good art and what is bad. “It is only from custom and prejudice that, in matters of taste, we speak of false and true. There is no taste which deserves the epithet *good*, unless it be the taste for such employments which, to the pleasure actually produced by them, conjoin some contingent or future utility: there is no taste which deserves to be characterized as bad, unless it be a taste for some occupation which has a mischievous tendency” (II, p. 254). Those who called themselves art critics “under pretence of purifying the public taste” usually reduced the sum of happiness by depriving “mankind of a larger or smaller part of the sources of their amusement”. By declaring some works of art to be inherently inferior to the alternatives and removing them from consideration, critics effectively reduced the choice set of consumers. “These modest judges of elegance and taste consider themselves as benefactors to the human race, whilst they are really only the interrupters of their pleasure – a sort of importunate hosts, who place themselves at the table to diminish, by their pretended delicacy, the appetite of their guests” (II, p. 254). Bentham spelled out clearly the policy implications of his doctrine. No one should declaim to others on matters of taste, nor should the state favor any of “the arts and sciences of amusement and curiosity”, one over any other. Rewards might occasionally be offered for the discovery of new knowledge in the sciences that promised direct public benefit, but in the fine arts the market should be counted upon to issue most of the signals necessary to induce appropriate levels of production. “Among rich and prosperous nations, it is not necessary that the public should be at the

expense of cultivating the arts and sciences of amusement and curiosity. Individuals will always bestow upon these that portion of reward which is proportioned to the pleasure they bestow" (II, p. 255).

The contextual circumstances that in conjunction with the methodological concerns expressed in the works of Bentham help to explain the neglect of culture and the arts in the evolving body of classical political economy were of two kinds. The first revolved around the distinction made explicit by Adam Smith between productive and unproductive labor, and the second around a similar distinction made by the Physiocrats between goods that emerged from the "productive" agricultural sector and the goods and services that came from the unproductive urban sector. Under both of these distinctions the arts suffered by comparison. They had their roots in the prejudices of the seventeenth and eighteenth centuries against certain kinds of economic activity, especially those deemed luxurious or vicious, as discussed above. But later they were intended to focus attention on sectors that might or might not yield a surplus of physical products such as buildings or grain that could be added to the total fixed or circulating capital (wages fund) and thereby increase the output of the nation in later periods. Smith defined "productive labour" quite carefully as "the sort of labour which adds to the value of the subject upon which it is bestowed" [Smith ([1776] 1976, p. 330)]. The manufacturing laborer was the most obvious productive laborer for Smith, whereas for the Physiocrats it was the farmer. Recall that the stereotypical unproductive laborer for both Smith and the Physiocrats was the menial servant. Smith included the arts in this category, a depiction clearly intended mainly as an aid in classification, but one that was not complimentary to the arts. Smith had certainly conceded that unproductive labor and its output had value, but the denigration of them in the language used was undeniable. In the circumstances of the times the argument that the arts could not contribute to economic growth was damning indeed. The advice was not lost on his readers in the evolving classical political economy that they would be well-advised to direct their attention and study away from the arts to "productive" activities such as manufacturing, agriculture, and international trade.

Nevertheless, the distinction between productive and unproductive labor did attract some powerful criticism soon after it was formulated by Smith. For example, James Maitland (Lord Lauderdale) pointed out in 1804 what he took to be some of the absurdities that flowed from the distinctions proposed by the Physiocrats: "it is impossible to subscribe to the opinion, that the labour of the manufacturer and the artist are totally unproductive of wealth" [Lauderdale ([1819] 1962, p. 141)]. He also took issue with the application to art markets of the Smithian theory that under competition price would gravitate naturally toward the cost of production, which in the case of the arts would mean the subsistence wage of the artist: "it is impossible to believe that a painter, whose works have sold for thousands of pounds, and the value of which has been known to have increased for a century after his death, added nothing more to the value of the canvas than the value of his sustenance and an equivalent for the expence of his education" ([1819] 1962, p. 142). Lauderdale argued that in the arts unlike most other occupations talent imposed barriers to entry and that prices of works of art over time

and under competition would reflect increasing rents to artists of rare talent. He found Smith's distinction of unproductive from productive labor just as unsatisfactory as that of the Physiocrats. "It appears, therefore, impossible to contend, that the labour of the manufacturer and artist, or even the labour of that class whose services perish at the moment, are not, as well as that of the husbandman, to be considered as productive of wealth" (p. 153). One wonders if perhaps Lauderdale's place in the aristocracy, where in the early nineteenth century the arts were valued more than in other strata of society, caused him to react so critically to the treatment of the arts and culture in the growing literature of political economy.

Despite the cogency of the criticism of it, the distinction between productive and unproductive labor did not die easily. It was still alive and well in John Stuart Mill's *Principles* in 1848. Mill denied that "unproductive" should in any way be interpreted as a "term of disparagement" or as casting a "stigma" on certain goods and services [Mill ([1909] 1964, p. 44)]. Yet his disclaimer is contradicted by the tone of the examples he uses. In the case of the theater he seems to suggest that it would be better if people stayed home and saved their money; the only net gain to a nation from expenditures on performances that he could see arose if they were held abroad and there were foreign remittances:

... but what is gained by an actor is a mere transfer from the spectator's funds to his, leaving no article of wealth for the spectator's indemnification. Thus the community collectively gains nothing by the actor's labour; and it loses, of his receipts, all that portion which he consumes, retaining only that which he lays by. A community, however, may add to its wealth by unproductive labour, at the expense of other communities, as an individual may at the expense of other individuals. The gains of Italian opera singers, German governesses, French ballet dancers, &c are a source of wealth, as far as they go, to their respective countries, if they return thither (p. 50).

It is no coincidence that in a body of economic theory that emphasized cost of production as the most powerful determinant of value, in attempting to understand the economics of the arts as of other segments of the economy the focus should be on the labor that produced the artworks rather than the artworks themselves.

This is not to suggest that in all of nineteenth century classical political economy there was no commentary at all on intriguing questions in the economics of the arts; rather it is to indicate that comments were few and that the negative heuristics conveyed to scholars about study of these topics were substantially stronger than the positive ones. David Ricardo had established early on that the arts provided some of the best examples of those exceptional goods that although produced with labor could not reliably be reproduced. Their value as a result was demand-determined. His oft-repeated discussion in *Principles of Political Economy and Taxation* ([1817] 1948), in which artworks are grouped for analysis with various other extravagances, ran as follows:

There are some commodities, the value of which is determined by their scarcity alone. No labour can increase the quantity of such goods, and therefore their value



cannot be lowered by an increased supply. Some rare statues and pictures, scarce books and coins, wines of a peculiar quality, which can be made only from grapes grown on a particular soil, of which there is a very limited quantity, are all of this description. Their value is wholly independent of the quantity of labour originally necessary to produce them, and varies with the varying wealth and inclinations of those who are desirous to possess them.

These commodities, however, form a very small part of the mass of commodities daily exchanged in the market [Ricardo ([1817] 1948, p. 6)].

John Rae, ever the speculative thinker, struggled in his *Statement of Some New Principles of Political Economy* ([1834] 1965) with the question of why people in fact demanded products of the arts. He was unhappy with the implicit suggestion that interest in the arts was just like any other consumer demand, and a luxurious one at that. He suggested that

There is a pleasure in the sight of certain shapes and colors, and arrangements of them, which is quite independent of their cost; there is a fitness also; in the texture of certain fabrics, to preserve from the extremes of heat and cold, to add to the beauties of feature or form, and to correct their defects, that, of itself, gives pleasure; there are pleasures too which the mind creates to itself, out of the associations of these . . . The sight of statues, paintings, flowers, is also capable of affording a high degree of gratification to many minds. The degree of pleasure thus experienced is different in different individuals, and it is scarcely possible to ascertain what its exact amount is in any one; hence the difficulty in most cases, of determining what is, or is not, luxury [Rae ([1834] 1965, p. 272)].

Finally in this overview of classical political economy in the nineteenth century, we can refer to John Stuart Mill. Mill's positions on social and economic policies regarding the arts are reflected in his social theory. He was concerned about how practitioners of the arts could survive in a modern market economy, and especially one like that of Britain where there had been very limited respect for the arts. He looked with special concern at the condition of creative writers: "although the highest pecuniary prizes of successful authorship are incomparably greater than at any former period, yet on any rational calculation of the chances, in the existing competition, scarcely any writer can hope to gain a living by books, and to do so by magazines and reviews becomes daily more difficult. It is only the more troublesome and disagreeable kinds of literary labour, and those which confer no personal celebrity, such as most of those connected with newspapers, or with the smaller periodicals, on which an educated person can now rely for subsistence" [Mill ([1909] 1964, p. 397)]. He urged that more attention be given to this problem of sustaining the arts where a tradition was lacking and times were hard. He observed that in the sale of many artworks by fashionable artists there was substantial producer rent, and like many after him he wondered if this could somehow be extracted to support artists who were not so fortunate [Mill ([1909] 1964, p. 443; 1986, p. 1240)]. Using a variety of arguments, he advocated public education in the arts at all levels. Above all the quality of life of the people would be improved thereby. "The races and



nations whose senses are naturally finer and their sensuous perceptions more exercised than ours, receive the same kind of impressions from painting and sculpture: and many of the more delicately organized among ourselves do the same. All the arts of expression tend to keep alive and in activity the feelings they express" [Mill (1984, p. 254)]. In addition cultivation of the arts would lead to a more moral and tolerant citizenry: "There is, besides, a natural affinity between goodness and the cultivation of the Beautiful, when it is real cultivation, and not a mere unguided instinct" (p. 255). And finally he argued that the arts could instill quality control in a people, help to demonstrate the possibility of sublime accomplishment and reveal "the Good made perfect". This was because "No other human productions come so near to perfection as works of pure art" (p. 255).

## 5. Humanist critics

While political economists in the nineteenth century shied away from a sympathetic approach to the arts and culture, others moved in to fill the void, in particular poets, novelists and essayists who came to constitute a community of criticism. The result was a widening gulf between the evolving discipline of political economy on the one hand and the humanities and the arts on the other. Some of the subjects addressed by the critics could be described today as positive externalities of the arts and culture; others involved more profound speculations about the place of the arts in national life. Many novelists and poets worried in particular that the capitalist market economy was squeezing aside the arts and culture in favor of more prosaic goals [Grampp (1973)]. For example, Thomas Love Peacock complained that the nineteenth century had become the age of the "huckster" (*l'épicier*) in which all goods were produced in response to consumer demand without regard to their inherent merit, in the "literary market" as elsewhere [Peacock (1926, pp. 294, 297)]. The "march of mechanics" [progress of technology!] had been accompanied by "the days of political economy", and the result was "what we call *l'esprit épicier*. Applied to literature, to the arts, to the mode of living, and manifesting itself in manner, style, and taste, by something obsolete, vulgar, and awkward, tinged with the ridiculous, this spirit has created what we call *le genre épicier*" [Peacock (1926, pp. 295, 300)]. Despite all his talk of freedom, the huckster was essentially authoritarian. "His ruling passion is the love of 'order', because he has observed that in the days of political disturbances there has been a fall of a per-centage on his operations. The apprehension of anarchy, or, to speak more correctly, the fear of diminished sales and falling prices, has made him a fanatic of 'l'ordre public' . . . For him, order is a positive result which must be obtained at any price, without regard to the causes which may have produced a feverish over-excitement in any portion of society" [Peacock (1926, pp. 300–301)]. Because of his fear of disorder the huckster was the enemy of the artist and the intellectual, indeed it "had never happened to him to have an idea in the course of his life" (p. 301).

Three humanist critics of political economy went beyond such denunciations to make notable contributions to the relationship of the arts to the economy: the poet and essayist Matthew Arnold, the prominent artist, critic, and art historian John Ruskin, and the leader of the Arts and Craft movement, William Morris. We turn to them in the following sections.

### 5.1. Matthew Arnold

The main question raised by Matthew Arnold was a direct challenge to the political economists: Were the arts and culture more than merely a distinctive collection of goods amongst which consumers might make a selection for their amusement? Were they not in fact essential to healthy and successful social development? Arnold addressed squarely what was becoming one of the most frightening questions in nineteenth century political economy: Could economic development continue without causing destructive conflict between owners of capital on one side and labor on the other? Poets and novelists before him had often addressed environmental degradation and sufferings of the working class caused by the new factory system;<sup>1</sup> Arnold asked the follow-up question of how society could be held together where these sufferings persisted while the franchise was increasing steadily through the reform acts of 1832 and 1867. What was to stop those without property from taking it away from those who had it, either legally through Parliamentary action or illegally through violence in the streets? What was to stop democratization from degenerating into anarchy? After all, the French Revolution was only a few decades behind and the revolutionary year of 1848 even more recent. The answer to the threat of anarchy, Arnold suggested, lay in civilizing the population and more specifically in extending “culture” throughout the populace, defined in his influential book *Culture and Anarchy* ([1869] 1903) in Enlightenment terms as the “study of perfection”, “turning a stream of fresh and free thought upon our stock notions and habits” (pp. xi, 7). He argued that culture is “possessed by the scientific passion as well as by the passion of doing good” (p. 8). It would lead human nature toward “a harmonious perfection, a perfection in which the characters of beauty and intelligence are both present”, “a human nature complete on all its sides” (pp. 19–20, 25). Arnold’s definition of culture seems rather close to what is meant today by the results of a “liberal education”. Culture would not come painlessly or without hard work and cost. Contrary to Bentham there must be expert critics who would endeavor to learn and propagate “the best which has been thought and said in the world” (p. xi). They must guide others in their quest for culture.

The type of beneficent cultured citizen that Arnold claimed was essential for a healthy, advancing society was in his view in sharp contrast to the grasping self-interested creatures modeled by the economists, and to whom he gave a scornful name

<sup>1</sup> Creative writers in Great Britain in the nineteenth century who deplored the industrialization that was taking place and the evolving science of political economy that they associated with it included Robert Southey, Samuel Taylor Coleridge, William Wordsworth, Thomas Love Peacock, Anthony Trollope, William Cobbett, Thomas Carlyle and, most notably, Charles Dickens; see Grampp (1973).

that was to remain with them, Philistines. “The people who believe most that our greatness and welfare are proved by our being very rich, and who most give their lives and thoughts to becoming rich, are just the very people whom we call Philistines” (p. 17). He described two bundles of behavioral pressures operating within humans that, following Heine, he called Hebraism and Hellenism. Hebraism included conscience, obedience, duty, and subjugation of self, some of the qualities treated by Adam Smith more than a century before in *The Theory of Moral Sentiments*; Hellenism involved spontaneity, curiosity, intellectual flexibility, and appreciation of beauty, subjects dealt with in *The Wealth of Nations*. “The uppermost idea with Hellenism is to see things as they really are; the uppermost idea with Hebraism is conduct and obedience” (p. 123). Arnold saw human history as mainly a struggle between these two sets of forces: “by alternations of Hebraism and Hellenism, of a man’s intellectual and moral impulses, of the effort to see things as they really are, and the effort to win peace by self-conquest, the human spirit proceeds; and each of these two forces has its appointed hours of culmination and seasons of rule” (p. 134). A healthy individual or society, Arnold insisted, maintained a proper balance between these two bundles of behavioral characteristics. Bennett (2005) sees the ideas of Plato and of the Romantic poets Goethe and Schiller behind Arnold’s conception of culture. It is not hard for the economist to see as well the economic thinkers of the Enlightenment, all of whom had been concerned about what should be the context in which a society, economy and polity could operate smoothly. The political economists of Arnold’s day had begun to believe that the economy could reasonably be explored as separate from the society and polity, and that if the natural proclivities of humans were left untrammelled, optimal social and political as well as economic results may well emerge from free markets. Arnold seemed to be following up some of the hints we find in the early Smith that the recipe for economic efficiency may not be that simple. Moreover, there may be a case for concluding that markets for art and culture are truly exceptional, both in how they operate and in their effects on society overall. Contextual preconditions are necessary for the smooth performance of a market economy, and these could require institutions and actions remote from the conventional concern with technical efficiency such as the cultivation of the arts and culture through a liberal education. This was a message that not many nineteenth century political economists wished to hear, or with which they were prepared to engage.

## 5.2. John Ruskin

Another message about the relations of the economy with culture and the arts that was profoundly disagreeable to nineteenth century political economists came from John Ruskin. Whereas Arnold had questioned one emerging sacred principle of classical political economy, that it was reasonable to focus on market interactions without attention to the cultural context, Ruskin disputed another, that utility functions of consumers are stable for the most part and that wants should be treated as exogenous to economic analysis. Arnold’s concern was essentially macro, about the stability of the entire social and economic system, whereas Ruskin’s was micro, about the happiness of the individ-

ual in society. Like Arnold, Ruskin was insulting to the political economists of his day seemingly by design. Although he published works with titles like “The Political Economy of Art” [1857; published in Ruskin ([1880] 1905)] he took pleasure in reporting that he had read not a word of political economy, except for *The Wealth of Nations* in his youth. Like Arnold he also constructed his own vocabulary when needed and made no effort to connect with political economists who might have had an interest in his subject. Also like Arnold, Ruskin was happy to reject laissez faire as a policy principle. Indeed, he looked to an authoritarian and interventionist leader to make society a better place. Ruskin’s policy norm was formation of the happy individual who had “perfected the functions of his own life to the utmost” [Ruskin ([1862] 1905, pp. 84, 105); see also Moore (2005)], but he proposed to make this individual happier not by piling more and more goods and services upon him but by training him to make a better selection among alternatives and to appreciate more those goods that he could afford. Ruskin believed that by this approach political economy truly became a moral subject, and the science–ethics dichotomy in political economy that Mill had pointed out no longer existed. For Ruskin the notion that value depended upon cost of production was absurd, and he anticipated the marginal utility revolution of the 1870s by insisting that wealth was relevant only to the way in which goods were used. He insisted that “wise consumption is a far more difficult art than wise production” ([1862] 1905, pp. 98, 104). The natural conclusion Ruskin drew from his theorizing was that the community had a responsibility to make sure of the “life availing” qualities of the goods that were consumed and to educate consumers to make the best use of them. The artist and the art critic both had important roles in this process, and the market could not be counted upon to guarantee the maintenance of the artist and critic. He was an early and energetic advocate of public support of arts education, of distinguished public buildings, of museums (to protect old art), and of a national arts purchase fund (to provide demand for new art). He returned to and agreed with Adam Smith’s theme that vanity, if not constrained, could become the mistaken determinant of what was produced in the arts. The omnipresence of bad taste and ostentation were all the proof Ruskin needed of the necessity of public intervention in the arts on a grand scale.

Even though Ruskin’s writings on the economics of art were denounced, deplored, or ignored by most of the economics profession in his own time and thereafter, they lived on in the thinking and literature of the labor movement in Britain and America. To give only one example, the popular economics textbook for workingmen by Henry Clay, *Economics: An Introduction for the General Reader*<sup>2</sup> began with this observation: “No study of Economics, therefore, it seems to me, is worth making which does not include some consideration of the relation of the economic organization to political and ethical aims and standards; which does not, in other words, indicate what light Economics can throw on Ruskin’s question, ‘What is wealth?’” [Clay (1918, p. viii)]. In his final chapter on “Wealth and Welfare” Clay turned again to the question of what constituted

<sup>2</sup> First edition 1918 published and republished many times both in England and the United States.

human progress and he noted the paradox that times when great art was produced in abundance were not necessarily years when consumer goods were increasing:

If we compare different ages and countries, we are struck by the apparent unimportance of wealth. The materialist regards the vulgar plenty of the twentieth century as greatness; others will regard its art as a better index of an age or a country's temper. Whichever is right, the two great ages of art, the two periods when craftsmen were artists and the appreciation of art was general, were ages of extreme material poverty; and few will assert that the age of Arkwright was greater than the age of Pheidias, the civilization of Chicago than the civilization of Athens; few, who have studied both and compare achievement with opportunity, will place the art of the thirteenth century below the art of the Victorian age, the age that built the cathedrals below the age that restored them. As society has grown richer, art has become more and more the concern of little cliques and coteries, less and less a part of the everyday life of ordinary people, until to-day we have countries like the United States, so rich that its Whistlers and Sargents fly to the poorer countries of Europe (p. 420).

Clay left this puzzle for his working-class readers to ponder; clearly he thought it deserved more reflection.

### 5.3. *William Morris*

The third writer on the arts who was anathema to respectable nineteenth century political economists was William Morris [Upchurch (2005)]. His sin was partly that he was a seeming turncoat. After all, Arnold and Ruskin might be dismissed as just an ethereal poet and a half-crazed art critic; Morris was by contrast a successful businessman who made a fortune from the applied arts. He was also himself a talented artist and designer, and much of his celebrity came from his pioneering leadership of the early Arts and Crafts movement. Perhaps because of his deeper immersion in the realities of art markets Morris was more gloomy than Arnold and Ruskin about the prospects for change from piecemeal reforms. Both Arnold and Ruskin envisaged the possibility of much beneficial reform through education and a rather modest intervention by the state to support and shape the arts. Morris had grave doubts. He believed that the arts could prosper only under conditions of collective ownership of the means of production, either through conversion to a socialist state or through a communist revolution. He was not precise in his picture of why the arts would necessarily flourish outside the free market, but his assertion that they would was another fundamental challenge to the economic orthodoxy of his time. It is noteworthy that Morris had more to say about the potentially positive effects of socialism on the arts than did Marx himself or most of his disciples [Solomon (1979)].

## 6. The marginal revolution

### 6.1. *The marginalists' treatment of arts and culture*

The marginal revolution of the 1870s was about the construction of general laws in economics, and about the Benthamite urge to postulate simple and universal principles of human behavior, rooted in utility calculations, that could be used to model the economy. So where in this new vision of economics stood the arts and culture? Not very prominently! Special attention to the arts and culture, as to other particular segments of the economy, did not fit well with the philosophy behind the new marginal economics. The search for exceptionalism in the arts, as in other areas, came into direct conflict with the search for generality.

Several observations may be made about the treatment of the arts and culture during the early years of neo-classical marginal economics. First, the marginalists made far more references to the arts than did their classical predecessors, but less as interesting and exceptional special cases than as manifestations of the general case. Favorite illustrations from the arts used in presentations of economic theory stressed the unique and non-reproducible nature of many artworks. Examination of these cases demonstrated that the price of all consumption goods was demand rather than cost determined. For example, Carl Menger wrote: "A large number of goods cannot be reproduced (antiques, and paintings by old masters, for instance) and thus, in a number of cases, we can observe value but no possibility of reproduction. For this reason, any factor connected with reproduction cannot be the determining principle of value in general" [Menger ([1871] 1950, p. 147)].

Second, there was greater willingness among the marginalists than among the classical economists to posit the possibility of positive externalities emanating from the arts. A moderate position on the salutary effects of the arts on social relations – the Arnold assertion – came to be seen as quite reasonable. Characteristically Alfred Marshall wrote to a correspondent in 1900: "the growth of towns makes it doubly urgent to supply wholesome thoughts & suggestions, lest unwholesome should prevail; and to turn music & painting & other fine arts to account in filling the void in man's life caused by the want of the free light & freshness & beauties of nature" [Marshall (1996, II, pp. 270–271)]. Third, there appeared more prominent economists with deep personal commitment to the arts who were prepared to testify to positive externalities from their own experience. Finally, over time the marginalists lost many of their inhibitions about commenting on the arts; some of their observations were prescient, others were absurd. None of the main neo-classical texts had a separate chapter, or even part of a chapter, devoted to the economics of culture and the arts. The references to the subject were widely scattered throughout. A closer look at the writings of two of the most prominent revolutionaries, Alfred Marshall and William Stanley Jevons, may be illuminating.

## 6.2. *Alfred Marshall*

In a taxonomy of wealth taken from F.B.W. von Hermann, Marshall claimed that all goods related to the arts may be placed in various categories. Conventional artworks are “material”, “personal” and “internal”; “the faculty of deriving recreation from reading or music” is “internal” but “non-material”. “The laws which govern the birth of genius are inscrutable” except that more genius per capita comes from the higher orders than from the working class [Marshall ([1920] 1964, pp. 45–46, 176)]. Marshall argued that the demand for artworks has roots in some fundamental set of preferences that will ultimately prevail over the pressure of fashion. “For though the crudest and most ridiculous fashions in art and in literature will be accepted by the people for a time at the bidding of their social superiors, nothing but true artistic excellence has enabled a ballad or a melody, a style of dress or a pattern of furniture to retain its popularity among a whole nation for many generations together . . . traditional instincts played a great part in preserving the purity of the industrial arts in Oriental countries, and to a less extent in medieval Europe” ([1920] 1964, p. 177). Education in the visual arts as distinct from the rest of the arts, Marshall thought, was desirable for the progress of the economy, but less so than education in other more practical areas. He suggested that education in art stood on a somewhat different footing from education in hard thinking: for while the latter nearly always strengthens the character, the former often fails to do this. Nevertheless he believed that the development of the artistic faculties of people was in itself an aim of the very highest importance, and was becoming a chief factor of industrial efficiency: “We are here concerned almost exclusively with those branches of art which appeal to the eye. For though literature and music contribute as much and more to the fulness of life, yet their development does not directly affect, and does not depend upon, the methods of business, the processes of manufacture and the skill of the artisans” ([1920] 1964, p. 177).

## 6.3. *William Stanley Jevons*

With Jevons we see the appearance among economists of a new phenomenon: the closet artist or aesthete, the economist who either has a separate life as an artist or acknowledges having had a profound aesthetic experience while seeming unwilling to incorporate the artistic dimensions of life into economics. For Jevons music was the fine art that mattered most. But he saw all the fine arts as capable of enriching the lives of the working class beyond their own expectations and therefore yielding positive externalities. He envisaged the arts as being initially unlikely to interest the masses and needing to be administered rather like a dose of tonic. The arts imposed on unsophisticated people in this way would, Jevons believed, ultimately have the aesthetic impact that he himself had experienced. But this aesthetic experience could not be anticipated by those who had not already experienced it, and therefore the arts would not sell well in the market to begin with and would have to be delivered for a while below market rates.



An unpublished manuscript by Jevons entitled “On the Functions of Music” makes it clear that he himself had had aesthetic experiences that he tried to put into words. He wrote that his own response to music was similar to “the contemplation of subjects of Interest, Beauty or Sublimity, and consists of a gentle excitement, an engrossment of the thoughts by agreeable subjects and a general removal of the mind from its ordinary course of duties and frailties, and its continual mixture of slight pleasures and pains” [Jevons quoted in Mosselmans and Mathijs (1999, p. 151)]. Jevons evidently felt that his reaction to music was rather like intoxication or a drug-induced euphoria, suggesting that the experience might be more like an addiction than a conventional act of consumption:

When rising to an unusual pitch, the feeling of which I speak becomes an intense delight; it absorbs the attention completely, and causing it to forget ordinary affairs and thoughts, elevates it to a region of pleasurable sensations nowhere else discovered. It is indefinite however, leads to no conclusion, suggests it may almost be said no leading or strong thought or purpose and even of one’s future path in life, or unavoidable duties happen to occur to the mind in the midst of this sort of intoxication, they appear smoother than ever before, the difficulties have entirely vanished and oneself feels such a sort of confident moral strength, as will only too soon be found to disappear when this state of feeling has ended and the petty or great difficulties of life are once again visible in all their reality (pp. 151–152).

Indeed, it seems that Jevons experienced something like guilt after a musical “trip” was over. “Who that has attended a concert, play or other public performance, has not experienced this sudden and almost disgusting revulsion of feelings at its termination when the music has finally ceased & no longer rivets the attention, and a dreamy progress home, and a renewed circle of duties and the sorrow are all that seem to await one” (p. 152). He was unwilling to probe deeply into what had taken place during his aesthetic adventures and he agreed with Bentham and some other early economists that the aesthetic experience could not be improved by the application of reason or through the services of scholarly critics:

To investigate its cause would be to enter one of the most difficult and least certain of metaphysical subjects and be quite beyond our purposes. It will suffice to say that it does not depend in any particular degree on knowledge . . . the essential nature of Poetry it is universally acknowledged to be almost impossible to define, and lastly the sense of beauty or of melancholy arising from a simple succession of sounds, that is to say an air or melody, is equally inexplicable (p. 152).

Had Jevons lived longer and continued to pursue his interests in the arts he might be remembered today as the father of cultural economics. He certainly made some promising beginnings on topics that would concern his successors thereafter. In 1878 he followed up his interest in the external consequences of “the amusements of the people” and concluded that one of the most promising “methods of social reform” would be encouragement to “good moral public amusement, especially musical entertainments”.



He distinguished very firmly between serious music and what he described as “the mixture of inane songs, of senseless burlesques, and of sensational acrobatic tricks, which make the staple of a music-hall entertainment” [Jevons (1904, pp. 2, 4)]. No advocacy of consumer sovereignty here! He knew what the working classes needed among the arts even if they did not know themselves. Jevons speculated that neglect of the arts for the masses might be the result of a conspiracy by the ruling classes who feared the consequences of a liberating culture. He wrote: “Now I believe that this want of culture greatly arises from the fact that the amusement of the masses, instead of being cultivated, and multiplied, and refined, have been frowned upon and condemned, and eventually suppressed, by a dominant aristocracy. Amusement has been regarded as in itself almost sinful, and at the best as a necessary evil. Accordingly, villages and towns have grown up in the more populous parts of the kingdom absolutely devoid of any provision whatever for recreation” (p. 6).

Following up on this conspiracy theory Jevons abandoned *laissez faire* policy as well as consumer sovereignty theory. He urged governmental attention to the arts as a high priority for public policy: “Among the means towards a higher civilisation, I unhesitatingly assert that the deliberate cultivation of public amusement is a principal one” (p. 7). To buttress his case Jevons claimed that the fine arts and culture would keep the working classes away from various temptations that might be personally and socially destructive, and would yield positive therapeutic effects.

What some seek at the cost of health, and life, and reputation, from alcohol, and from opium, that they might obtain innocuously from music, if they could cultivate true musical taste. Of course there is some nervous waste even in the enjoyment of music, and it is greater as the attention is more excited. Tedium must usually follow an entertainment of two or three hours; but so soon as tedium approaches, the attentive attitude of mind is destroyed, and the corresponding nervous waste ceases. The music, in short, holds the mind enchained just so long as there is energy of thought to spare; in the meantime the body remains in a perfect state of repose (p. 10).

Jevons went so far as to charge that the authorities in Britain, in contrast to those on the Continent, had responded to the demands of the aristocracy and had sabotaged public arts by tolerating within them criminals of every kind – especially pickpockets and prostitutes. He asked indignantly: “Why do we tolerate a state of things under which a young man cannot seek an hour’s recreation without meeting an evil magnet at every turn?” (p. 25).

Another public policy topic in the arts that interested Jevons was what he called “museum economy”, an area in which he found himself a pioneer. He endeavored to arrive at some conception of the purposes in creating public collections and of the means by which those purposes may be most readily attained. “Although the subject has hardly received any attention as yet, I believe it is possible to show on psychological or other scientific grounds that much which has been done in the formation of Museums is fundamentally mistaken” (p. 53). He was deeply critical of British museums and the theory

on which they were based. “There seems to be a prevalent idea that if the populace can only be got to walk about a great building filled with tall glass-cases, full of beautiful objects, especially when illuminated by the electric light, they will become civilised . . . The well-known fact that the attendance at Museums is greatest on wet days is very instructive” (p. 54). He complained that most museums were over-crowded and chaotic: “a great multitude of diverse things is not only useless but actually pernicious, because it tends to destroy that habit of concentration of attention, which is the first condition of mental acquisition . . . It does not seem, however, to have occurred to the creators and managers of Museums, that so far as education is aimed at, a certain unity of effect is essential” (p. 55). Although Jevons’ extended commentary on museums does not in the end consist of much more than shrewd insights and unsupported opinions, he was surely a pioneer in suggesting that these institutions be opened to scientific study.

For the most part the marginal revolutionaries based their economic theory around the general proposition that a competitive market system in which certain conditions hold will yield the maximum human welfare from a given endowment of resources and technology. In principle the determining wants on the demand side of product markets were the demanders’ own business. As Jevons put it “In the science of Economics we treat men not as they ought to be, but as they are” (1905, p. 38). In practice, however, the implementation of this principle was not so straightforward. These Victorian moralizers might agree that in theory pushpin equals poetry but they did not really believe it. They had their own prejudices about what goods were morally superior to others and they did not try to conceal their views. They spoke of a “natural” sequence in which consumption would (should?) take place over the course of economic growth, and this came to be treated as self-evident. Food, clothing, housing was the presumed sequence, and if a consumer announced that a glass of gin and a lottery ticket were high priority consumer goods to precede food and clothing, this would undoubtedly be deemed unacceptable. The vocabulary used to describe consumption often revealed the approved hierarchy of values. Menger referred to “needs” and “requirements” rather than to “wants”; he also distinguished between “true” and “imaginary” goods, the latter being especially common among the demands of people at lower “levels of civilization”. These included the following items of which he clearly did not approve, even though they were demanded freely in the market: “most cosmetics, all charms, the majority of medicines administered to the sick by peoples of early civilizations and by primitives even today, divining rods, love potions, etc. For all these things are incapable of actually satisfying the needs they are supposed to serve” [Menger ([1871] 1950, p. 53)]. He also disapproved of goods that were produced “when non-existent human needs are mistakenly assumed to exist”. In this category he included “medicines for diseases that do not actually exist, the implements, statues, buildings, etc., used by pagan people for the worship of idols, instruments of torture, and the like”. Menger concluded optimistically that “As a people attains higher levels of civilization, and as men penetrate more deeply into the true constitution of things and of their own nature, the number of true goods becomes constantly larger, and as can easily be understood, the number of imaginary goods becomes progressively smaller” (p. 53).

In the same vein Jevons talked of the “higher pleasures” which were “almost incomparable in power and authority” but noted somewhat reluctantly that, because of the current stage of economic development, “It is the lowest rank of feelings which we here treat” [Jevons ([1871] 1965, pp. 26–27)]. Jevons cited the authority of T.E. Banfield on the natural sequence of goods to be consumed: “the satisfaction of every lower want in the scale creates desire of a higher character. If the higher desire existed previous to the satisfaction of the primary want, it becomes more intense when the latter is removed . . . The highest grade in the scale of wants, that of pleasure derived from the beauties of nature and art, is usually confined to men who are exempted from all the lower privations. Thus the demand for, and the consumption of, objects of refined enjoyment has its lever in the facility with which the primary wants are satisfied” [Jevons ([1871] 1965, pp. 42–43)].

The significance of these comments by Menger and Jevons is to show that they tended to see art and culture as human activity that occurred only after sustained economic growth. In their day, they believed, art works were demanded by consumers who from good fortune had already reached high levels of income. Since such goods were few they could safely be ignored.

#### 6.4. *Lionel Robbins*

The last great neo-classical economist who attended to the economics of the arts before the appearance of the cultural economics sub-discipline in the 1960s was Lionel Robbins. Like Jevons, Robbins had a deep personal interest in and commitment to the arts; in his youth he had hoped to become a poet and he rejoiced in the visual arts, the opera and the ballet. He was one of the most energetic and successful advocates for the arts in twentieth century Great Britain; his efforts on behalf of the National Gallery, the Tate Gallery, Covent Garden and other institutions in the arts are legendary. Susan Howson, Robbins’ biographer, reports that over his career Robbins expounded the full range of neo-classical economic arguments for public support of the arts [Robbins (1963); Howson (2005)]. He explained that some art works at least may be seen as public goods and because of their particular character and the external benefits to society that they yield will be underproduced if left entirely to the private sector. He also favored free public access to the arts on grounds of equity. But he struggled with the eternally challenging question of whether the arts are truly exceptional – different in kind from other consumption goods. Clearly he thought they were, but he recognized that the economics he practiced could not help him very much in making the case. He concluded it was a matter of “ultimate values”; clearly the arts imparted “quality and meaning to life on this planet by reason of their mere existence” but to understand why and how they did so was probably beyond the scope of economic science. Robbins was confident that through public education and exposure to the arts the taste of the public could be formed and they could be led to levels of appreciation that would not otherwise be attained. Robbins supported wider use in Britain of encouragement to private support of the arts through tax concessions to private donors, and he made a case for emergency

public purchase of great works of art at times when important private collections were being dispersed.

## 7. The arts in American economics

If treatment of the arts as truly exceptional – as a component of the economy that must be analyzed separately and with its own conceptual apparatus – could be found anywhere in the late nineteenth and early twentieth centuries, surely it would be in the work of those who would later be called the American Institutionalists. Typically these economists were in open revolt against the marginal revolutionaries of the 1870s and against Benthamite utilitarianism as an explanation for economic behavior and as a way of measuring changes in human welfare. They objected by and large to the universalizing strategy of picturing economic agents on both sides of the market as maximizing well-understood quantities like satisfaction and profit, and minimizing pain and cost. They postulated a much more complex system in which human behavior and its objectives are heterogeneous and not easily understood. Accordingly they examined many different types of economic agent, from the captain of industry to the trade union leader, the farmer, and the housewife confined to the home. They attended to behavioral drives that ranged from the instinct of workmanship through the propensity for emulation and the parental bent. They studied particular segments of the economy that especially interested them: agriculture, higher education, and the family, to mention only three. But how about the arts and the market institutions that sustained the arts in a competitive economy? Were the arts really distinct from other segments of the economy? Was pushpin, indeed, different from poetry? What made the arts tick? Rather surprisingly on these questions the Institutionalists had little to say.

In Thorstein Veblen's discussion of the arts in his *Theory of the Leisure Class* ([1899] 1934) there was a move backward to the cynical and dismissive tone often heard in the commercial pamphlets of the seventeenth century. Veblen described works of art from the past as having been intended mainly to demonstrate the wealth and power of the patron, something that had been accomplished most effectively by making art works novel, wasteful, idiosyncratic and expensive. Aesthetic considerations were usually subsidiary or absent. "The canon of beauty requires expression of the generic. The 'novelty' due to the demands of conspicuous waste traverses this canon of beauty, in that it results in making the physiognomy of our objects of taste a congeries of idiosyncrasies; and the idiosyncrasies are, moreover, under the selective surveillance of the canon of expensiveness" [Veblen ([1899] 1934, p. 153)]. He grouped the arts and culture loosely within the large body of wasteful and unproductive activities pursued by those concerned mainly with wealth and power. As in Bentham, the arts were arrayed with games and sport, but now for different reasons. For Veblen pushpin equaled poetry not for Bentham's reason that they were both just items in a consumer's choice set, but because they were both intended mainly to demonstrate someone's conspicuous leisure, past or current. Demonstration of the capacity to waste time on either pushpin or poetry was

what both these activities were all about; neither contributed to improvement of human welfare in any real sense. “Such immaterial evidences of past leisure are quasi-scholarly or quasi-artistic accomplishments and a knowledge of processes and incidents which do not conduce directly to the furtherance of human life” (p. 45). In effect Veblen revived the notion put forth by Bodin and Mandeville that the arts are something with which the idle rich rather than humans in general are mainly involved, and must be, therefore, wasteful and without redeeming social value.

Some of the later writers in the Institutionalist tradition showed more regard for the arts, and they developed the implications of a rejection of the Benthamite calculus for the valuation of art. John Kenneth Galbraith had a lifelong interest in the arts and even wrote a book on *Indian Painting* during his time as ambassador in New Delhi [Galbraith and Randhawa (1968)]. He was much devoted also to “the building of bridges between economics and the arts” but he encountered many obstacles. In the 1960s he inaugurated a seminar at Harvard on the Economics of the Arts [Galbraith (1986, pp. 137, 144)]. He reported that his “artist friends saw it – to the extent that they saw it at all – as a rather philistine performance; nothing could so surely degrade art as an association with economics . . . My fellow economists saw my enterprise as essentially frivolous, possibly even eccentric” (p. 144).

Galbraith’s views on various topics in cultural economics were presented in a chapter entitled “The Market System and the Arts” in his book *Economics and the Public Purpose* (1973). He explained the inattention to the arts among economists as a result of the necessarily anachronistic character of most artistic production. Firms producing art works were operated still by independent entrepreneurs and not by managers bent on the division of labor. Indeed in the large corporations the word “artist” was used as a term of contempt for someone unwilling to submit to discipline. At the same time the inability of most large corporations to make use of the artist constructively was reflected in their typically poor standards of design. “The automobile industry, the mass producers of furniture, the household appliance industry, the container industry and numerous others amply illustrate the point” [Galbraith (1973, p. 63)]. Galbraith focused mainly on how economic aspects of the arts had become different from the norm in the modern economy. For a start, since the arts were typically anathema to the “technostructure” of the large firm, when they were found to have some commercial value they were often hived off into small independent companies. Second, modern art works, even though necessarily the result of creative change, were typically resisted by demanders who had been trained to be very conservative; most modern demanders had been conditioned by their culture to reject artistic change. They came to accept and to celebrate change in due course, but only after a struggle. Galbraith’s contributions to thinking on the economics of the arts never went beyond shrewd observations and provocative speculations of this kind, and he never proposed an agenda for research as did the Bloomsbury Group, discussed next.

## 8. Keynes and the Bloomsbury Group

More than any economist before or since, John Maynard Keynes spent much of his life close to, and at times even embedded within, a community of artists and writers. He experienced the arts first hand. The core of this community was the Bloomsbury Group, but his contacts and interests extended well beyond this collection of close friends. Keynes did not himself make significant contributions to the vigorous and original thinking within Bloomsbury on the place of the arts in society and the economy. But by word, and most importantly by deed, he demonstrated that he was fully aware of and in sympathy with it [Moggridge (2005)]. Why did he himself not take this body of thought and inject it into the economics of his time, where he was such a prominent figure? We can only speculate. It may have been because of his concentration on the more critical issues of the decades in which he lived: international conflict, depression, and the need to construct a new and viable world order. Or it may have been that he inferred the likely lack of receptivity to the Bloomsbury way of thinking within the economics discipline and could not confidently see a road ahead in that direction.

There were at least five topics on which Bloomsbury thinking might be categorized as relevant to cultural economics today:

- (i) the place of the arts in human life, including economic life;
- (ii) the character of the artistic experience, both for artists and for their audience;
- (iii) the use made of the arts by artists and others to shape behavior in society and the economy;
- (iv) the nature of the demand for and supply of art works in art markets; and
- (v) opportunities for experiments with policy change and institutional reform.

### 8.1. *The arts in human life*

The Bloomsburys denied vociferously the view in much of the economics literature to date that artists were simply the producers of superior goods that were consumed only when incomes rose, or that they produced merely luxuries for times when other more important wants had been filled. The Bloomsburys believed the exact opposite, i.e. that everyone in society of any class or economic station could and should be engaged in or at least exposed to the arts. In fact they were persuaded that modern society, being the result of economic progress, was more likely to stultify than to stimulate the artistic impulse in humans. The arts, they insisted, were an essential element and building block in the achievement of true human civilization; the arts were a cause not a condition or consequence of that peculiar and precious set of civilizing circumstances that only liberated humans can produce. Indeed, “the economic possibilities for our grandchildren” toward which Keynes looked ahead wistfully in 1930 consisted necessarily of “the arts of life as well as the activities of purpose” [Keynes ([1930] 1972, p. 332)].

The Bloomsburys were among the first to praise the extraordinary artistic accomplishments of non-western peoples, denying thereby the assertion that the arts could flourish only at high levels of income. One of the first illustrations in the celebrated

book *Civilization* by Roger Fry's friend and disciple Kenneth Clark (1969) is not of a Michelangelo or of a Da Vinci but of an African mask owned by Roger Fry [Goodwin (2005)]. Fry also held two exhibitions of the drawings of children – one of drawings by children of artists and the other of drawings by uninhibited inner-city children – at his Omega Workshops (1913–1919) to demonstrate that before they had extensive exposure to the modern economy the children's capacity for artistic expression was at its peak [Collins (1984, pp. 144–145)].

## 8.2. *The character of the artistic experience*

One of the contributions for which the Bloomsbury Group became best known was a “formalist” theory of aesthetics, expressed originally in Fry's “An Essay in aesthetics” ([1909] 1998) and extended by Clive Bell in his extremely popular book *Art* ([1914] 1947). Fry and Bell claimed from the evidence of introspection that the aesthetic experience was fundamentally different from the satisfaction achieved from consumer goods and services. In effect they insisted upon exceptional status for the arts and set out to explain why. They conceded that the Benthamite utilitarian model might be appropriate for understanding the satisfaction of biological needs of humans and other animals. But, using the term “imagination” to refer to the activity of the human mind as Hume and others had done before, they argued that an understanding of the “imaginative life” required more powerful analytical devices than the felicific calculus. Indeed, many of the Bloomsburys concluded that the simple application of Benthamite principles to the arts had impeded the development of a richer and more truly enlightening body of theory. Keynes made this point in his posthumously-published memoir “My Early Beliefs”. He wrote: “we were amongst the first of our generation, perhaps alone amongst our generation, to escape from the Benthamite tradition”. And he continued: “It can be no part of this memoir for me to try to explain why it was such a big advantage for us to have escaped from the Benthamite tradition. But I do now regard that as the worm which has been gnawing at the insides of modern civilisation and is responsible for its present moral decay” [Keynes (1949, p. 96)].

A few highlights of the efforts of Bloomsbury to move beyond the Benthamite paradigm may be mentioned. The original impetus for inquiry came from an awareness among these artists and writers that their responses to works of art seemed fundamentally different from those experienced during their consumption of food or other conventional goods and services. Roger Fry in a brief autobiographical fragment talked about how on reflection he saw himself “always groping my way towards some kind of a reasoned and practical aesthetic” [Fry ([1920] 1998, p. 87)]. For Fry the groping led first to rejection of the notion of beauty inherent in an artwork. He was inspired by an essay of Leo Tolstoy that portrayed art as the communication of “aesthetic emotion” rather than some sort of production and consumption of a good or service. Fry and Bell went on to speculate that the “form” of an art work rather than the content helped to determine the effectiveness of artistic communication. Some of the Bloomsburys reached out to such pioneer psychologists as Sigmund Freud and Wilfred Trotter to help them



understand the aesthetic experience; others, including Keynes, looked to philosophy and especially to the ethicist G.E. Moore. Regardless of their success in developing a new understanding of the artistic experience, high on the priority list of the Bloomsburys was always a desire to demonstrate the exceptional nature of the arts, a task that had been largely ignored by thinkers in the English-speaking world since the eighteenth century.

### 8.3. *The arts in society*

Most economists by the twentieth century had come to view the arts and culture as merely a rather uninteresting dimension of human consumption, perhaps to be viewed as luxurious and with or without externalities depending on your judgment, but unlikely in any case to have much impact beyond themselves. By contrast the Bloomsburys were fascinated with the ways in which the arts and culture through the course of history had conditioned and helped to interpret society, polity, and economy [Goodwin (2000)]. On this topic the artists, the novelists, and the social scientists in the Group could find common ground. Great works of religious art and literature fascinated them in particular. They were intrigued especially by Old Testament stories such as the accounts in the Book of Genesis of Adam and Eve in the Garden, and of Noah enduring the flood. These art works, they believed, contained hidden but intentional messages that affected human behavior. From the account of the Fall in the Garden readers, or viewers of pictures, could infer that hard labor under difficult agricultural conditions was the inevitable lot of man and was partly at least the result of female miscalculation. From the Noah story, as from Marxian doctrine that the Bloomsburys so much abhorred, readers were intended to conclude that catastrophe – a flood or a revolution – must precede fundamental improvement in the human condition. They were fascinated also by classical mythology that played such a prominent role in English public school education. The myth of Cupid and Psyche strengthened the portrayal of the unreliability and fecklessness of women; by contrast the stories of Pan and of Daphne and Apollo warned of the need to protect the natural environment. The connection between works of art and literature and social and economic affairs was explored by nearly all the Bloomsburys but especially by the artists Duncan Grant, Vanessa Bell and Roger Fry, the novelists Virginia Woolf, E.M. Forster and David Garnett, the biographer Lytton Strachey, and the political theorist Leonard Woolf. In the writings of Keynes there were repeated references to biblical and mythological characters and events. For example, he noted that “the old Adam” in humans keeps them from spending money when expenditures are needed, and the delusion that struck King Midas still causes humans to hoard rather than to circulate gold [see also Skidelsky (1992, p. 425)].

### 8.4. *Art markets*

The Bloomsburys paid close attention to various aspects of the art markets, but they found the conventional models in economics unsatisfactory to explain them. On the supply side they found that price was seldom the most important determinant of supply



– artists were as anxious as anyone for a decent living but for the most part they seemed impelled more by internal psychological pressures than by financial reward to produce their best work. The Bloomsburys came to believe that the main challenge for those concerned about the health of art markets should be to generate sufficient demand of the right kind to sustain the artists. Here again they found that price was not the most powerful independent variable. Roger Fry led the movement within the Bloomsbury Group to examine the demand for the arts in its various components and to probe the psychological circumstances of all the actual or potential providers of support for the arts. Fry and the others identified a number of distinct categories of demander to whom they gave, with their typically irreverent style, names such as “snobbists” (following Thackeray), “Philistines” (following Arnold), “Classicists” and “the herd” (following Trotter) [Goodwin (1998)]. The Veblenian instinct of emulation played an important part in the demand for art, but there were also the church, the monarchy, the landed aristocracy, big business, and an aesthetically-sensitive component of the middle class. In the Bloomsbury decomposition of the demand for works of art and in their attention to the psychological motivations of the various components, it is not hard to see a methodological precursor to the later Keynesian identification of the components of aggregate demand in the *General Theory of Employment Interest and Money* ([1936] 1947). There is even a parallel in the compensating role for the state proposed in both cases; public support should be introduced in the arts or the macro-economy, the Bloomsburys insisted, when private alternatives had been exhausted.

### 8.5. Policy reform

The topic of policy reform was where Keynes’s leadership was so important. Most of the Bloomsburys felt a strong sense of social responsibility – perhaps inherited from their Victorian families – to respond to policy challenges as they saw them and not just to sit on the sidelines and theorize. Roger Fry’s unceasing travels across Britain lecturing to large audiences on art and Virginia Woolf’s commitment to the “Common Reader” were manifestations of their sense of social obligation. They were advocates above all of private sector solutions to problems when these seemed possible.

The Bloomsburys concluded from their extensive personal experience that most artists were ill-equipped by training and emotional make-up to set up or operate firms that could effectively market their products to the complex mix of potential demanders. Accordingly they set out to experiment, often collaboratively, with a range of innovative market devices. The Hogarth Press and the *Nation* magazine, the latter directed by Keynes, were two publishing outlets designed to insulate the writer as much as possible from intrusive business and editorial pressures. The Omega Workshops started by Roger Fry and the London Artists’ Association managed by Keynes attempted to secure a regular minimum income for those artists who were willing to limit their outlets and trade off some sales opportunity for economic security. One of the most interesting private sector experiments (with which Keynes was involved as well) was designed to take account of the role of the critic, a matter about which the Bloomsburys felt strongly. The Con-

temporary Art Society was established in 1909 with Fry and Bell among the founders, to provide seals of approval to living or recently deceased British artists through certification by critical authorities. They hoped thereby to “educate the artistic opinion of the country” and engender confidence in inexperienced buyers of art. CAS members paid annual subscriptions and certain designated experts made purchases with CAS funds or made gifts of the works of artists whom they favored; these works were then exhibited, publicized and passed along on loan to local museums. Like the Book of the Month Club founded a few years later in the United States, the CAS was a pioneering attempt to correct what was perceived to be a unique market failure in art markets. The founders appreciated that it typically took years to establish an artistic reputation and by this time an artist could be impoverished. The CAS would correct an information gap and thereby help to assure support for those artists who deserved to survive.

Education and research in the humanities and the arts were desiderata for the Bloomsburys that they hoped could be accomplished by activities in the public and private sectors. Education in the liberal arts was justified on the ground that it was training for life, just as much so as vocational education. Technical education was justified typically on the ground that it increased the output of goods and services from a given set of resources. In exactly the same way education in the arts and literature was justified for its capacity to enable humans to experience a better life through gaining access to the best in the arts and literature. Research in the arts and literature was predicated on the proposition that there *was* a legitimate distinction between good and bad art and writing, and that good quality could be discerned and explained by trained people after sustained study. Just as education made possible extension of the *breadth* of use of the arts and literature, so research could increase the *depth* of this use. To the extent that enough people with resources perceived the social significance of education and research in the arts and culture, their nurture could be left to the market and private philanthropy. To the extent that the private sector proved inadequate, however, it would be necessary for the government to intervene.

The best-known public sector policy initiative with roots in Bloomsbury was the British Arts Council, established after World War II as a public funding mechanism with Keynes and Kenneth Clark as chairmen [Moggridge (2005); Goodwin (2005)]. By channeling public support for the arts mainly through private sector intermediaries the Council hoped to strengthen public support for the arts but lighten the heavy hand of governmental bureaucracy.

It may not be too much to speculate that if Keynes had successfully introduced the ideas of his Bloomsbury friends on the economics of art and culture to the economics profession, as he did his ideas on macro-economics, the result might have been just as revolutionary. Indeed the two approaches are strikingly similar. Both call for more contact with the discipline of psychology, and with other disciplines too, for explanations of how human actors behave in different market situations. At a time when applied micro-economics was beginning to assert its pervasive explanatory power because of the generality of its models, this call for a return to particularities and the evidence of experience would have been heresy. But the explorations started by this new line of in-

quiry might have been stimulating beyond a narrow focus on the economics of the arts and culture; they might have extended at least to the sub-field of industrial organization where the theory of the firm seemed of limited usefulness in the arts, and to the sub-field of public economics where it might have been conceded that the arts as an area of economic activity had established its exceptional nature and deserved scrutiny beyond conventional considerations of externalities and public goods. The parallels between Keynesian macro-economics and the Bloomsbury intimation of a cultural economics also go beyond economic theory. The richness of the Bloomsbury array of policy experiments in the arts is reminiscent of the recommendations of the young Keynesians as they explored ways to increase and manage aggregate demand.

## 9. Conclusion

By Keynes's death in 1946 attention to the economics of art and culture had not gone far. As this chapter has shown, there had been promising starts on important questions made by eighteenth century Enlightenment thinkers and in the twentieth century by the Bloomsbury Group, but they were not followed up. The development of a distinctive sub-discipline to address this subject was held back initially by suspicions of luxury, vice, and conspicuous consumption and subsequently by claims that all significant questions could be answered satisfactorily by the simple application of the emerging tools of marginalist micro-economics. All the same, numerous issues had been raised that remain before the economics profession still, notably whether positive externalities generated by the consumption of art justify public support. Those who today address either theoretical puzzles in cultural economics, or policy challenges, would do well to begin by reviewing the efforts of those who went before.

## References

- Arnold, M. ([1869] 1903). *Culture and Anarchy*. Macmillan, London.
- Baumol, W.J., Bowen, W.G. (1966). *Performing Arts: The Economic Dilemma*. Twentieth Century Fund, New York.
- Bell, C. ([1914] 1947). *Art*. Chatto and Windus, London.
- Bennett, O. (2005). "Beyond machinery: The cultural policies of Matthew Arnold". *History of Political Economy* 37 (3), 455–482.
- Bentham, J. (1962). *The Works of Jeremy Bentham*. Russell and Russell, New York. Edited by J. Bowring.
- Bodin, J. ([1578] 1947). *The Responses of Jean Bodin to the Paradoxes of Malestroit and the Paradoxes*, second ed. Country Dollar Press, Washington. Translated by G.A. Moore.
- Bodin, J. ([1606] 1962). *The Six Bookes of a Commonweale*. Harvard University Press, Cambridge, MA. Edited by K.D. McRae.
- Burke, E. ([1757] 1869). *A Philosophical Inquiry into the Origin of Our Ideas of the Sublime and Beautiful with an Introductory Discourse Concerning Taste*. Harper and Brothers, New York. Edited by A. Mills.
- Clark, K. (1969). *Civilisation*. Harper and Row, New York.
- Clay, H. (1918). *Economics: An Introduction for the General Reader*. Macmillan, London.
- Collins, J. (1984). *The Omega Workshops*. University of Chicago Press, Chicago.

- Fry, R. ([1909] 1998). "An essay in aesthetics". In: Goodwin, C.D., *Art and the Market: Roger Fry on Commerce in Art*. University of Michigan Press, Ann Arbor.
- Fry, R. ([1920] 1998). "Retrospect". In: Goodwin, C.D., *Art and the Market: Roger Fry on Commerce in Art*. University of Michigan Press, Ann Arbor.
- Galbraith, J.K. (1973). *Economics and the Public Purpose*. Houghton Mifflin, Boston.
- Galbraith, J.K. (1986). *A View from the Stands – of People, Politics, Military Power and the Arts*. Houghton Mifflin, Boston.
- Galbraith, J.K., Randhawa, M.S. (1968). *Indian Painting: The Scene, Themes, and Legends*. Houghton Mifflin, Boston.
- Galiani, F. ([1751] 1930). "Della Moneta". In: Monroe, A.E. (Ed.), *Early Economic Thought*. Harvard University Press, Cambridge, MA, pp. 279–307.
- Goodwin, C.D. (1998). *Art and the Market: Roger Fry on Commerce in Art*. University of Michigan Press, Ann Arbor.
- Goodwin, C.D. (2000). "Economic man in the Garden of Eden". *Journal of the History of Economic Thought* 22, 405–432.
- Goodwin, C.D. (2005). "Kenneth Clark: His case for public support of the arts". *History of Political Economy* 37 (3), 557–592.
- Grampp, W.D. (1973). "Classical economics and its moral critics". *History of Political Economy* 5 (2), 359–374.
- Howson, S. (2005). "Lionel Robbins' 'Art and the State' ". *History of Political Economy* 37 (3), 617–646.
- Hume, D. ([1752] 1965). "Of refinement in the arts". In: Lenz, J.W. (Ed.), *Of the Standard of Taste and Other Essays*. Bobbs-Merrill, Indianapolis.
- Hume, D. (1825). *Essays, Moral, Political, and Literary*, vol. I. Bell and Bradfute, Edinburgh.
- Hutcheson, F. (1729). *An Inquiry into the Original of Our Ideas of Beauty and Virtue*, third ed. J. and J. Knapton, London.
- Jevons, W.S. ([1871] 1965). *The Theory of Political Economy*, fifth ed. Augustus M. Kelley, New York.
- Jevons, W.S. (1904). *Methods of Social Reform, and Other Papers*. Macmillan, London.
- Jevons, W.S. (1905). *The Principles of Economics*. Macmillan, London.
- Jones, E. (1736). *Luxury, Pride and Vanity*. Hanson, London.
- Keynes, J.M. ([1930] 1972). "Economic possibilities for our grandchildren". In: *Essays in Persuasion, The Collected Writings of John Maynard Keynes*, IX. Macmillan, London, pp. 321–332.
- Keynes, J.M. ([1936] 1947). *The General Theory of Employment, Interest and Money*. Macmillan, London.
- Keynes, J.M. (1949). *Two Memoirs*. Rupert Hart-Davis, London.
- Lauderdale, J.M. Earl of ([1819] 1962). *An Inquiry into the Nature and Origin of Public Wealth and into the Means and Causes of Its Increase*, second ed. Augustus M. Kelley, New York. Edited by M. Paglin.
- Mandeville, B. ([1732] 1924). *The Fable of the Bees: Or, Private Vices, Publick Benefits*, sixth ed. Clarendon Press, Oxford. Commentary by F.B. Kaye.
- Marshall, A. ([1920] 1964). *Principles of Economics*, eighth ed. Macmillan, London.
- Marshall, A. (1996). *The Correspondence of Alfred Marshall, Economist*. Cambridge University Press, Cambridge. Edited by J.K. Whitaker.
- McCulloch, J.R. ([1845] 1938). *The Literature of Political Economy*. London School of Economics and Political Science, London.
- Menger, C. ([1871] 1950). *Principles of Economics*. Free Press, Glencoe. Translated and edited by J. Dingwall and B.F. Hoselitz.
- Mill, J.S. ([1909] 1964). *Principles of Political Economy*, new ed. Augustus M. Kelley, New York. Edited by W.J. Ashley.
- Mill, J.S. (1984). *Essays on Equality, Law, and Education*. University of Toronto Press, Toronto. Edited by J.M. Robson.
- Mill, J.S. (1986). *Newspaper Writings by John Stuart Mill*. University of Toronto Press, Toronto. Edited by J.M. Robson.
- Moggridge, D.E. (2005). "Keynes, the arts, and the state". *History of Political Economy* 37 (3), 535–555.

- Moore, G. (2005). "Evangelical aesthete: Ruskin and the public provision of art". *History of Political Economy* 37 (3), 483–508.
- Mosselmans, B., Mathijs, E. (1999). "Jevons's music manuscript and the political economy of music". In: De Marchi, N., Goodwin, C.D. (Eds.), *Economic Engagements with Art*. Duke University Press, Durham, pp. 121–156.
- Peacock, T.L. (1926). *Critical and Other Essays. The Works of Thomas Love Peacock. IX*. Constable, London.
- Rae, J. ([1834] 1965). *Statement of Some New Principles on the Subject of Political Economy*. In: James, R.W. (Ed.), *John Rae: Political Economist, vol. II*. University of Toronto Press, Toronto.
- Ricardo, D. ([1817] 1948). *The Principles of Political Economy and Taxation*. J.M. Dent, London.
- Robbins, L. (1963). "Art and the state". In: *Politics and Economics: Papers in Political Economy*. Macmillan, London.
- Ruskin, J. ([1862] 1905). *Unto This Last*. George Allen, London. Edited by T.E. Cook and A. Wedderburn.
- Ruskin, J. ([1880] 1905). *A Joy For Ever (and Its Price in the Market)*. George Allen, London. Edited by T.E. Cook and A. Wedderburn.
- Skidelsky, R. (1992). *John Maynard Keynes: The Economist as Saviour 1920–1937*. Macmillan, London.
- Smith, A. ([1759] 1976). *The Theory of Moral Sentiments*. Clarendon Press, Oxford. Edited by A.L. Macfie and D.D. Raphael.
- Smith, A. ([1776] 1976). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Clarendon Press, Oxford. Edited by W.B. Todd.
- Smith, A. ([1795] 1980). *Essays on Philosophical Subjects*. Clarendon Press, Oxford. Edited by W.P.D. Wightman.
- Solomon, M. (1979). *Marxism and Art*. Wayne State University Press, Detroit.
- Turgot, A.R.J. ([1750] 1973). *Turgot on Progress, Sociology, and Economics*. Cambridge University Press, Cambridge. Edited by R.L. Meek.
- Upchurch, A. (2005). "William Morris and the case for public support of the arts". *History of Political Economy* 37 (3), 509–534.
- Veblen, T. ([1899] 1934). *The Theory of the Leisure Class*. Modern Library, New York.
- Williams, R. (1958). *Culture and Society 1780–1950*. Chatto and Windus, London.

## THE HISTORY OF ART MARKETS\*

NEIL DE MARCHI and HANS J. VAN MIEGROET

*Duke University, Durham, North Carolina, USA*

### Contents

Abstract	70
Keywords	70
1. Approach, focus and scope	71
2. Early local art markets, primary and resale	72
2.1. Forms of exchange	72
2.2. The primary market in 15th-century Florence	73
2.3. Florentine estate sales: Early specialization among second-hand dealers	76
2.4. The primary market in 15th-century Bruges: Scale; foreign demand; artistic emulation	77
2.5. Wealth, and the ownership of paintings: Florence versus Amsterdam	84
3. Distributing paintings across markets	86
3.1. The <i>panden</i> in 15th- and 16th-century Antwerp	86
3.2. Dealer-led incursions into Paris, and to Dutch cities	90
3.3. The market in Holland as a network	93
3.4. Trading in paintings: The traffic from Antwerp to New Spain, via Seville, c. 1540–1670	95
4. Selling paintings I: Artist–dealer relations	97
4.1. Securing payment for invention	97
4.2. Selling paintings in Venice and Rome	99
5. Selling paintings II: Auctions	104
5.1. “Dutch” auctions in Amsterdam	104
5.2. Magistrate-imposed rules for transparent auctions in 17th-century Lille	104
5.3. Auctioneer-contrived rules, London, late 17th century	106
5.4. Creating value through auctions: The dealer Gersaint in Paris, 1733–1750	107
6. Coda	108

\* Thanks are due to many colleagues for conversations, advice, information, access to their unpublished research, and key references, and for their critical suggestions on portions of or the whole of our text. We are particularly indebted to Linda Borean, Isabella Cecchini, Miguel Falomir, Richard Goldthwaite, Frank Heims, Dorothea Herreiner, Loredana Lorizzo, Christopher Marshall, Louisa Matthew, Michael Montias, Sophie Raux, Richard Spear, Mickaël Szanto, Filip Vermeylen and Ilya Voytov. Special thanks to Mike Scherer for a critical reading of a near-complete draft, and to Victor Ginsburgh for encouragement, suggestions, and extraordinary editorial forbearance.

*Handbook of the Economics of Art and Culture, Volume 1*

*Edited by Victor A. Ginsburgh and David Throsby*

Copyright © 2006 Elsevier B.V. All rights reserved

DOI: 10.1016/S1574-0676(06)01003-9

7. Some developments since 1750	112
Appendix A	114
A.1. Number of painters per '000 of population, by city, 1470–1764	114
A.2. Sources and methods	115
References	116

## Abstract

Treating markets as arenas where relative advantage is contested, this entry explores the emergence and evolution of Western markets for paintings, 1450–1750, in terms of the players, their creative moves to secure gain, and the rules they devised to maintain order. Primary markets for paintings arose as a derivative of the commission market for one-off, mainly religious paintings in places such as Florence and Bruges, in the second half of the 15th century. Demand from foreign merchants eager to obtain works in the new medium, oil, gave Bruges an edge. So did a demand for easel paintings on thin linen, and even in oil on panel, as cheap substitutes for tapestries. Variety and cost also played a role. Emulation among differently-trained artists generated novel products plus extraordinary cost reductions, and painters discovered a latent demand among the less wealthy. Some novel products were exported, as were new techniques. The retail market in Florence was limited in size and largely confined to serving a need for cheaper versions of unique, public commissions. A mass demand for paintings across the social spectrum occurred principally in Northern cities: e.g., Antwerp, and later Amsterdam, though also in Spain. Resale markets followed retail with a lag, recycled paintings being handled by second-hand clothes dealers. This sequence – commission nexus, cost-reduction and novel sorts of paintings, mass retail, then resale markets – occurred in cities across Europe. As mass markets emerged, so too did specialist dealers. A large part of the entry is devoted to detailing their creative marketing moves. There were tensions as to whether only artists might sell, but demand mostly overrode guild reluctance to relinquish control of distribution. Widespread distribution came to require efficient sales mechanisms, hence public sales and auctions. The entry explores auction rules and techniques within the broader sequence identified above.

## Keywords

rules vs. plays, guilds, dealers, auctions

*JEL classification:* D4, L8, N8, Z12

## 1. Approach, focus and scope

Competition is a notion linked to a particular sort of market structure and its associated set of equilibrium properties; in the practical realm, however, competition is very clearly a creative struggle for advantage. Focusing on the latter, as we will, it makes sense to think of markets as arenas within which various players engage flexibly in experimental strategizing for relative gain.<sup>1</sup> Historically, at an early stage, regulations or rules tend to be put in place to govern such competitive struggles. Those rules, along with certain structures – avenues for seeking the redress of grievances, for example – constitute institutions which, together with the creative strategies adopted by the players, come to define the structure of a market and the way markets function.<sup>2</sup> The gaming analogy implied in talk of players and rules derives its force from the fact that games too have a structure, and typically it is one that allows for probing, rule-challenging and rule-modifying plays.<sup>3</sup> Our history of art markets will be guided analytically by the categories roles, rules and creative plays, though we will also try to maintain a sense of chronology.

Some disclaimers about our scope are in order. Though certain mechanisms of art markets have quite ancient and non-Western origins – auctions, for example – we will confine ourselves to the emergence of such markets in Europe during the Renaissance and Early Modern period, roughly 1450 to 1750. As to regional concentration, the Low Countries have been the focus of much of the available research; as to terminal date, all the main innovatory elements we now know as standard were in place by about 1750. Finally, most research deals with paintings, and we will follow the available research.

Research on the history of art markets depends heavily on surviving guild records, inventory studies and legal documents. Survival being a matter of historical accident, the record is patchy. Our overview, unavoidably, will reflect this, and will thus be both episodic and somewhat disjointed. A nine-point synthesis is provided in a coda (Section 6) that may usefully be read in advance.

Prices have not been adjusted for inflation but have been expressed in terms of Florentine gold Florins, a money of account. Conversions are indicated in-text and in notes.

<sup>1</sup> De Marchi and Van Miegroet (1994, p. 452).

<sup>2</sup> Rules and structures may be informal and depend for their effectiveness on mechanisms such as loss of reputation or the ostracizing of offenders. As Grief and others have shown [Grief (1993); Grief, Milgrom and Weingast (1994)], such forms of self-regulation can work even in the absence of an official legal basis.

<sup>3</sup> Our conception of markets derives in part from North (1990), and is in part inspired by Grief's work. For illustrations of our conception at work see De Marchi and Van Miegroet (2000b).



## 2. Early local art markets, primary and resale

### 2.1. *Forms of exchange*

Makers and buyers meet in exchange relationships, which take various forms: commission exchange and gift exchange are at one end of a spectrum, market forms at the other. Both commission and gift exchange involve a more or less unique product, and contact is between one buyer and one seller, often direct, or mediated at just one remove by an agent. Valuation is involved in both these forms, but in neither is the primary motivation of a purchase asset value – the painting as potentially re-saleable, for gain. Power, decorum, honor and display were important within gift exchange and these criteria were culturally rather than market-determined.<sup>4</sup> Artistic and functional (mainly religious) value entered significantly into commission exchange, along with reputation (on both sides: purchasers and artists). Prevailing norms also affected the value of artists' services. Yet there was room in these parameters for negotiation and for individuating the terms of an exchange. Thus, historically, in the Central Italian commission nexus of the 15th century, prices fixed in advance were rare; more usually a range, or a maximum, or a minimum, was specified. Price – and implicitly the artist's reward – would then be determined as part of an assessment of the end product, the estimation usually being done by a committee assembled for the purpose. Both commission and gift exchange were practiced within certain strata and segments of society in our period; here, however, we can do little more than acknowledge that fact and use these forms as a foil.<sup>5</sup>

In contrast to both commission and gift exchange, market exchange – our principal concern – involves buyers and sellers who are relatively numerous; moreover, valuation takes place via bids and offers for products that, while they may be somewhat unique, are also regarded as reproducible.<sup>6</sup> Markets may be primary (first sale) or secondary (resale), and transactions in both may be primarily retail or mixed retail/wholesale, but the basic distinction is between primary and resale. Historically, primary markets involved artists doubling as dealers – in their own and other artists' products – long before specialist, professional dealers emerged. We find the first substantive European evidence for primary markets and the retailing of paintings in 15th-century Florence and Bruges. Resale markets, temporally, followed primary markets with a lag of fifty years or more. The clearest evidence concerning the auctions that served resale markets pertains to early 17th-century Amsterdam, late 17th-century London, and to Paris in the first half

<sup>4</sup> For the values implicit in gift exchange see Welch (2003) and references there cited.

<sup>5</sup> Commission exchange is best approached through the study of contracts, for which the basic source is Glasser (1977); see O'Malley (2005) for a valuable study of altarpiece contracts and their prices in 15th- and 16th-century Central and Northern Italy. Details on commission contracts for the Netherlands in the same centuries, especially the 15th, are scarce.

<sup>6</sup> This implies recognizing, as was the case, (a) that techniques for copying were part of best workshop practice, (b) that copies were accepted – even asked for – by the best collectors, and (c) that prints were a widespread form in which imagery was enjoyed, and not only at the low end of the market.

of the 18th century. But their earliest manifestations were in the form of estate sales and forced – court-ordered, and bankruptcy – sales. Those who traded in these early resale markets were dealers in second-hand goods, initially used clothes. Specialist dealers in paintings in the resale or secondary market followed, but for a long time appear to have shared the trade

- (1) with artists, who frequently dealt in paintings on the side and indeed were often the only people allowed by the guild to market paintings;
- (2) with agents who served as intermediaries and advisers to those who did not want to get personally involved in finding paintings or haggling over price; and
- (3) with street dealers in new and old paintings, stallholders at fairs, and with shopkeepers.

For the latter, more often than not, paintings were just one line within a broader range of merchandise.

The sequence just sketched applies across Western Europe, north and south of the Alps, though much of the creativity that interests us was expressed, as we would expect, with significant local differences. Indeed, as secondary markets join primary sales, it is the ingenuity of specialist dealers and auctioneers, or dealer-auctioneers, bent on expanding or maintaining markets, who supply much of the fascination. In some cities the primary art market always had a sizeable import component; other cities tended to be, or to become over time, net exporters of paintings. This difference is apparent from the information assembled in [Appendix A](#). Dealers in net-importer cities – pragmatically, those with ratios equal to or less than one – tended to function primarily as intermediaries and arbitragers across space. Dealers in net-exporter cities, on the other hand, often had to devise strategies to secure market share in other places, near and far, in order to sustain a production capability in paintings at home following a slowdown in the growth of local demand. Antwerp and Mechelen/Malines are prime examples.

We shall begin with the primary market first observable in Florence, probably a net-importer city, in the second half of the 15th century. The early phase of a secondary market – when second-hand dealers dominated – can also be observed there, from around the turn of the 16th century. This experience will then be contrasted with that of contemporary Bruges, the pre-eminent *entrepôt* of the western world, and a net exporter of paintings. Thereafter we will introduce local developments, in roughly chronological sequence, as noted, though emphasizing creative dealing, and related behaviors which inject intriguing variants into the general sequence sketched above.

## 2.2. The primary market in 15th-century Florence

Florence was perhaps the single most prolific center of the new painting, and of ideas concerning painting, that we associate with the Renaissance. Much of the work that has survived and is widely admired was commissioned: for churches and their chapels, or for religious orders, or for charity hospitals, guilds, and civic buildings. Some was intended for distinguished private houses (*palazzi*). In many instances involving a public commission, a competition was conducted, and a selection made from among mod-

els, fashioned according to published guidelines and submitted for consideration. The guidelines, however, did not preclude artistic ingenuity; indeed, it was favored. This made each commission unique, even if there was a certain formalization of modes of representation and an expanding number of well-rehearsed subjects. Uniqueness, or discreteness of design, meant that demand was discontinuous. Moreover, though in the course of the 15th century the Medici family acquired more and more power, Florence remained a republic; there was no Ducal Court.<sup>7</sup> For this reason and because commissioning bodies were many, with a variety of goals, there was no “single dominating authority” around which taste might coalesce.<sup>8</sup>

The resulting fragmentation of, and discontinuity in, demand was accentuated by artists themselves: their very competitiveness conditioned buyers to expect ever more ingenuity.<sup>9</sup> Within the commission circuit, therefore, competition was confined to the artistic; it could not express itself in cost-saving innovations in the way production was organized. Yet, as the surviving business records of the Florentine painter Neri di Bicci illustrate, a retail market existed in ready-made and made-to-order altarpieces, painted wooden tabernacles (containing one or more devotional scenes or images) and painted devotional forms in gesso and terra cotta relief. Often these were based on works produced within the commission nexus, though Neri di Bicci’s altarpieces were also recognizably his own.

As we would expect, the average price of the “derivatives” made in this market was below that of commissioned works or those chosen for a private collection. Table 1 shows this to have been the case. Neri di Bicci’s clientele ranged from barbers to patricians; the largest segment (40 percent), however, were middling sorts of persons: artisans and politically successful members of the lesser guilds, or members of lower-status patrician families.<sup>10</sup> Such clients could limit their outlay by opting for smaller variants of standard items and restricting the number of figures, though without forgoing ornamentation such as richly carved surrounds, the use of gold and silver leaf, gold filament for the detailing of fabric, etc.<sup>11</sup> In these ways Neri di Bicci managed to supply devotional imagery inspired by the publicly-displayed creations of artists in the commission circuit, and sharing some of their artistic values, yet suited to lower budgets. And he was successful at this: a wealth tax assessment of 1480 placed him ahead of all other painters in Florence.<sup>12</sup> Moreover, he was included in a listing of 1470 by Benedetto Dei of artists who maintained a workshop in Florence, along with some famous names.<sup>13</sup> It

<sup>7</sup> This changed in the 1530s, when the ruling member of the Medici declared himself a Duke.

<sup>8</sup> Goldthwaite (1982, p. 412).

<sup>9</sup> *Ibid.*, p. 421.

<sup>10</sup> Holmes (2003, p. 218; see also Note 40, p. 223), in Fantoni, Matthew and Matthews-Grieco (Eds.) (2003).

<sup>11</sup> *Ibid.*, p. 217. Neri di Bicci’s most popular altarpiece is said to have comprised a panel of c. 2 × 2 meters, with the *Madonna and Child* flanked by 2 to 4 saints, the whole costing just 2.75 to 8.5 florins (*ibid.*, Note 36, p. 222); but he also made many much smaller panels [Kubersky-Piredda (2003, p. 117)].

<sup>12</sup> Commanducci (2003, p. 106).

<sup>13</sup> Gilbert (1980, pp. 182–183) contains a translated extract of Dei’s observations.

Table 1  
Prices, commission circuit vs. derivative market, 15th-century Florence

	Commission circuit <sup>5</sup>			Derivative market <sup>5</sup>	
	Av.	Range		Av.	Range
<i>Devotional images</i>					
De' Medici <sup>1</sup> ( <i>N</i> = 3)	24.7	4–40	di Bicci <sup>2</sup> ( <i>N</i> = 100)	16.0	2.75–45
<i>Altarpieces, 4–6 sq. meters</i>					
Central Italy <sup>3</sup> ( <i>N</i> = 8)	120		di Bicci <sup>4</sup> ( <i>N</i> = 5)	28.9	

<sup>1</sup>Three Flemish panels, two attributed: Nuttall (2004, Appendix A).

<sup>2</sup>Works made 1454–1474: Kubersky-Piredda (2003, p. 121).

<sup>3</sup>O'Malley (2003, Table 4, p. 176).

<sup>4</sup>Kubersky-Piredda (2003, p. 121).

<sup>5</sup>All prices in Florentine florins.

is worth noting that original paintings in Amsterdam inventories, 1620–1660, averaged about 2.2 florin-equivalents, below Neri di Bicci's minimum prices.<sup>14</sup>

Neri di Bicci shared two difficulties with artists in the commission circuit: the difficulty of maintaining a workshop between sales, and the need to advertise.<sup>15</sup> Unlike those artists, however, he was able to organize his business so as to reap the cost savings associated with repetition: he sub-contracted for bulk orders of standardized gesso and terra cotta reliefs, and for similarly standardized wooden tabernacles.<sup>16</sup> He differed from those in the commission circuit also in that his margins, generally, were quite low – Kubersky-Piredda estimates 10–15 percent of the price, to cover his fee as well as minor materials and food costs. This was significantly less than the margins paid painters of superior ingenuity and skill, such as Sandro Botticelli who, late in the century, could earn 35–45 percent of the total cost of commissioned altarpieces.<sup>17</sup>

<sup>14</sup> Montias (2002a, p. 118). The estimated average price in Holland was about 13 guilders (of 20 stuivers). The guilder of 20 stuivers had been equivalent to the gold Carolusguilder when the latter was first issued in 1521, and the Carolusguilder was equivalent to 0.7 of a Rhineguilder, which in turn was equivalent to 0.75 of the Florentine gold florin. The Carolusguilder disappeared from circulation in the course of the 16th century; but, based on gold content, the original guilder of 20 stuivers had slipped to the equivalent of just 0.17 of a gold florin by the 1620s. In terms of equivalences in 1521, the following relationships had held: 1 guilder (20 stuivers) = 1 Carolusguilder = 0.7 of a Rhineguilder = 0.75 of a gold florin; so that 2 guilders (of 20 stuivers)  $\approx$  1 florin. Because of debasement, however, in the 1620s it would have taken roughly 5.9 guilders of 20 stuivers to purchase 1 gold florin. Silver guilders were of much lower value. See De Vries and Van der Woude (1997, pp. 80–81, and graph on p. 85).

<sup>15</sup> See Wright (2003, pp. 225–236) on the goldsmith-artist Antonio Pollaiuolo and his marketing devices. Thomas (1995, pp. 88–93), discusses variations in di Bicci's workshop size, and Matthew (2006), in De Marchi and Van Miegroet (Eds.) (2006), notes Neri di Bicci's promotional pricing of portraits.

<sup>16</sup> Kubersky-Piredda (2003, p. 116).

<sup>17</sup> Blume (2003, pp. 152–153). Two observations only. Data are scarce, so these figures may or may not be representative.

We possess only anecdotal evidence of dealers in the primary market in Florence. Kubersky-Piredda supplies information on three such dealers in the period 1450–1500. Two of the three purchased from Neri di Bicci, but for each, paintings were merely a side line. Their primary business lay elsewhere: in banking, haberdashery, and jewelry, precious stones and other high value-added objects.<sup>18</sup> Specialized dealing, as noted, is first observed in the trade in second-hand goods.

### 2.3. Florentine estate sales: Early specialization among second-hand dealers

Regular estate sales were common in most cities in Europe and they were dominated by second-hand dealers, corporations of which date from the late 13th century. In most places these dealers were given a name that reflected their main line, old clothes and cloth goods in general, in various states of disrepair, often just rags: hence *stracciaroli* or *strazzaroli* in Venice, *rigattieri* more generally across Italy; *oudecleerkopers* (second-hand clothes dealers) in the Netherlands; *fripriers* in France; rag and bone men in England. Sometimes more generic designations were used, for instance, to indicate that one of their principal roles was to clear out the house of a deceased person (hence *uitdraagsters* in the Dutch Republic), or to repair household goods preparatory to recycling them (hence “upholders” in England).

In Florence, in the last decades of the 15th century, signs of specialization begin to appear in the meticulous records kept there of the public auctions of estates. Records of the sales include the names and bids of all bidders for each lot, including of course but not limited to the winning bid. In principle, therefore, we are able to identify concentrations of types of bidder for particular sorts of goods. By late in the century lots at estate sales were being organized so as to separate cloth goods from other household movables (e.g., furnishings, including paintings), suggesting that a specialization of sorts had already begun among the *rigattieri*. The 1479 inventory of a *rigattiero* contained only artworks and furniture, while an auction of 1498 listed a lot comprising 26 works of art, including paintings, and the winning bidder was from a family of known second-hand dealers.<sup>19</sup> With a lot of 26 objects, we should add, a sale almost qualifies as part of a wholesale market.

The distinction in Florence within the ranks of the second-hand dealers became sharper with time, and presumably was mirrored elsewhere, though a terminological distinction to match the situation on the ground seems to have been rare.<sup>20</sup> In any case,

<sup>18</sup> Kubersky-Piredda (2003, p. 117).

<sup>19</sup> See Lydecker (1987, pp. 214–216).

<sup>20</sup> In Spain, France, England and the Netherlands (north and south) no distinct terminological acknowledgement was made of the specialization among second-hand dealers. Nor was there any in Rome; while even in Naples, far from the specialization being named, it was not even accepted: *li mastri della Giudeca* were confined, late in the 17th century, to dealing in clothes and *not* allowed to deal in other household goods [Allerston (2003, p. 311, n. 57), citing information supplied by Christopher Marshall].

as far as we know, nothing comparable occurred in Bruges.<sup>21</sup> It was not until 1588, and then in Antwerp, that we find two men registered in the painters' guild – officially required of those wishing to deal – but as “second-hand clothes dealer, also trading in paintings” (*oudecleercoper ende handelt met schilderijen*).<sup>22</sup> By the middle of the 16th century the recycling market in Florence had developed to the point where the city designated a sales venue in the city center, opened a warehouse to store goods pending their sale, hired estimators to value each lot (which estimates became the starting point for bids), and ordained that any citizen might hold sales – not only estate sales – using the facilities.<sup>23</sup>

#### 2.4. *The primary market in 15th-century Bruges: Scale; foreign demand; artistic emulation*

In the fourteenth century Bruges was the most important trading center in Western Europe, and the second-largest city (after Ghent) north of Paris, with a population of some 46,000. It was surrounded by textile-producing towns, but never itself became an industrial city. Its strength was trade – *entrepôt* trade – and the city developed sophisticated institutions to maintain its pre-eminence in that sphere: space for displaying the textiles immediately preceding its May fair; inspection arrangements and quality guarantees; and a system of licensed brokers, through whom deals outside fairs were to be closed. Innkeepers, many of whom also hired brokers to close deals, supplied storage and provided financial services.<sup>24</sup> These institutional arrangements were strictly supervised for the most part, brokerage fees, and tolls, being important sources of income to the city.<sup>25</sup> The same predilection for oversight, order and control was also reflected in guild regulations, and in the strict boundaries between crafts. The city's painters fell under the same restrictions as other crafts though, as will be seen, the magistrates could also wink at breaches of guild regulations when the city's trade was at stake. By the mid-15th-century Bruges' population had fallen to about 40,000, and in the next five decades the city's economic prospects were damaged by serious, continued silting of its access river and, after 1582, by political differences with the claimant to the privileges exercised by successive Burgundian dukes, the future Holy Roman Emperor Maximilian I. The dispute resulted in the decamping to Antwerp – at first temporary, but later permanent

<sup>21</sup> We know of no documentary indications of specialization or re-naming among the *uitdraagsters* in the Dutch Republic even in the 17th century, though we know a good deal about individuals and their participation in sales involving paintings from Montias' study of Amsterdam Orphan Chamber Auctions, 1597–1638 [Montias (2002b)]. By the mid-17th century the Friday (estate) Market in Antwerp had become a place to acquire old paintings, many of them not cheap. But at that time it was frequented for the purpose by knowing, specialist dealers, and their agents, not just old clothes dealers.

<sup>22</sup> See Vermeylen (2003, note 37, p. 132).

<sup>23</sup> Lydecker (1987, p. 217).

<sup>24</sup> See Murray (2000), Stabel (2000) and Greve (2000), all in Stabel, Blondé and Greve (Eds.) (2000).

<sup>25</sup> Nicholas (1992, pp. 204, 296).

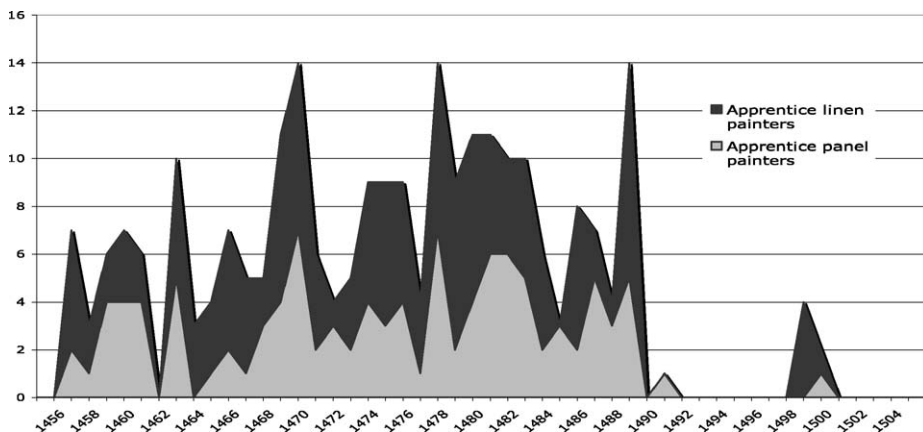


Figure 1. Numbers of newly apprenticed cloth and panel painters, Bruges, 1456–1500. Note: Our source excludes masters’ sons. Our numbers therefore differ from some other counts, but omitting masters’ sons captures more accurately the incentives in play. Source: [Vanden Houte \(1913\)](#).

– of the foreign “nations” (resident foreign merchant communities). Painters fared better than many crafts during the decades of decline, partly because they were able to generate new demand at the low end of the retail market.

That demand can be inferred from the numbers of new master painters registering during the second half of the 15th century; and even more clearly from the numbers of new apprentice painters. As [Figure 1](#) shows, in some years the number of new apprentice painters exceeded twenty. No comparable record exists for Florence. By way of crude comparison, however, the total number of painters with a workshop in Florence in 1470, according to Dei’s list, was under 30. That estimate roughly accords with a separate list of “figure painters” in the Company of St. Luke, a confraternity of painters, in 1472.<sup>26</sup>

One factor clouding the numbers for Florence is that artisans there could move freely between crafts without acquiring official membership in any guild but that of their original training. Andrea del Verrocchio, for example, having failed as a goldsmith, moved successfully into painting (as well as carving and casting). Antonio Pollaiuolo, on the other hand, who also trained as a goldsmith, maintained a successful goldsmithing workshop whilst freely collaborating with his brother on painting projects.<sup>27</sup> Lower recorded numbers of painters registered as such in Florence may thus understate the true supply of artists’ services.<sup>28</sup> Both Verocchio and Pollaiuolo, however, made it onto

<sup>26</sup> [Wackernagel \(1981, p. 300\)](#). Wackernagel notes that 42 were listed as members adding, however, that some were not painters at all, while others were engaged in related crafts such as gold leaf production. Dei’s own list included several painters already deceased.

<sup>27</sup> [Wright \(2003, pp. 227–228\)](#).

<sup>28</sup> It also disadvantages Florence to count only figurative painters there; the Bruges lists are more inclusive.

Dei's list, suggesting that he was a knowing observer. In any case, there are strong indications that painters in Florence faced less favorable conditions than artists in Bruges. They found themselves having to advertise, as noted earlier. Moreover, many of them moved to other cities in pursuit of work.<sup>29</sup> By contrast, in Bruges, between 1466 and 1496, almost one-third of new masters in the image makers' guild, to which painters belonged, were immigrants.<sup>30</sup> Since new painters accounted for more than two-thirds of all new master image makers in this period, it seems safe to assume that there was also substantial net immigration of painters.

There is also evidence of a quite different sort that distinguished Bruges from Florence. Demand in Bruges included a component contributed by foreign merchants. And in that respect Bruges reaped collateral benefits that Florence lacked. An annual international fair, plus resident foreign "nations", meant a whole infrastructure of institutional arrangements to facilitate exchanges.<sup>31</sup> These, as noted earlier, were the city authorities' conscious contribution towards maintaining Bruges' pre-eminence as a trading center. Textiles predominated but the two-way exchanges of goods in Bruges differed according to the offerings and needs of participants, and the merchants present included Catalans, Castilians, Portuguese, Lucchese, Milanese, Genovese, Florentines, representatives of the Hanseatic towns, French, English and Scottish. It is thought that foreign merchants and their assistants might have numbered 1000 outside the period of the May fair in Bruges, and perhaps double that during the fair – or roughly one in ten heads of households there.<sup>32</sup> Florence, by contrast, sent bankers and merchants abroad, but there was no established mechanism or tradition of foreign merchants visiting or congregating in Florence, much less staying there for extended periods. Florence, paradoxically, was a city of home-grown merchants, yet not the host of an international market. Or, as Goldthwaite puts it, "the Florentine marketplace ... for all the city's position in international banking and commerce, had only a regional scope".<sup>33</sup>

Nor was the contribution of visiting and resident foreign merchants to Bruges limited to their numerical strength; on average, they had greater buying power than the local population.<sup>34</sup> Moreover, the wealth-elasticity of demand for art among them may

<sup>29</sup> Goldthwaite (1982, p. 420). Goldthwaite notes that in provincial Verona they even outnumbered locals.

<sup>30</sup> Blockmans (1996, p. 20).

<sup>31</sup> Goldthwaite (2003, p. 432). Stabel (2006) notes that the paintings market in Bruges could draw on a whole history of such infrastructural elements as part of the city's longstanding contribution to exchange.

<sup>32</sup> Blockmans (1995, p. 15), and Stabel (2001, p. 191). A moneychanger's records for 1366–1368 contain the names of 990 foreign merchants: Murray (2000, p. 7).

<sup>33</sup> Goldthwaite (2003, p. 432).

<sup>34</sup> Direct evidence of this for Bruges is lacking – though see Nicholas (1992, p. 296) – but it is a fair inference from what is known of the Amsterdam merchant population, which in the early 17th century included a significant proportion of immigrants from the Southern (Spanish) Netherlands. Montias has shown, from inventories, that "virtually all" the paintings valued at 10 gulden (1.7 florins) or more were owned by individuals who paid a minimum of 5 guilders in tax (0.8 florins) in a 1631 wealth assessment in Amsterdam. "The bulk of this demand [moreover] originated with merchants, along with a few brewers, sugar refiners, and other small-scale 'makers of goods'. They are the people who paid 50 gulden [c. 8.5 florins] or more in



well have exceeded 1.0. The available estimates of the wealth elasticity of art are for 17th-century Holland; there, most paintings were purchased by merchants, and several separate data sets put the elasticity above unity.<sup>35</sup> Though we lack such estimates for Bruges, an illustration of the significance of merchant demand is to be found in the case of Hans Memling, a German-born artist who settled in Bruges. The socio-economic status of the buyers of almost half of the works attributed to Memling is known; over 60 percent of them were foreign merchants.<sup>36</sup> But if local demand in Bruges was supplemented importantly by the demand of foreign merchants, things were the other way around in Florence. There, the merchant/bankers *imported* paintings made in Flanders on commission or obtained at the annual fairs. Inventories of the Medici family in the 15th century show that close to one-third of the 142 paintings registered in the 1492 inventory of Lorenzo the Magnificent were imports from Flanders, or Antwerp in the neighboring Duchy of Brabant.<sup>37</sup>

There is yet another factor that distinguished conditions in Bruges: competition among different sorts of specialist painters. We have mentioned artistic emulation among artists vying for success in competitive commissions in Florence. The competition in Bruges was different, and had a quite different impact on the primary market for paintings there. Three distinct crafts of painting were recognized in Bruges. The image makers' guild included two of them; there were painters in oil on wood panels and painters using pigment mixed with rabbit skin glue – a mixture called “watercolor” – on thin linen. These were the *clederschrijvers*, or *cleerschrijvers* (literally, “writers” on cloth/clothing). The third group was the illuminators or miniaturists, who had been shunted off into the *librariers'* guild. The number of the illuminators is not known, but as an artistic force they were significant; no less a painter than Jan van Eyck was trained as an illuminator, though that is not the kind of influence we have in mind here. The three crafts competed for market share, and the competition was by product type as well as by price. Moreover, for all that the groups maintained a separate identity, their engagements with each other as competitors issued in mutual borrowings and adaptations of each other's subject matter, techniques, formats and supports. The result was increased variety, including many items not seen before, and products available at every level of price. Again, to judge from the numbers of apprentices taken on by *cleerschrijvers* and panel painters in the image makers' guild – a whopping 242 between 1456

taxes, the top quarter of the distribution of taxpayers” [Montias (2002a, p. 117)]. Gelderblom (2000, p. 227), has shown further that of the 477 taxpayers assessed 100 guilders (c. 17 florins) or more in the 1631 assessment, 56 percent were merchants, more than two-fifths of them of southern origin. Montias' conclusions give substance to the shape of the tax “pyramid” in Amsterdam: see Figure 2.

<sup>35</sup> Montias estimated 1.23 in a study based on Delft inventories: (1982, p. 265); 1.37 for “works of art” sold at auction in Amsterdam, 1598–1638: (1996, pp. 75–76); and 1.46 from a random sample of Amsterdam inventories (1620–1660): (2002b, p. 34, table). No comparable studies exist for Bruges, or Antwerp, both because surviving inventories do not contain valuations and because no wealth tax assessments were conducted in either city for “normal” years.

<sup>36</sup> Blockmans (1996, pp. 23–24), citing the work of Maximiliaan Martens.

<sup>37</sup> Nuttall (2004, p. 106).

and 1490 – the changes must have elicited fresh demand.<sup>38</sup> Thus we may think of the foreign merchants as adding to final demand for a given range of products, while competition between the three groups of painters issued in novel products, which in turn also spawned new demand.

The differences in technique and end-product between the three sorts of painter in Bruges were so clear – each craft of painting occupying its own market segment – that the three ought to have been able to co-exist peacefully. Thus, by tradition the miniaturists' business was manuscript illumination; they decorated texts, but did not intrude on the making of easel paintings or hangings. The cloth painters' medium enabled large surfaces to be covered easily and quickly. They painted on thin linen, readily available in Flanders, and sometimes even known as *draps de Bruges*. And, since their material support was cheap, and relatively little labor time was involved in this sort of painting, cloth painters, for the most part, occupied a distinct segment of the market for easel paintings and hangings: banners, flags, temporary decorations, for example. But not all their products were fleeting or inexpensive. They also made the cartoons for tapestries and some of their larger, narrative scenes were bought as tapestry substitutes. Those could be costly: one such cloth painting owned by the Medici was valued at 10 florins, and another at 25. Nonetheless, oil painting on panel was the preferred technique for altarpieces, portraits, and devotional paintings of superior artistic pretensions. That technique called for layers of pigment interspersed with multiple lightly-tinted glazes. The work was tedious and time to dry long, and the labor-cost therefore relatively high. At the same time the contrasts, depth, shading and intensity achieved with oil were greater, giving a satisfying impression of richness to warrant, and match, a higher price.

Despite these traditional, technical distinctions, two changes in Bruges in the second half of the 15th century created strains. Of the 242 new apprentice painters between 1456 and 1490, 134, or 55 percent, were aspiring cloth painters, reversing a slight balance in favor of the panel painters at the start of that period. The second was that the illuminators began to make and sell single-leaf miniature paintings. Detached miniatures were new, and might have seemed an intrusion on the product space of the cloth and panel painters. The growth of apprentices among the cloth painters, meanwhile, must have made the panel painters uneasy.

Both developments at first produced blocking moves by the oil painters. They charged the illuminators in 1457 with importing detached miniatures. Probably this was mistaken – the miniaturists replied that they actually *exported*, and on a daily basis – though that would hardly have calmed fears of displacement among the panel painters.<sup>39</sup> Then, in 1459, the panel painters alleged of two cloth painters that they were displaying panels in oil, which was not permitted them, though nor might they display *any* paintings for

<sup>38</sup> Compare this with just 14 new inscriptions in the Company of St. Luke in Florence, 1492–1502, at the height of artistic activity in that city: Wackernagel (1981, p. 300, n.1).

<sup>39</sup> Campbell (1976, p. 190) records this incident.

sale, even of their own craft.<sup>40</sup> Initially the oil painters won support from the magistrates against the cloth painters, but in 1463 the magistrates reversed their own ruling, affirming that the cloth painters might display and sell in the open, “as they are used to doing”, on and near the St. John’s bridge, where ships tied up and unloaded, a step from the toll-house and close to the foreign “embassies” and the Exchange (*beurs*).<sup>41</sup>

These episodes involving the cloth painters, and the magistrates’ 1463 decision, tell us two things. Firstly, that the cloth painters were attempting actively to interest foreign captains and merchants in their wares; they were not content simply to hope that the foreigners would visit their workshops, as the regulations assumed. Secondly, the resulting deals must have grown to the point where to suppress open display might have harmed the city’s revenues from brokerage.<sup>42</sup> This is inference on our part, but something like that must have moved the magistrates. It is not known how large the traffic in cloth paintings was, but 17th-century records of Antwerp dealers show them purchasing cloth paintings made in nearby Mechelen by the hundreds and shipping them abroad on the same scale.<sup>43</sup> In this context it is probably significant, too, that, whereas the Medici in the 15th century had only four Flemish panels among their household movables, they had 38 cloth paintings of Flemish origin.<sup>44</sup> In the years 1461–1470, in Bruges, no fewer than 40 new apprentices attached themselves to cloth painters, while only 25 joined workshops of panel painters, a startling shift. Perhaps the magistrates detected early signs of this and went along with it, in the greater interest of the city.

Was price a factor in the improving fortunes of the illuminators and cloth painters? Undoubtedly. The average valuation of the Medici cloth paintings was a mere 3.38 florins, whereas that of their four Flemish panels was 23.5 florins. Again, Antwerp dealers in the 17th century paid around 1.6 guilders on average for cloth paintings, the equivalent of about 0.3 florins.<sup>45</sup> Interestingly, the price was about the same for 1 as for 100. For further comparison, a small sample of Flemish altarpieces commissioned by Italian buyers in the second half of the 15th century, averaged 23 to 42 florins per square meter.<sup>46</sup> Michelle O’Malley’s database of prices for Central Italian altarpieces generates a comparable range, though it also reveals that there were fewer altarpieces commissioned at higher prices. That relationship is not a demand curve, since each altarpiece was unique, but it is a price-sensitivity curve of form similar to a demand curve. Cloth paintings, in any case, clearly were very much cheaper, on average, than high quality panel paintings. The same must have been true with respect to the illuminators’ single-leaf miniatures. Matthijs Musson, one of the 17th-century Antwerp dealers

<sup>40</sup> *Ibid.*; and transcript of 1459 proceedings, kindly supplied by Peter Stabel.

<sup>41</sup> Martens (1998, p. 21).

<sup>42</sup> Unlicensed brokers were not unknown, but there is no hint that the cloth painters were actually closing deals themselves, to evade brokerage.

<sup>43</sup> See, for example, Duverger (1969, pp. 83 and 96) for purchases and a shipment by the dealer Musson.

<sup>44</sup> Nuttall (2004, Appendix 1).

<sup>45</sup> See Denucé (1931, pp. 28, 34, 63, 64, 73, 284, 289) for the dealer Forchondt; and Duverger (1969, pp. 83, 95, 96, 98, 102, 103) for the dealer Musson.

<sup>46</sup> De Marchi and Van Miegroet (2003).

just alluded to, obtained paintings on parchment for 0.04 guilders each, the equivalent of just 0.007 of a florin apiece.<sup>47</sup> If we were to translate the information just given for cloth paintings in Antwerp and for panels/altarpieces in 15th-century Florence and Bruges into price-quantity space they would reveal a horizontal demand curve for cloth paintings with a very low price intercept (in fact a narrow range, centered on 0.3 florin-equivalents), and a distinct price-sensitivity line covering a much higher price range. And were we to use the cheapest Medici panel (4 florins) as the low point of that price-sensitivity line, it would *start* at a point 13.3 times higher than the price intercept of demand for cloth paintings.

Such price differences had always been present. What really was new was competition in terms of products, supports, sizes, subjects and technique at the low end of the market. Here are a few instances. The illuminators, as noted, began selling single-leaf miniatures, painted on sheets of parchment or paper – a shift towards easel painting but still using their traditional skills in manuscript illumination. And they began making pocket-size portraits – as small as 6 × 9 cm. The panel painters responded to the latter by miniaturizing their own oil portraits on panel. Meanwhile, the panel painters freely imported the naturalistic landscape elements of illuminated Books of Hours, into their own narrative religious scenes. Cloth painters, for their part, began to create substitute panel paintings, attaching cloth to cheap panels. Some panel painters – Van Eyck, for example – also experimented with seccatives in their oil medium to shorten drying time; others, Memling among them, further reduced time to completion through a more economic use of paint: fewer paint layers, less glazing and an increased deployment of lead white for highlights to create surface effects.<sup>48</sup> Cloth painters, as noted, had long been responsible for making the cartoons for tapestries. In the case of the horizontal (*basse lisse*) looms commonly used in Flanders, the cartoons could be laid under the warp threads and the pattern directly transferred into the woven textile; but drawings for such cartoons could also be used to make series of painted substitute tapestries on linen. The substitute tapestry was a new product for the cloth painters, but techniques for transferring designs or motifs were employed by all three sorts of painter, saving costs for everyone, and benefiting the buyer, whichever kind of work was fancied. Many of these innovations could also be realized, at still lower cost, in the form of printed sheets, the technology for which was available in Bruges (and Venice) from mid-century, and was intensively applied in Antwerp in the sixteenth century.

These product innovations and cross-craft adaptations generated in Bruges an array of imagery not previously available and – importantly – affordable by those in the lowest income groups. The three sorts of painters found themselves, whether by intention or merely to retain or increase market share, tapping into a previously latent demand. This new market was distinct from the commission nexus: prices were much lower, but for the most part the products were not artistically derivative in the same way as were those

<sup>47</sup> Duverger (1969, p. 99, folio 14).

<sup>48</sup> Galassi (1997, p. 346).

of di Bicci in Florence. The former division between expensive, commissioned, panel paintings and cheaper cloth paintings of course remained, and cost-conscious buyers such as the Medici continued to purchase both sorts. At the same time, developments in Bruges marked the beginning of what would soon become a widespread habit: owning and displaying pictures of relatively low value, in considerable numbers, even in modest dwellings. This spread throughout the Low Countries, and in lesser degree elsewhere.

With the beginning of selling to the masses, new mechanisms for display and sale were called for. Before turning to some of them, however, it is worth dwelling for a moment on differences between Florence and one of the 17th-century northern seats of mass production and widespread ownership of paintings, Amsterdam.

### 2.5. *Wealth, and the ownership of paintings: Florence versus Amsterdam*

There was a comprehensive wealth assessment for tax purposes in Florence in 1427, and one in Amsterdam, in 1631; both imposed a tax rate of 0.5 percent. Unfortunately, no similarly comprehensive assessment exists for a contemporary Flemish city. For our purposes, however, a Florence–Amsterdam comparison serves. After converting the wealth and tax classes in Amsterdam from guilders into florin-equivalents, we obtain Figure 2. We adopt four tax ranges: class IV: <5 florins (<29 guilders); class III: 5–7.5 florins (29–44 guilders); class II: 7.5–50 florins (44–295 guilders); and class I: 50–580 florins (295–3422 guilders). In all 4110 persons were taxed in Amsterdam, and 6745 in Florence. Since the population ratio was c. 115,000 to 37,144, or 3 : 1, in favor of Amsterdam, it is clear that the tax base was much narrower there. The proportion of Amsterdam taxpayers in the highest three wealth categories, however, was greater than in Florence – 52 versus 24 percent – as were the absolute numbers: 2128 persons in Am-

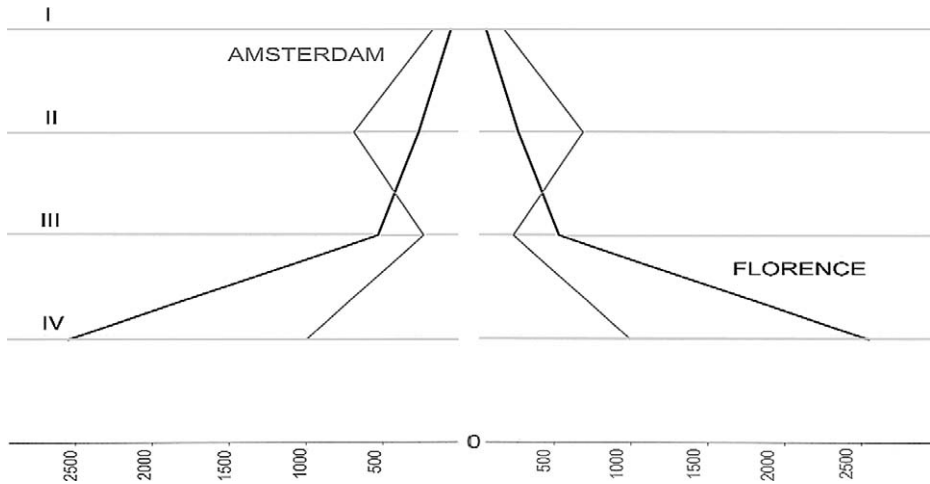


Figure 2. Numbers in wealth and tax classes (I–IV), Florence and Amsterdam.

sterdam, versus 1649. If we could also show that paintings in Amsterdam were cheaper, the combination of more persons of wealth and cheaper paintings would suggest that, other things being equal, there should have been more paintings owned by the top three *taxpaying* groups of households in Amsterdam than in Florence.

The data to check that expectation are lacking at the Florence end; but in any case the culture of filling the walls of a house with easel paintings was not one that developed at all strongly there. It is telling, however, to take as reference point the value of the *cheapest* panel in the Medici collection, an unattributed *Head of Christ*, at 4 florins, and compare it with the minimum average value of *attributed* paintings in Amsterdam inventories, 1620–1660, which was 9 guilders, or roughly 1.5 florins. This suggests a significant unit cost advantage in Amsterdam.

Did that translate into more actual purchases? Again, unfortunately, the question can be addressed comparatively only for cities north of the Alps. Table 2 gives the picture for Antwerp and Amsterdam, and for Paris and Metz. We do not possess ownership data for Bruges, or for Florence. However, taken together with the estimates given earlier for wealth-elasticities of demand for paintings in the Dutch Republic – all greater than 1 – Table 2 can be used to infer a pattern of increasing ownership over time in the Netherlands, north and south, both absolutely and relative to France. The pattern for the Netherlands thus establishes a standard of sorts against which to measure ownership in Florence and other Italian cities in future inquiries. It is perhaps worth noting that

Table 2  
Ownership of paintings in the Low Countries, and in France

	1530s/1540s	1565–1585	1620–1645	1630/1631	1645–1672
<i>Antwerp</i> (per house)	≈5.75 <sup>1</sup>	1–11 rooms: 6.8 <sup>2</sup> ≥12 rooms: 22.2		13.3 <sup>3</sup> 45.3	
<i>Amsterdam</i> (per inventory, by wealth-tax class): <sup>4</sup>					
Low end of lowest bracket:				8–15	
Upper end of lowest bracket:				9 originals +20 cheaper	
Top 2 brackets:				23 originals +23 cheaper	
<i>Metz</i> (per inventory) <sup>5</sup>					5.5
<i>Paris</i> (per inventory) <sup>6</sup>			8.7		

<sup>1</sup>Per house, judicial inventories, houses of all sizes: Martens and Peeters (2002, p. 881).

<sup>2</sup>Per house, various sorts of inventory: Vermeylen (2003, p. 148).

<sup>3</sup>Various inventories, excluding only houses of 16 or more rooms: Blondé (2002, pp. 382–383).

<sup>4</sup>Montias (2002a, pp. 122–123, Table 1). “Originals” were priced at the florin-equivalent of 2.3 and up.

<sup>5</sup>Benedict (1985, pp. 103, 105).

<sup>6</sup>Schnapper (2001, pp. 425–426), citing unpublished work by Mickaël Szanto.

the average of 23 original paintings per inventory in the very top two tax brackets in Amsterdam were valued at just under 6 florin-equivalents. The Medici – certainly in the top tax bracket in Florence – owned only 6 comparably cheap paintings, out of the 142 recorded in the 1492 family inventory.

Against this background of a rising ownership of paintings in the Netherlands – over time, and with increasing wealth – we will now trace some of the means by which sales were effected in numbers greater than would have been possible via direct sale from an artist’s workshop. For this purpose we shall turn, first, to the two cities Antwerp and Mechelen/Malines, which, in our terms (see [Appendix A](#)), comprised a net-exporter complex. Then we shall turn to Haarlem and Amsterdam in the province of Holland. Amsterdam, the dominant urban center – population c. 1660 of 200,000 – was a net importer of paintings and was linked with other cities via a canal network, which allowed that situation to continue. De Vries estimates that the population of other cities accessible within a day’s travel from Amsterdam was c. 350,000.<sup>49</sup>

### 3. Distributing paintings across markets

As production in volume became normal – in 16th-century Antwerp and Mechelen, 17th-century Holland and 17th-century Rome – institutions and mechanisms for the bulk marketing of images also developed, or were adapted for the purpose. These display an interesting variety of form as well as in the rules used to structure sales. In the case of Antwerp the city authorities played a central role, adapting to paintings and prints an older marketing institution – the display hall, used for textiles at fair times in Bruges.

#### 3.1. *The panden in 15th- and 16th-century Antwerp*

Antwerp in the 16th century took specialization and division of labor to new heights. The city benefited from natural and political forces that caused a decline in the Bruges economy during the 15th century. Antwerp’s river access, long restricted by sand bars, was improved thanks to freak floods in 1375–1376 and 1404, eliminating Bruges’ monopoly up to that point on traffic using larger sea vessels.<sup>50</sup> As to the political, as noted already, following a decision to oppose the centralizing moves and claims to dominion (and power to tax) over Flanders by the future Emperor Maximilian I, in the 1480s, Bruges was punished. Maximilian offered incentives and safe passage to foreign merchants to relocate to Antwerp, which had sustained his cause. After some hesitancy and a temporary return by some merchants in the early 1490s, by the 1620s all the foreign “nations” had moved permanently to Antwerp. In subsequent decades Antwerp

<sup>49</sup> De Vries (1981, p. 75, Table 3.7).

<sup>50</sup> Nicholas (1992, p. 390).

quite superseded Bruges as the world's *entrepôt*. This reorientation was greatly assisted by a monopoly in the distribution of spices granted to the city by the Portuguese at the start of the 16th century, plus a realignment of commercial traffic along land routes in accordance with the new mining wealth and burgeoning money markets of West-Central Europe. The adjustment occurred in several waves, between 1460 and 1540. Italian bankers suffered an eclipse in that region, and Venice became linked to Antwerp chiefly via land routes that ran through Cologne, Nürnberg and Vienna, as well through Cologne, Augsburg and the Tirol.<sup>51</sup> The distribution of Antwerp and Mechelen paintings in the 17th century can be traced partly along these routes, which complemented an older traffic – still maintained – to Paris, the Iberian Peninsular and the Americas.

The resettlement of the foreign “nations” to Antwerp contributed positively to the development of an art market there, in the same way as it had in Bruges; but there were additional elements that were unique to the Brabantine city. One was an extraordinary openness within the Antwerp St. Luke's guild, which was all-inclusive, coming to comprise, by early in the 16th century, some twenty crafts including painting, and as many as forty-eight by mid-century.<sup>52</sup> During that century dealers also begin to show up as registered members, though not all dealers bothered with guild membership.<sup>53</sup> More importantly, the guild did not forbid different sorts of masters cooperating with each other.<sup>54</sup> This all but eliminated the boundaries set according to technique and product that were officially espoused in Bruges, even if, as we have seen, practice could be less rigid.

A second distinctive feature in Antwerp was the deployment of cloister-like structures known as *panden*, some of them operating for the selling of paintings on a year-round basis. Strikingly, the city authorities took an active role in promoting these dedicated sales venues.<sup>55</sup> The first recorded *pand* involving paintings was in the cloister of the Dominican friary. It dates from 1445, and – unusually – was a combined initiative of the Antwerp jewelers' guild and the painters' guilds of Antwerp and Brussels. This *pand* lasted for more than a century, though in 1481, following a dispute, the painters switched to a new building near the cathedral church of Our Lady. There, from 1484, they enjoyed a monopoly, guaranteed by the city, on the sale of paintings during fairs. Our Lady's *pand* was also the first to function as a continuous or year-round sales site,<sup>56</sup> an interesting outgrowth of Antwerp's two annual fairs, in mid-May and late August,

<sup>51</sup> On these developments see Blanchard (1996).

<sup>52</sup> Martens and Peeters (2002).

<sup>53</sup> Vermeylen (2003, p. 134).

<sup>54</sup> *Ibid.*, p. 138. Martens and Peeters (2002) note that, contrary to belief, there were very few large workshops in Antwerp, and that many practitioners of other crafts who registered with the painters' guild also held membership in another guild. It would appear that within-workshop division of labor was relatively less important than openness and cooperation across crafts.

<sup>55</sup> Vermeylen (2003, p. 28).

<sup>56</sup> *Ibid.*, p. 26. The pioneering study of this *pand* is that by Ewing (1990).



each of which lasted for about six weeks. At those fairs luxury items of every sort could be had, including paintings.<sup>57</sup>

This last marks two other important differences from Bruges. It is unlikely that paintings were displayed in the open for sale at the Bruges fair. Guild regulations envisaged display and selling principally from shops; moreover, it is improbable that paintings were shown in the damp outdoors at the spring fair, unless perhaps during the three official viewing days.<sup>58</sup> In addition, though the city opened a *pand* in 1482, at the request of the jewelers' guild, it was not until 1511 that the restriction on painters just mentioned was lifted and they were allowed to participate in the annual *pand*.<sup>59</sup>

Various other *panden* were set up in Antwerp, for the display and sale of specialty products such as tapestries, but the last and greatest such enterprise for paintings was the one in the new Exchange that opened in 1540. Undertaken at the city's behest, this *pand* comprised 100 stalls on the upper floor of the Exchange building for the exclusive display and sale of paintings, prints, frames, etc.<sup>60</sup> One artist-dealer who rented a slightly larger-than-standard space in this artists' *pand* was Jan Van Kessel; at his death in 1581 he had a stock of 610 paintings there, exclusive of prints and various other things such as copper plates, paper, and chests for storing paintings.

Van Kessel's inventory gives us an opportunity to guess at the possible volume of production in Antwerp. If he was representative, and if we assume that his stock turned over completely every three years, there could have been something like 17,661 paintings added every year to the stocks displayed above the Exchange – at least in the good years up to the 1580s.<sup>61</sup> Most of these paintings would have been relatively standard, on canvas or cloth, and quite cheap. They would also have been new and anonymous;

<sup>57</sup> Campbell (1976, p. 196).

<sup>58</sup> Blockmans (1996, p. 22). This suggestion concerning display prior to the spring fair is not at odds with the fact discussed earlier that cloth painters found ways to display their paintings in the open, in the port area.

<sup>59</sup> Wilson (1983, p. 477 and Note 10).

<sup>60</sup> Vermeylen (2003, p. 50ff).

<sup>61</sup> For the information from which we have calculated Van Kessel's share of rented space in the *pand* see *ibid.*, pp. 60–61 and Appendix 2, p. 194. Van Kessel had space 1.152 times larger than the standard stall, so of the 100 stalls available, there could have been only 87 of Van Kessel's size.  $87 \times 203$  new paintings per stall-holder per year = 17,661 paintings. This is the crudest possible way to arrive at an estimate of total output. Montias has articulated a more sophisticated method: first calculate the number of artists in a city; then, using archival information on average yearly earnings for artists, plus average prices of paintings (from inventories), infer the weekly productivity needed to generate such an income. Applying this to Amsterdam, he arrives at a figure of one painting every three days (or 90 paintings in a year of 270 working days) for more expensive paintings (those averaging 13 guilders or florin-equivalents of 2.2 each and measuring c.  $63.5 \times 89$  cm), and 2.5 to 3.3 per week (or 135 to 180 a year) for cheaper paintings (5–9 guilders or 0.85–1.5 florin-equivalents apiece). Using Montias' estimated proportions for paintings in each category, the weighted average total for a year comes to 124 paintings. For an estimated artist population of 100, this would have meant a total output of 12,400 paintings per year, though with a maximum of 18,000 – assuming all paintings made were of the cheaper sort [Montias (2002a, pp. 117–120)]. We can check our crude calculation based on Van Kessel's inventory by applying elements of Montias' method. Thus, if we assume that the paintings sold at the exchange in Antwerp were mostly cheap, and employ the total number of registered painters in 1585–1586 – 108 – then total production at that point in time could have been  $108 \text{ artists} \times 180 \text{ paintings per year} = 19,440$ . In fact we are

indeed, the inventory of Van Kessel's stock for sale at the *pand* contained just one older work, "an old painting on canvas by . . . Bosch".

Guild openness, civic encouragement, widespread dealing,<sup>62</sup> plus specialization and division of labor practiced in the many crafts making up the painters' guild, and especially between masters' workshops, as well as a marketing and exporting orientation – all these features marked Antwerp as an environment for the production and sale of paintings the like of which had not been seen before. Moreover, Antwerp was complemented in the production of paintings by its near-neighbor Mechelen/Malines. Mechelen was a much smaller city, but it had been advantaged by the placement there during the rule of the Burgundian Charles the Bold (1467–1477) of a court of appeal (*parlement*), as well as of the Chamber of Accounts for all of Burgundian Flanders. The court was diminished under his successor and the Chamber returned to Lille. Charles' granddaughter Margaret of Austria, retained Mechelen as a seat till, in 1530, the court removed to Brussels. After a half-century of such losses Mechelen might have retreated into itself; instead a decision appears to have been taken to specialize in watercolor paintings on cloth, including cartoons for the tapestry industries in Brussels and, possibly, Oudenaarde, as an export business. The guild of painters reorganized in 1533 and was soon recruiting apprentices in extraordinary numbers for a city of maybe 30,000. Close to 80 apprentices were added between 1538 and 1568, and another 192 between 1598 and 1619 – a recovery after the re-capture of Antwerp by Spanish forces in 1585 and resulting emigration. The data on new master painters move similarly to, though they led those in Antwerp over the period 1568–1630.

The former of those two numbers for new apprentice painters in Mechelen – 80 – may be taken as evidence of the success of the Antwerp *pand*; it is known that paintings from Mechelen were sold by stallholders at the *pand* in the 16th century. The surge in

inclined to think this number, higher though it is than our own first guesstimate or Montias' calculated total for Amsterdam, is too low. The reason is that some paintings on cloth, made in Mechelen, were sold through the Antwerp *pand* – Van Kessel's records refer to such paintings (private communication from Filip Vermeylen), and Antwerp dealers are known to have constituted a major conduit for exports of paintings from Mechelen (cf. Footnote 43 above). Cloth paintings were certainly quickly made and very cheap, as we have seen – no more than 1.45 to 1.78 guilders apiece (weighted averages) to two 17th-century Antwerp dealers. A sales contract between a Mechelen painter and an Antwerp dealer in 1654 gives 1.8 guilders for a "large" and 1.25 for a "small" painting. We adopt these as proxies for more and less costly paintings. We also have a contract specifying the day wage of a painter, in 1634: 0.8 guilders. The relation of wages to prices remained roughly constant even though the levels of both changed between the mid-16th century and the early 17th. Assuming these contracts were representative, and applying both day wage and prices to the 1560s – Mechelen's zenith as a center for paintings production – an artist there could produce 103 more costly or 148 cheaper paintings in a year of 38.5 six-day weeks. We estimate that there were 159 painters in the city in the 1560s, yielding total annual output in the range 16,377 (more costly) to 23,532 (cheaper) paintings. There remained something like 69 painters in 1632. For documents see Van Autenboer (1949) and Monballieu (1971). A conservative annual average total production for Mechelen and Antwerp combined therefore might be 30,000, and more in years when the combined artist population exceeded 200, as it probably did at times.

<sup>62</sup> Vermeylen (2003, pp. 66–67 and Appendix 3), counts 17 persons registered in the guild as specialist dealers in the sixteenth century plus 24 more so named in other (non-guild) documents.

numbers early in the 17th century, on the other hand, marks a change in the style of marketing. The *pand* was then no more, but by the 1620s its place was being taken by international dealers such as Chrisostomo Van Immerseel and the brothers Goetkint, and presently by firms such as the partnership of Matthijs Musson and Maria Fourmenois, and the family business of Willem Forchondt and his sons. Records of the latter two firms, and letters of Van Immerseel, attest to the hundreds of Mechelen paintings they bought, for re-export.

The economic situation in Antwerp had begun to decline by the 1580s. First, in 1576 there was a destructive riot by unpaid Spanish troops garrisoned in the city. Then, in 1583, a fire destroyed the *beurs* and with it the artists' *pand*. Though rebuilt, the new exchange was finished only in May, 1584, by which time Antwerp (at the time Calvinist) was under siege and merchants had begun to withdraw from the city. The steady decline of the *pand* from this date,<sup>63</sup> combined with the huge emigration – 50 percent – in the four years after the city capitulated to Spanish forces in 1585, including some hundreds of merchants who moved to Amsterdam,<sup>64</sup> obliged artists and artist-dealers to look to external markets. And while new, rebuilt, or refurbished churches and religious orders in the period from about 1590 through 1625 created a fresh local demand for sculpture and painting, the subsequent slowdown renewed the pressure to seek sales abroad. We shall consider here incursions made by Antwerp dealers into Paris and Holland.

### 3.2. Dealer-led incursions into Paris, and to Dutch cities

Given the circumstances in Antwerp after its re-capture by the Spanish in 1585 we would expect the number of new apprentices to have fallen off, or artists to have switched professions, or to have emigrated. All those things occurred, yet artists there and in Mechelen quickly regrouped, first restoring and thereafter sustaining a production capability in paintings well in excess of the needs of local buyers. A tradition of family artist dynasties possibly slowed the adjustment process in both places, while the cooperative guild tradition in Antwerp also contributed towards a flow of novel products even in the face of, or possibly as a response to, the slowing of demand.<sup>65</sup> However, the dominant response was to turn more aggressively to dealing abroad, maintaining Antwerp and Mechelen as production bases.

Paris was one target of this aggression. There is evidence that Antwerp dealers visited the internationally-important pre-lenten fair of Saint-Germain-des-Prés, held just outside Paris, in the late 16th century. By early in the 17th century Antwerp dealers had become major stallholders at the fair. This also gave them influence in the Paris market itself, in part because they began leaving stock not sold during the fair with associates

<sup>63</sup> Annual rents tell the story: see Vermeulen (2003, p. 54, graph 3).

<sup>64</sup> Gelderblom (2000, Appendix 4), records some 285 in the decade from 1585 and many more in the subsequent decades up to 1630.

<sup>65</sup> For discussion of some of these see De Marchi, Van Miegroet and Raiff (1998).

in Paris for subsequent sale, in contravention of regulations.<sup>66</sup> The annoyed response of the corporation of painters in Paris, the *maîtrise*, was to request, in 1619, that foreigners be excluded from trading, except at the fair, and that, even there, the paintings they offered be subject to assessment by *jurés* of the *maîtrise*. There is evidence of fines being imposed on foreign dealers if their paintings were deemed inappropriate. Independently of this deterrent, dealers from Antwerp, by the late 1620s, feared sequestration of their goods should war break out between France and Spain, and they began to switch from owning extended leases at the fair to renting stalls on an annual basis. Several also pursued naturalization or other means that would allow them to settle legitimately in Paris.<sup>67</sup> Foreigners never were excluded from the fair, but it remained difficult for them to obtain permissions to establish in Paris. The corporation of painters, sustained in the crucial early decades of the century by the magistracy, was strong enough to continue to threaten and to restrict entry by outsiders. Antwerp dealers, with few exceptions, were confined to visiting during the fair, though some who did settle continued to be supplied from Antwerp. We shall see below that, in the case of Lille, where, in the third quarter of the 17th century, there were also almost annual incursions, the numbers, relative to the local population of artists, were only slightly more alarming than in Paris: 10 dealers relative to maybe 40 guild artists in Lille; perhaps 15 Antwerp dealers compared an estimated painter population in Paris (in 1611) of 75.<sup>68</sup> Tellingly, however, in Lille the local magistrates withheld their support from the painters (see Section 5.3 below).

There came an interesting reversal in the way the one-way flow of paintings from Antwerp to Paris throughout the 17th century was managed. That traffic, dominated by Antwerp dealers in the early decades of the century, is reflected in significant Netherlandish holdings in many late 17th- and early 18th-century Parisian collections.<sup>69</sup> Throughout the 17th century, the marketing of paintings in Paris was more or less controlled by painters, French-born or foreign, including some from Antwerp. Early in the 18th century, however, some trader-merciers (*marchands-merciers*) in Paris began to ease control of the market away from artists. They took the initiative in bringing Netherlandish paintings to Paris themselves. One such merchant-mercier was Edme-François Gersaint, but he was followed by others, among them Alexandre-Joseph Paillet and Jean-Baptiste Pierre Le Brun; all three went regularly to Flanders and Holland to view collections and buy at auction there, for resale in Paris.

Returning to the 17th century, Antwerp dealers also tried smaller-scale incursions into the cities of Holland – Dordrecht, Delft, Gouda, Leiden, Amsterdam. This continued

<sup>66</sup> Szanto (2002; 2006).

<sup>67</sup> Szanto (2001).

<sup>68</sup> The number of visitors in Paris is based on the size of a group of 14 in 1628 [for which see Szanto (2006)]; the figure 10 for Lille is an average drawn from signatories to requests to hold sales in the years 1677–1688. The Paris artist population is from Schnapper (2001), taking the ratio of painters to sculptors as two to one, in a combined population – painters and sculptors – of 100–115 in 1611; that for Lille is based on the guild's 1679 complaint that 35–40 families had been ruined by the incursions: see De Marchi and Van Miegroet (2005c).

<sup>69</sup> Schnapper (1994); see also Van Miegroet (2005).

for a period of maybe three decades or so, starting very soon after 1585. In this case, as in Paris, the local painters protested. After initial attempts to address the problem by obtaining ordinances against the interlopers, however, local artists' guilds and informal groups of painters gradually began to act more like Gersaint and his ilk, conducting sales under their own auspices. In Amsterdam the local guild received only tepid support from their magistrates, probably because the dominant buyers of paintings were merchants, many of them immigrants from Antwerp, and the city's commercial prosperity was more importantly served by them than by its painters.<sup>70</sup> In this broad respect, and in related details, the guild misconstrued the changes underway in Amsterdam: from a static environment in which guilds could expect to safeguard the income of members by protecting their market against foreign intrusions, to an open, dynamic world of changing ideas, products, processes and tastes.<sup>71</sup>

The Amsterdam guild alleged that the southern dealers brought with them, and sold in illegal auctions, only inferior paintings – mere copies and student work – yet represented them as originals.<sup>72</sup> Local buyers, they added, being unable to tell the difference, were being duped. By implication, local artists would suffer, because, as buyers became skeptical, they would hold back and the market would collapse, as per Akerlof's "lemons" model.<sup>73</sup>

This did not happen; but nor does the evidence support the key charge, that buyers were duped. Montias has shown for contemporary, legal auctions – those held by the Amsterdam Orphan Chamber – that, in the years 1597 to 1619, 54 percent of known buyers were themselves immigrant southerners.<sup>74</sup> Of known buyers, too, one in three was a merchant, 60 percent of them of southern origin.<sup>75</sup> It is highly likely, therefore, that the majority of buyers at the illegal sales organized by the interlopers were also immigrants from the south, happy to purchase paintings of a sort with which they were very familiar. Far from these buyers being undiscerning, then, their origin and their enthusiastic participation in the official Orphan Chamber auctions suggests that they must have known exactly what they were getting. The demand in Amsterdam, we may infer, was segmented: knowing southerners, versus allegedly undiscerning locals. This seriously restricts the possible scope of application of the guild's claim that the market would collapse. In addition, since the allegedly undiscerning could observe what

<sup>70</sup> Gelderblom (2000, p. 227) shows that merchants of southern origin in Amsterdam were wealthy, if less so than local merchants.

<sup>71</sup> It is worth noting that, despite the Amsterdam guild's public face, cost-saving techniques pioneered in the Southern Netherlands made their way in the next few decades into Dutch cities, via Amsterdam: see Montias (1986) and Sluijter (2000).

<sup>72</sup> See De Marchi (1995) for references to original sources.

<sup>73</sup> Akerlof (1970). De Marchi (1995) applies Akerlof's model to the Amsterdam situation on the assumption that the guild's arguments were well-founded. Montias' research (see below) has since provided an empirical basis for concluding that they were not.

<sup>74</sup> Montias (2002b, p. 63).

<sup>75</sup> Ibid. and p. 47.

knowledgeable buyers were willing to pay at auction, it is simply implausible that the asymmetrical information driving Akerlof's model was present.

As noted, gradually guilds in other Dutch cities began to explore the possibility of benefiting by public sales. In fact from the 1640s guilds and informal organizations of painters in Leiden, The Hague and Haarlem became active in setting up display-facilities for their members and/or in holding regular public sales. The external threat of the first decades of the century was tamed in Holland by the artists' organizations accepting public sales and bringing them under their own auspices.<sup>76</sup> In the 1680s, however, there was a new threat from Antwerp dealers, directed to the south. Lille (and neighboring cities) was the prime target, but a group of dealers, including many of those who traveled to Lille in the same period, also appeared in Ghent, with the same request: that they be allowed to hold auctions outside the times of the fair. In Ghent, however, the painters held more power than in Amsterdam, and their magistrates denied the request.<sup>77</sup> Again, the Lille experience will be addressed below in Section 5.3.

It is worth inserting here that for all its openness the Antwerp painters' guild reacted just as negatively as guilds elsewhere when foreign dealers came to Antwerp and sold outside the twice-yearly fairs and the weekly "free" Friday market-day. The language used by the Amsterdam guild in its first protest, of 1608, could have come from a 1575 request by the Antwerp guild that an older statute proscribing dealing in any form by non-guild members be re-affirmed. Similar language was retained in later revisions of the statute. Officially, then, openness in Antwerp was limited. The practice, however, was very flexible. Close to one third of known Antwerp dealers in the 16th century were not registered with the guild, yet prosecutions were rare; non-registered dealers held office in the guild, and they were among the stall-holders at the city-backed *beurs pand*.<sup>78</sup>

### 3.3. *The market in Holland as a network*

In recent years ingenious estimates have appeared of the output of paintings in the Dutch Republic during its decades of growth, roughly 1590 through the 1660s, and for Amsterdam separately. If we extend the period to a full century, the total for the whole Republic, in the period 1580–1680, is put at 4.7 to 5.3 million paintings.<sup>79</sup>

Those numbers are impressively large, though in relative terms not out of line with our estimates for 16th-century Antwerp–Mechelen, which rest on comparable productivity per artist.<sup>80</sup> Yet, throughout the period, Amsterdam was also an importer of

<sup>76</sup> See Romein and Korevaar (2006) for details.

<sup>77</sup> Raux (2004).

<sup>78</sup> See Vermeulen (2003, pp. 66–67 and 131–136), and Appendix 5 for the 1575 document.

<sup>79</sup> Two very different methods have been used. One is due to Van der Woude (1991), whereby the number of paintings is inferred backwards from surviving paintings and estimated rates of depreciation; then the number of active painters is calculated for particular years. Finally the productivity necessary to connect those two sets of estimates is derived. The other is that of Montias, for which see Footnote 61 above, and Montias (1990, p. 70).

<sup>80</sup> See the discussion in Footnote 61 above.

paintings (attributed and unattributed, that is, originals and copies), to the tune of perhaps 20 percent of its demand.<sup>81</sup> How was the lack supplied?

In a sample of Amsterdam inventories covering the period 1620–1679, and attending only to original paintings, Montias discovered that just 35 percent of these were attributed to artists who worked solely in Amsterdam. The percentage rises to 53 if artists are included who worked in Amsterdam *and* other places.<sup>82</sup> The percentage of Haarlem artists in Haarlem inventories, however, was much higher, 74.<sup>83</sup> But whereas tiny fractions of holdings in Haarlem came from other cities in the Republic, 14 percent of holdings in Amsterdam came from artists residing solely in Haarlem (24 percent if artists who worked in Haarlem and elsewhere are included), and another 12 percent from Antwerp. An additional 8 percent came from artists who worked only in Utrecht.<sup>84</sup>

How did Haarlem come to supply such a solid portion of Amsterdam's net imports of paintings? Boers has shown that the supply of paintings in Haarlem began to outgrow demand there around 1625,<sup>85</sup> and that two complementary changes occurred to right the imbalance. First, dealers stepped in: 12 were recorded in Haarlem in 1634, two of them women, in an artist population of only 90. Several of those dealers either originated from elsewhere (two from the southern Netherlands) or were actually based in another city (two in Rotterdam). These dealers took off surplus paintings, reselling them in markets such as Rotterdam and Amsterdam. Second, several Haarlem artists migrated to nearby Amsterdam.

Montias' basic finding implies that some 65 percent of attributed originals in Amsterdam inventories were made by out-of-town artists. There is a pattern to where the paintings came from. It correlates loosely with the system of canals built from the late 1620s onwards, in two periods of investment. The canals were used for slow but reliable passenger-carrying, horse-drawn barges (the *trekvaart* system).<sup>86</sup> There was an earlier, strictly local market-boat system, and an inter-city system as well, but the latter involved circuitous routes imposed by local water-toll authorities, who restricted travel possibilities so as to maximize their receipts.<sup>87</sup> The *trekvaart* system was the first water-based passenger service between cities using canals that followed the most direct routes. The Utrecht–Amsterdam link was in place by the late 1620s, and in Amsterdam inventories from the next two decades, 1630–1649, Utrecht painters account for 11 percent of the attributed originals (more than a third above the 8 percent average for Utrecht over the

<sup>81</sup> Montias (1985; 2002a, pp. 127–128).

<sup>82</sup> Montias (1991, Table 8).

<sup>83</sup> Montias (1985, Table 8).

<sup>84</sup> Montias (1991, Table 8).

<sup>85</sup> Boers (2000). Painters in Haarlem as a percentage of the city's population grew from 0.05 in 1605 to 0.13 in 1625, and 0.21 in 1635: figures from Sluijter (2000, p. 137), note 19 citing Boers estimates of the number of painters, including artists whose name appears in archival records but for whom there is no known surviving work.

<sup>86</sup> De Vries (1981, pp. 27ff).

<sup>87</sup> *Ibid.*, p. 19.



whole period 1620–1679). The Haarlem–Amsterdam link – just twelve miles long – came into service in 1632 and in the succeeding three decades 16 percent of holdings in Amsterdam were accounted for by Haarlem artists who worked only in that city (above its 14 percent average for the whole period) – a one-seventh difference, less dramatic than for Utrecht, possibly because the distance by road was so small to begin with. Cities such as Leiden, Rotterdam, Delft and The Hague, for many years without direct canal links to Amsterdam, continued to account for tiny percentages of the attributed originals in Amsterdam inventories.<sup>88</sup>

This is weak evidence of a network of paintings flows facilitated by a transportation infrastructure. It has been put to us that unframed paintings are low bulk, high value-added goods, so that freight costs cannot have been responsible for the geographical pattern of production. But our argument is not that; it is that local specializations among artists emerged for a variety of non-economic reasons, and were reinforced by local guild membership fees, which everywhere favored the sons of local masters. With an efficient transport infrastructure, local surpluses in specialist paintings could be redistributed, to align supply more closely with demand without change being needed in where production occurred. In other words, and although some artists did migrate, the *trekvaart* system made it quicker and easier for arbitraging dealers to function – as they did in Haarlem immediately after the Haarlem–Amsterdam barge link was forged – instead of forcing artists themselves to move. Even a short trek like that between Haarlem and Amsterdam was eased – made quicker, more reliable, and less dangerous – by the barge service, since it supplanted a route that required passage over a stretch of sometimes dangerous open water and via an equally squally river.

The canal network complemented and assisted the dealers, while the dealers facilitated collecting by individuals. Dealers and the canal network together enabled Amsterdam to sustain a paintings trade imbalance throughout its “golden” age. The network also permitted local specialization in paintings to occur in many small cities – too small for efficient specialization, considering only local demand. Thus it fostered collective diversity with excellence by supplementing local demand to the point where specializing was feasible even among small populations of artists and buyers.

### 3.4. *Trading in paintings: The traffic from Antwerp to New Spain, via Seville, c. 1540–1670*

Finally, in our set of illustrations of dealing to supplement inadequate local demand, we look at traffic in paintings from Antwerp, through Seville, to the Americas. Such a trade existed from at least the second half of the 16th century. This trade has not been systematically studied, but according to Falomir it passed through three stages.<sup>89</sup> These chart a transition from a commission nexus through artist–dealers to specialist,

<sup>88</sup> Montias (1991, Table 8).

<sup>89</sup> Falomir (2006).



professional dealers, much like the progression encountered in Antwerp – perhaps not surprising, since the traffic for many years was dominated by merchant–dealers based there. The stages are as follows:

- (1) Prior to the 1540s paintings for New Spain and other parts of Spanish America were commissioned directly from Spanish artists by the *Casa de Contratación* and the finished works shipped to the viceroalties.
- (2) From about the mid-sixteenth century painters began to take the initiative. They also assumed the risks, producing paintings that had not been commissioned which they then shipped through intermediaries, in many cases ships' captains, with instructions to sell as well as possible in the ports of America and return the proceeds, for a percentage.
- (3) Though intermediaries continued to be used, the potential profits on the traffic with America – prices there might be 3 times higher than those in Spain – increasingly drew in professional merchants, who shipped paintings and prints along with *escritorios* (writing desks), cabinets with doors and tiny drawers to which small paintings on copper were affixed, painted keyboard instruments, books, textiles (including tapestries) and clothing.<sup>90</sup>

Prominent among these professionals were Flemish merchants. Some stayed in Antwerp and filled orders for paintings received via agents abroad, receiving payment prior to or not long after shipment. Others, such as Chrisostomo Van Immerseel, were more vertically integrated; he initiated commissions in Antwerp himself – though in response to advice from his wife, Maria Fourmestraux, in Seville – and the couple carried the risk themselves.<sup>91</sup> If final sale was to occur in America, paintings (and other goods) would be shipped to Seville, and later Cadiz, for forwarding to “the Indies”. Shipments across the Atlantic were sometimes prepared to order, sometimes on spec (without a known client or clients in view). Limited evidence relating to the latter suggests that shipments were deliberately mixed as to supports, sizes, subjects and prices, as if the merchants involved were consciously spreading their risks so as to achieve on average, as with an investment portfolio, a target rate of return.<sup>92</sup>

How large was the traffic in paintings from Antwerp to America via Seville? Falomir suggests that it may have peaked just before or around 1600, which would roughly coincide with the peak in the tonnage of shipping between Spain and New Spain.<sup>93</sup> Nonetheless, though with ups and downs, the average shipments per year even during the second half of the 17th century remained substantial. Between 1583 and 1599, on average, 144 paintings each year were exported, with occasional exceptional numbers:

<sup>90</sup> *Idem.*, citing Páramo (1999) and Echeverría (2000).

<sup>91</sup> Van Immerseel started in his father's trading business, which included trading in tapestries. Typically, in that industry, merchants supplied capital to sustain the workers during the long preparation process, as well as handling orders and distribution/sale. It would have been natural for him to treat his trade in paintings in a similar way.

<sup>92</sup> De Marchi and Van Miegroet (2000a).

<sup>93</sup> Lynch (1969, p. 283, Table B).

e.g., some 637 shipped in a single three-month period in 1586. Between 1651 and 1700 the numbers ranged from 150 to 910 per year.<sup>94</sup> Yet the highest number, 637 in three months, or 2185 on an annual basis, is still only 7.3 percent of the very conservative estimate we offered above (note 61) for annual production in Antwerp/Mechelen combined, of 30,000 paintings per year. We should note that orders were also placed in Italian cities – Naples and Rome – and in France and Madrid, not just in Antwerp, though on those flows there is only scattered evidence.<sup>95</sup>

Taking just exports from Antwerp and Mechelen, the Americas trade shows this production complex in a new light. The professional merchants involved were not dealers making forays into markets a day or two distant, but integrated international traders in paintings, involved in production, yet obliged also to be sensitive to buyers' wishes. In having to translate preferences expressed in Spanish into the Flemish of the artists who would fill the orders, dealers found themselves mediating not just between different language communities but between distinct production and visual traditions. That is what the international agents of patrons and collectors had always done, only now it was being accomplished in bulk, not item by item.

#### **4. Selling paintings I: Artist–dealer relations**

We move now to some less neatly-packageable aspects of markets in paintings, involving relationships between artists and dealers or dealers and collectors. They are included because our emphasis on markets as experimental plays leads us to think that a certain degree of messiness was normal and should not be shunned; we also value the variety of routes taken as markets moved towards maturity. We shall look first at artists' pay and how it was shaped by traders who viewed original paintings they commissioned as capital, capable of generating an income, via copies. Then we turn to the less studied situations of dealers and how they related to artists in Venice and Rome.

##### *4.1. Securing payment for invention*

One aspect of the Van Immerseel–Fourmestraux vertically-integrated trading business enabled a change in the way artists were paid, in Antwerp certainly, and perhaps elsewhere. Painters in 15th-century Florence were regarded as skilled artisans and paid accordingly. To illustrate, we noted that Neri di Bicci showed up in the tax assessment of 1480 as the wealthiest painter in Florence. Goldthwaite points out, however, that this would have left him “entirely within the ranks of artisans and shopkeepers as delineated

<sup>94</sup> García Fuentes (1988, p. 58); Kinkad (1984, p. 305); Echeverría (2000). These numbers all depend on how many paintings it is assumed went to make up a “roll”, which was the unit of record. We have followed Falomir in assuming 100 paintings per roll.

<sup>95</sup> Falomir (2006).

by the more authoritative *Catasto* [tax assessment] of 1427".<sup>96</sup> Day-wage equivalents for painters in Florence are extremely hard to come by, but even the highest salary in the construction industry, paid to foreman of the cathedral Brunelleschi, amounted only to something like 1.5 to 1.75 times the de facto average rate for skilled artisans.<sup>97</sup> Little changed in Florence over the course of the 15th century where, as noted earlier, artists' services seem to have verged always on being in over-supply.

In terms of rewards for artists relative to those of other artisans, the situation in late-14th-century Bruges was similar. A partial survey of Bruges tax lists for 1394–1396 records no painters in the two highest tax categories; instead, painters and saddlers – members of the same guild – appear in the lowest or second lowest tax bracket, a biased distribution paralleling the income structure for the population as a whole.<sup>98</sup> Over the course of the 15th century the picture probably altered little in Flanders.<sup>99</sup>

Shift forward to early 17th-century Antwerp, however, and there we find reproducibility being exploited to sustain payments for artists' services in excess of the rates paid other skilled artisans. Jan II Brueghel, grandson of the famous Pieter I Brueghel, and a second-tier painter, claimed payment for his invention. In itself this was not novel. Invention – *ingegno* – was expected by patrons in 15th-century Florence but often not paid for.<sup>100</sup> Jan II Brueghel, however, priced original paintings by themselves at roughly twice the cost of a copy of the same by his own hand. Assuming that size, support, and medium were the same, only invention in the original could account for the price difference. But how could a second-level painter make such a claim to payment for invention stick? The circumstances are somewhat special, though perhaps telling. Brueghel supplied paintings to Van Immerseel, who recorded at one point that he regarded the originals as capital assets. As such, they were not to be sold – Van Immerseel likened them to a shoemaker's last – but rather to be used to generate a (limited) number of copies.<sup>101</sup> The creator of such income-generating capital assets could easily be paid something extra when it was averaged over the entire stream. This was impossible within the commission nexus in Florence, with its one-off projects. Neither was

<sup>96</sup> Goldthwaite (2003, p. 437).

<sup>97</sup> *Ibid.*, pp. 318, 321.

<sup>98</sup> Blockmans (1995, p. 12).

<sup>99</sup> A large decorative project at Lille, in 1454, for the Burgundian Court, employed 38 painters at day rates. The rates for ordinary free masters on this project were one third above those for skilled artisans generally: see Martens (1999, Table 1). However, this is not conclusive evidence of an improvement in the situation of Flemish artists precisely because of the Court origin of the commission.

<sup>100</sup> Securing payment was always an uncertain business [Thomas (1995, p. 183)], but *ingegno* introduced special difficulties. This was because patrons – rightly, in many cases – perceived the originating idea as theirs. Moreover, patrons might even claim, as the poet Carbone put it, that “the merits of princes are reflected in the talents of their subjects”, hence that the talent of an artist is just a channel for transmitting their own worth and merits no special reward. See Syson and Thornton (2001, pp. 135, 142–143) for the example of the patron Isabella d'Este, and p. 156 for the case of the artist Cossa, who complained that he was paid as a journeyman.

<sup>101</sup> For details see De Marchi and Van Miegroet (1996).

it present in 15th-century Bruges, where artists sold directly out of their workshops, keeping only motifs, and drawings of heads and hands, to be copied.

As if in acknowledgement of the potential involved in arrangements such as those set up by Van Immerseel, however, the Flemish and Dutch word for an original painting of some quality from at least the late 16th century onwards was *principael*, in English principal, a capital sum, though the term applied to paintings primarily meant first-class. Dealers in 17th-century Rome also retained originals for copying, but it is unclear whether they treated the income stream from copies as a fund for paying artists for their invention.<sup>102</sup>

In Antwerp the implied mechanism took. Brueghel was not alone among his contemporaries in securing payment for invention, while there is fragmentary evidence there from later in the 17th century that puts the day-rate-equivalent of common master painters at between two and four times that of skilled artisans.<sup>103</sup> It is possible that the special conditions which allowed Jan II Brueghel and others successfully to claim such a differential became conventional during the ensuing period because numerous artists found themselves working for vertically integrated trader-dealers. Montias suggests that a comparable differential was enjoyed by artists in 17th-century Holland, where the putting-out system was also practiced.<sup>104</sup> But putting out could also result in slave wages, as seems to have been common in 18th-century Venice (see Section 4.2 below). A lot depended on local circumstances, and in Flanders and Brabant there was a tradition of viewing as capital cartoons for tapestries and engraved copper plates in the print industry. An 18th-century English example combines both painting and printing: the portraitist Joshua Reynolds calculated the value of a painting made for the publisher Boydell in terms of the revenue expected from the print run.<sup>105</sup>

#### 4.2. Selling paintings in Venice and Rome

Late in the 16th century, in Venice, a second-hand dealer (*strazzarolo* or *rigattiero*) was charged with stocking some new paintings. A few years later one of the city's official auctioneers was charged with having new paintings displayed above his stand, presumably to sell on own account. Of the second-hand dealer it was said that he actually had an arrangement with painters, something additional documents show to have been not uncommon.<sup>106</sup> Such arrangements were illegal in that only artists were supposed to sell in the domestic market – non-artists could buy for resale abroad.

<sup>102</sup> Lorizzo, private communication.

<sup>103</sup> De Marchi and Van Miegroet (1996, pp. 60–61).

<sup>104</sup> Montias (1990, p. 64ff; 2002a, p. 119; 2001, n. 3). Montias' generalization is not consistent with all available data; see, e.g., Van Zanden (1993, Table 3.1).

<sup>105</sup> The example of the painter Sir Joshua Reynolds is particularly striking. For his negotiations with the enterprising print publisher Boydell see Zablotney (1999). William Hogarth presumably made similar calculations in arriving at the subscription price to charge for prints after his paintings.

<sup>106</sup> Favaro (1975, pp. 71–72); Cecchini (2000, p. 194). Matthew (2006) details the use of such arrangements by the 16th-century artist Lorenzo Lotto.

The rationale used in Venice for insisting that new paintings be sold in the domestic market only by their maker or by another guild-registered artist had two parts to it. On the one hand, it was feared that, without such a restriction, unscrupulous individuals, including unregistered masters, or even non-masters, might garner a share of the market properly belonging to legitimate guild masters. On the other hand, if this occurred it would undoubtedly involve means that could not be countenanced – the employing of non-registered artists, and apprentices, probably in hidden locations, to turn out works of dubious quality.<sup>107</sup> Unfortunately the restriction took no account of a market that was expanding – as it was in 16th- and 17th-century Venice – and in which the specialized services of intermediaries might be needed to allow painters to concentrate on what most of them did best, making pictures.

Venetian artists had long enjoyed the freedom to make paintings other than those commissioned, and, during the 15th century, additional concessions were granted affecting both display and sale.<sup>108</sup> Local and mainland artists could sell at weekly markets in the quarters of San Polo and San Marco; moreover, as elsewhere in Europe, even foreigners were free to sell during the *Sensa* fair. Yet, as late as 1513, the prohibition against selling by others than artists was repeated, and even a century later, in 1607, the same proscription was renewed.<sup>109</sup> While the earlier concessions might be read as attempts to adjust to a growing demand for Venetian paintings, the guild, supported by the civic authorities, by continuing to insist that only registered artists should sell, actually heightened the potential for conflict in the way artists' allocated their time.

That conflict, however, was just one effect of the ban. Proscribing selling by non-artists – exports excepted – invited evasion, confirmed by the examples cited at the outset of this section. At the same time, to the extent that the ban was respected and enforced, it obviously favored exporting. In the seventeenth century both local and foreign demand was strong. A repayment of public loans from 1577 injected a large amount of spending power, which fed an emerging preference to put pictures on the walls of houses. Through the mid-seventeenth century paintings large and small came to adorn Venetian houses of every sort.<sup>110</sup> At the same time, however, the population of Venice grew, but only fitfully; by the late seventeenth century, and in the eighteenth, the market was increasingly driven by foreign demand. The problem of artists having to divide their time and intellectual capital between making and selling was masked while demand grew strongly. Nevertheless, it did put artists at a disadvantage in the export market, compared to merchants and specialized dealers in paintings, both of whom had the advantage in experience, networks of contacts with foreigners, and access to capital. This meant that, when the growth in local demand slowed relative to export demand,

<sup>107</sup> Cecchini (2000, p. 193). We find this rationale spelled out only in a document from 1638, but it seems reasonable to assume that it also informed earlier constraints on selling.

<sup>108</sup> Ibid., p. 192; Favaro (1975, p. 73).

<sup>109</sup> Cecchini (2000, p. 192).

<sup>110</sup> Cecchini (1998) and Borean (1998).

Venetian artists could not just turn to the export trade and expect to compete there as dealers on equal terms.

The result was that, by the early decades of the eighteenth century, the business of trading paintings had fallen significantly under the control of merchants and specialized dealers, and artists, in many instances, began to find themselves working, not directly for the market, but to fill orders placed by these traders, as if “in the galley”. This last is backward inference. A tax list of the College of Artists from 1719 shows that the dealers who belonged to the College were much wealthier on average than the artists. Those among dealers who paid the minimum tax still paid more than did two-thirds of the artists.<sup>111</sup> The situation appears not to have improved in subsequent decades: a document of 1757 reported that many painters worked for dealers, producing “quadri dozenali”, an expression meaning shoddy work but also encompassing the idea of serial painting. The latter sort of production is exemplified by a shipment by an eighteenth-century merchant, to Spain, of paintings that included a series depicting saints, and by a sale in 1751 by a dealer to a visiting German nobleman of 49 readymade portrait and cityscape paintings all within a week, many of the same dimensions, and all at prices between 5 and 6 ducats.<sup>112</sup> Dealers, the 1757 report alleged, often paid despicable amounts (“prezzi vilissimi”) for “quadri dozenali”, but sold them on at prices that were much higher (“assai migliori”).<sup>113</sup>

This view of things is further sustained by the fact that specialist dealers are first mentioned in the late seventeenth century, in records of the new College of Artists, which admitted dealers who were trained as artists, though not on an equal basis.<sup>114</sup> That is very late, compared to, say, Antwerp, where the rule that marketing be in the hands of artists (except for exports), as we have seen, was conveniently ignored, at least during the good years of the mid-sixteenth century. But this does not define the extent of the difference. We find the artist–dealers who were members of the College in Venice, as late as 1769, caught in a sudden lurch toward greater strictness: the College insisted that from that moment the dealers should no longer register with them but with the guild of furniture painters, since they had shown that they did not know how to use the brush.<sup>115</sup>

Why were the painters of Venice, supported by the city authorities, bent on maintaining a policy towards the marketing of paintings that was at odds with their own long-term interest? Any answer must be speculative; Cecchini suggests, however, that there might have been less deliberation than inadvertence involved. The other luxury trades – Turkish cotton, Persian silk, spices, and domestic manufactured and finished goods such as textiles, glass and lace – on which Venice’s wealth rested much more

<sup>111</sup> Montecuccoli degli Erri (2003, pp. 145–146).

<sup>112</sup> Cecchini (2006); Montecuccoli degli Erri (2003, pp. 152ff). The Venetian ducat was equivalent to the Florentine florin.

<sup>113</sup> Montecuccoli degli Erri (2003, p. 146).

<sup>114</sup> Cecchini (2006); Shaw (2006).

<sup>115</sup> Cecchini (2006).

than on paintings, probably absorbed the lion's share of attention.<sup>116</sup> And, after all, the export of paintings did not suffer, nor the wellbeing of the more successful artists. A comparison with the glassmaking industry, however, shows that masters on the lagoon island of Murano, where the industry was concentrated, insisted that dealers in Venice itself only be allowed to sell the most common sorts of glassware. Fear of losing their independence led the masters to retain the exclusive right to sell all high-value sorts of glass, and something similar may have motivated the painters.

Whatever the reason, the marketing of paintings in 17th-century Rome was much less ordered than in Venice, and in part because it was beyond the ability of any authority to control it, the retail market flourished there. The Academy of Saint Luke in Rome prohibited full members from dealing, and, from 1633, levied a tax on all in the city who did engage in marketing paintings. But dealing could not be effectively policed, and there is evidence both that Academy members did trade in pictures, while many others who dealt in them did not pay the tax.

Haskell, in his classic study of the seventeenth-century art market in Rome, noted that artists and buyers were plentiful, and that there was "a growing appreciation of pictures as pictures rather than as exclusively the records of some higher truth".<sup>117</sup> Nonetheless, he passed quickly over dealers, because they "played little part in the lives of the more important painters". That is to say, there are few instances of "distinguished artists" who worked "systematically" for dealers, and few dealers who "ever ventured much beyond the relatively unknown painter".<sup>118</sup>

Recent research, however, unconstrained by Haskell's focus on "important" artists, has begun to reveal numerous instances of artisans selling paintings on the side at low prices, and of quasi-specialized dealers in the mid- to upper reaches of the market.<sup>119</sup> At those levels, archival finds relating to one specialized dealer, Pellegrino Peri (1624–1699), in particular, reveal five new facets of dealing activity in Rome.<sup>120</sup> First, Peri was not just a professional dealer but specialized in "genre" paintings. (He was also the only dealer known to date who made his money solely from paintings.) Then too, of at least six artists known to have worked for him under contract, at least four graduated fairly quickly to independent status, a dynamic for which we have no comparable evidence in the case of Venice. Third, Peri, though his paintings were mostly modest in price, also supplied prominent collectors, including the Pamphilj and Pietro Gabrielli in Rome and Giuseppe Maria Durazzo in Genoa, his home city. Fourth, he sold a *bloc* of paintings to the Pamphilj, a mode of transacting that became more frequent in the course of the seventeenth century, at times in the form of purchasing paintings, though at times also renting, by the room. *En bloc* purchases, and rentals, we might note, were made both by

<sup>116</sup> Ibid.

<sup>117</sup> Haskell (1980, p. 130).

<sup>118</sup> Ibid., p. 122, and p. 123, n.1.

<sup>119</sup> For the former see Cavazzini (2004) and – for the early years of the 18th C. – Coen (2004).

<sup>120</sup> For the details that follow see Lorizzo (2003b, 2006).



those knowledgeable about paintings and by those who sought pictures only as decoration. Fifth, Peri, at his death in 1699, left a stock of 2491 paintings, four times as many as the dealer Van Kessel in Antwerp in 1581, hinting at a market in Rome the scale of which might have matched, or even exceeded, that of Antwerp.

Still other recent archival finds suggest that even renowned collections in Rome might have been formed largely by acquiring paintings on the street, or at dealers' shops, but in either case by direct purchase and *not* via commissions.<sup>121</sup>

These revelations point towards a large, vibrant, and mainly unfettered retail market in 17th-century Rome, of the sort that Venice might have seen but did not, for the reasons we have identified. It is impossible to quantify elements in this market, but we do know that the market in Rome was driven with help from a large transient population, as well as by patronage stemming from the Papal Court. As to residents, Rome grew only modestly, from c. 110,000 in 1625 to c. 120,000 in 1675. But grow it did, while Venice declined, from c. 180,000 in 1550 to perhaps 150,000 in 1600, and c. 140,000 in 1625. Moreover, the resident population in Rome was enlarged by around one-quarter each year because of pilgrims and other visitors, and in a jubilee year it could more than double.<sup>122</sup> The annual demand for newly-made, cheap religious images, therefore, was substantial, quite apart from the patronage of wealthy families and of the Papacy and individuals close to the Popes. As to the population of artists, available contemporary records are not very helpful.<sup>123</sup> Nonetheless, to construct a possible point of reference, we might start with the Academy's list of dealers (*rivenditori di dipinti*) who submitted to the tax on dealing in 1675, a list likely to be more accurate than earlier ones because the tax had recently been reduced. This number was 40. Only some of this number would count as specialist dealers, but if we take an arbitrary fraction, 0.625, or 25 artist-dealers, and multiply it by seven, the average ratio of artists to specialized dealers in Venice from the tax lists of 1712 and 1719, this yields an estimate of 175 artists in Rome in 1675, a number we have used in our [Appendix A](#). It produces a ratio of about 1.5 artists per thousand (resident) population, which agrees with recent estimates by Bonfait based on studies of the residents of two parishes.

Whatever the true number of artists, there were important advantages stemming from the fact that dealers and buyers in Rome (as in Venice) were concentrated by location. The Piazza Navona was a major node.<sup>124</sup> This enabled a wealthy and knowledgeable collector, local or foreign, to shop around within a narrow physical compass and compare offers. The papal palazzo occupied (from 1651) by Prince Camillo Pamphilj, for example, was on Piazza Navona, a step from Peri's shop and those of several other dealers. It was also to this Piazza that John Evelyn went regularly during a visit to Rome in February 1645, drawn by those there who sold "medals, pictures, and such curiosities".

<sup>121</sup> This is argued persuasively, for the famous Colonna collection, by Natalia Gozzano (2003, 2004).

<sup>122</sup> Spear (2003, p. 310); Hibbert (1985, p. 202).

<sup>123</sup> For a selection of the totals in Academy's records see Lorizzo (2006, Figure 1).

<sup>124</sup> Ibid.



## 5. Selling paintings II: Auctions

We move, finally, to resale markets using the auction method. Here our focus will be on the evolution of rules.

### 5.1. “Dutch” auctions in Amsterdam

The interlopers from Antwerp who held illegal auctions in early 17th-century Amsterdam were said to have used the descending price method. In the same document it was noted that descending price was a method used only in estate sales.<sup>125</sup> In fact it seems likely that only in estate sales where second-hand clothes dealers were expected to predominate was the method used.<sup>126</sup> But perhaps the complainants knew that and mentioned it only to cast a slur on the interlopers’ auctions by association. In any event, the method is no longer employed for sales of paintings. It works best where speed is important (as in sales of perishables); where the goods are more or less interchangeable – so that someone who misses out on a lot can expect to bid for another just like it – and where buyers are very knowledgeable: they do not need the extra information provided in the ascending price, open outcry system. That last sustains our earlier conclusion that the buyers at the interlopers’ sales in Amsterdam were not likely to have been duped.

### 5.2. Magistrate-imposed rules for transparent auctions in 17th-century Lille

Perhaps the most concerted and sustained grab for market share by Antwerp dealers and artists occurred in 17th-century Lille.<sup>127</sup> It began, as far as surviving records allow us to state with certainty, in 1667 – though it probably started much earlier – and continued, more or less regularly, for at least the next forty-five years. There was a long-standing ban against foreigners selling outside the times of the fair, but the Lille magistrates, at a certain point, decided that dealers and some accompanying artists from Antwerp could better serve local buyers than could the city’s own artists. They therefore granted permission to these interlopers to hold auctions after the fair, and extended the period of those sales till all the items brought by the visitors were sold. Surreptitious selling after a fair was not uncommon,<sup>128</sup> but conducting auctions outside the fair openly and with full authorization might have been unique.

The decision to grant permission for post-fair sales was a concession by the Lille magistrates to powerful local buyers, who must have persuaded them that the painters of the city could not supply them with paintings comparable to the “rare and excellent things” brought in by the dealers from Antwerp. Implicitly therefore, the magistrates

<sup>125</sup> See De Marchi (1995) for details.

<sup>126</sup> Montias (2002b, pp. 21–22, pp. 259–260, notes 45 and 46).

<sup>127</sup> See De Marchi and Van Miegroet (2005c). This is our reading of documents rediscovered by Raux, and first analyzed by her in relation to public sales and collecting in Lille, for which see Raux (2002, 2005).

<sup>128</sup> Antwerp dealers did the very same in Paris, as we have seen.

were acknowledging a mutuality of interest between Lille buyers and Antwerp dealers; their permissions in effect promoted a coalition between those two groups, to the disadvantage of Lille artists.

The dealers' overriding interest was to get rid of all their stock – of course at prices that would give them the usual rate of return in their line of business. What did they stand to gain from selling everything? The dealers themselves said that it would save them money on exit tolls and carriage; but they might have added that they would also avoid being stuck, back in Antwerp, with old and tainted stock, and that they would save on interest costs by turning over their capital in a single season. There was also a certainty attached to those gains compared to selling during the fair, and it meant that the dealers could be satisfied with a return from the auctions even a bit below normal for dealers.

It was the buyers' interest as collectors to be supplied, on a sustained basis, with quality items – the Antwerp dealers did not only bring paintings with them. And if clearing everything was so important to the sellers, then buyers too could not but support after-fair sales.

At the fair, sellers, wanting to retain an attractive set of objects for the after-fair sales, set prices artificially high. A few buyers – those with a weak budget constraint, or those who, being risk averse, did not want to lose a piece they admired – would purchase during the fair. But the fair was basically an occasion for display and viewing. Once after-fair auctions became normal procedure, buyers could expect to get a better price for a desired lot at the sales than the deterrent prices posted during the fair.

This was a reasonable expectation, because the magistrates proscribed the devices that sellers commonly use to skew sales in their favor. The Antwerp dealers had to conduct their sales without reserves. That meant everything had to be sold – something dealers wanted anyway – but the magistrates also insisted that everything sold must be by open outcry: no private sales on the side. Finally, the magistrates prohibited false bidding.

These three rules ensured transparency and fairness on the part of the sellers, which in turn gave buyers confidence that the sales would not be rigged against them. As a result, buyers too would have been inclined to play fair. They would have no reason not to bid up to their true valuations, lot by lot, and – equally important – would have felt no need to resort to bidding rings to keep winning bids down. But under those incentives, sellers would also be less inclined to seek out new ways of gouging buyers. In short, the magistrates, knowingly or not, created conditions for the sales that probably improved the chances of mutual fair play, and prices that both buyers and sellers would find acceptable.<sup>129</sup>

<sup>129</sup> This is in line with economic auction theory, which suggests that where auctioneers fully share information with potential bidders their revenue is higher, since no rents accrue to winning bidders who secure a lot with a bid lower than the true value simply because they know – or are assumed to know – something that others do not. See *Milgrom and Weber (1982)*. The transparency insisted upon by the Lille magistrates seems to approximate the full information assumption.

No mechanism could guarantee such an outcome, certainly not lot by lot, or even sale by sale. But, given the desire of both parties to meet each other again and again, dissatisfactions could be signaled – bidding not serious; auctioneer seeming to favor certain bidders – and slight adjustments be made in the course of a selling season. Group interest here dominated individual interest, and any recalcitrant individual ignoring a complaint would be punished. Buyers who did join a ring might face aggressive bidding from other buyers as a way of imposing a “winner’s curse”, while a wayward and non-compliant seller would not be invited to form part of the group from Antwerp the next time around. We are not privy to such behind-the-scenes behavior, but one way or another the system worked: the sellers did come back, and buyers were pleased.

There are hints that Antwerp dealers also sought permission to hold auctions after the fair in Tournai, Douai, Cambrai, Valenciennes, Arras and Ypres. And, as noted earlier, they tried the same aggressive tactic in Ghent in the 1680s, though without success.

### 5.3. Auctioneer-contrived rules, London, late 17th century

The rules insisted upon by the Lille magistrates are unusual in the history of paintings auctions. A set more familiar, at least to Anglo-Saxon audiences, is the rules contained in printed catalogues for auctions held late in the 17th century, in London. Printing rules was done at the initiative of the auctioneer-organizers, who did not pretend expertise – they came to paintings via book and print sales – and who, it seems, were mainly concerned with turnover, on which their fee was based, and with protecting themselves against irresponsible bidders: those who bid for fun and left without paying for lots they had won.

Thus a basic rule required buyers to leave part of the purchase price before vacating the premises, a precaution against irresponsible bidding, or bidding to drive the price up for fun, but where the winner had no intention of paying. Further rules were soon added. One, widely copied, was a minimum increment-of-bid rule. The auctioneer who first introduced this rule stated that it was for his own convenience. The average time per lot in these sales was 1.4 to 2.4 minutes (in Sotheby’s auctions today it is less: 0.5 to 1.2 minutes). Thus the auctioneer had little time to raise the bid, and specifying a minimum increment would have both helped him move bids to a satisfactory level and saved him from the temptation to pull bids from the ceiling. Still other sales advertised that no false bidding would be used and that no commissions – bids by the auctioneer on another’s behalf – would be accepted.<sup>130</sup> The latter rules were appended to a very few sales, implying that the norm was precisely their opposite; that bidders could expect false bids and bids by the auctioneer acting for a seller or for an absent buyer, in both cases for a cut. Finally, in larger sales or more ordinary paintings it was often announced that there would be no undisclosed reserves, possibly a move to encourage and sustain

<sup>130</sup> In modern auctions the seller may not bid; the auctioneer may do so on the seller’s behalf, by prior arrangement, but only up to an agreed minimum. But the auctioneer may bid on another’s behalf.

attendance.<sup>131</sup> Again, by implication, in other sales bidders could expect undisclosed reserves. Today, the major international auction houses announce the relation between reserves and the lower pre-sale estimate, Christie's promising that the reserve will not exceed that figure, Sotheby's that as a rule it will be 75 percent thereof.

The London rules no doubt were an improvement on having no advertised rules at all; nonetheless the auctioneer's hand, and interest, is apparent in them. Intriguingly, modern international rules bear a closer family resemblance to the early London rules than they do to those imposed in Lille which, in turn, are distant kin to modern French rules which, among other practices, preclude buy-ins which end up in the hands of the auctioneer himself.

#### 5.4. *Creating value through auctions: The dealer Gersaint in Paris, 1733–1750*

As we have seen, one control exercised with unusual effectiveness by the Paris painters' corporation in the 17th century was the requirement – akin to the rule in Venice – that would-be dealers first be accepted as artists. A way around the requirement was to train as a *marchand-mercier*, specializing in paintings and related wares.<sup>132</sup> This route was taken by Edme-François Gersaint (1694–1750), and by some before and after him. But Gersaint used his status to effect major changes.

He started out modestly enough but through astute joint ventures and by acquiring expertise in several fields (print-making, porcelain, shells, antiquities, semi-precious stones, and so on, as well as paintings), he was able to realize economies of scope. The great range and exotic character of the goods he stocked (*chinoiserie*, Indian screens, *naturalia*, etc., as well as things in the fields already listed) made of his shop a sort of non-stop fair. He attracted clients at all levels, including those sensitive to price. On the whole, however, his paintings were expensive. He could justify this by claiming special knowledge which enabled him to be selective and could back the claim with an unmatched experience: he made frequent buying trips north to acquire Flemish and Dutch paintings, knew collections and collectors in Flanders and Holland, and had intimate knowledge of auctions there. Gersaint introduced Parisians to the ascending price auction without private side-sales; his sales had printed catalogues; and there were pre-sale viewing days. His catalogues were discursive and instructive, and his sales leisurely, allowing potential buyers to become familiar with paintings, and to form their valuations partly by discussing with, and observing the direction and strength of bidding by, others. In these several ways Gersaint eased control of the Paris market away from artist-dealers.

Among Gersaint's more radical notions was the idea that anyone, in principle, could become amateur lovers of paintings and form a respectable collection, even those without the money, leisure or learning to become true connoisseurs. In pursuit of this ideal

<sup>131</sup> On all this see De Marchi (2004).

<sup>132</sup> The *marchands-merciers* or generalist trader/mercers in Paris were not to produce, but could sell anything made outside Paris.

he sought to wean viewers and buyers from an excessive absorption with the painting *per se*, and especially with authorship and subject. Instead he invited them to consider paintings as bundles of pleasure-yielding properties, an idea similar to Lancaster's fresh approach to demand, adumbrated in the 1960s.<sup>133</sup> Gersaint's marketing ploy was so novel and sophisticated that it merits being spelled out in full. If the viewer's pleasure in a painting is made the primary ground for preference, it becomes possible to argue that paintings by any one of several artists supplying similar desired characteristics – even if differently-bundled – might be considered substitutable for one another. Such paintings constitute a “family” cluster. (In applying his ideas, of course, Gersaint made it clear that any paintings he offered for sale should be understood to be of collectable quality.) The paintings in such clusters, needless to say, would likely fetch different prices. A buyer could choose from among the paintings in a cluster that subset for which each painting yielded equal pleasure for the price. (Price here is to be read as expected price at auction; in other words, Gersaint's informed pre-sale estimates.) A final choice could be made by a bidder from this subset according to what his or her budget would permit.<sup>134</sup>

It was integral to Gersaint's convictions that individual viewers have the right to impart their own value to a painting. This was radical at the time but became a formative idea among artists of the Romantic and later Modern Art movements. The importance of the ideal is that it held the potential to extend the demand for paintings. This was taking dealing beyond the acceptance of marketing as a natural complement to making; it expanded the concept of the dealer to include one who nurtures the love of painting, a role now taken for granted among those who deal in the secondary market.

## 6. Coda

Our overview has been thematic but unavoidably also episodic. We have been reporting on research that is based on scattered documents which have survived or been uncovered largely by good luck, or which exist in large numbers but have not yet been fully explored. There are therefore serious gaps, and the record, such as it is, is always for individual cities. Do the truncated segments of a history we have been able to offer add up to an integrated whole? We began with a temporal sequence in the maturation of an art market, but we can now also attach dates and place names to the several stages, and link them across locations. As a result, too, some new details and emphases can be introduced here.

- (1) In the second half of the 15th century we observe, in Florence, made to order and on spec production in a nascent retail market, parallel to, though artistically derived from, the paintings generated within the commission nexus. A retail market

<sup>133</sup> The parallel is very close. See, e.g., Lancaster (1966).

<sup>134</sup> This decision process is illustrated for a cluster of artists painting Italianate landscapes, and for a cluster of artists who produced Flemish peasant scenes, in De Marchi and Van Miegroet (2005a, 2005b), respectively.

emerged at the same time in Bruges. It was larger, more vibrant, and offered a wider range of novel and differentiated sorts of imagery. Much of the production was cheap, eliciting demand from lower income groups and creating a deeper market than in Florence. The innovativeness of the visual imagery in Bruges stemmed from emulation among artists and competition for market share between three well-defined crafts of painting. The market was also supported by a large, resident and temporary population of relatively wealthy foreign merchants; by the existence of an annual international fair; and by an institutional infrastructure created by the city to facilitate exchanges – elements lacking in Florence.

- (2) Where Florence excelled was in the structuring of estate sales, out of which, around the turn of the 16th century, there grew something resembling a wholesale recycling market. This was dominated at first by old clothes dealers, certain of whom, over time, began to specialize in household movables, paintings among them. The city organizers facilitated this development by separating out lots with art objects. Late in the 16th century, in Antwerp, dealers appeared for the first time as members of the painters' guild, described, however, as second-hand clothes dealers who also traded in paintings.
- (3) The development of new products at the low end in Bruges fostered the habit of buying and displaying easel paintings in large numbers in the home. That particular transition was easier in the Low Countries than in 15th-century Florence, where easel paintings had to compete with paneling and with intarsia (wood inlay) "paintings".
- (4) In the 16th century various designated sales venues for paintings are to be found in Antwerp, the chief of them on the upper floor of a new trading Exchange. Antwerp and nearby Mechelen comprised a formidable production complex, painters in Antwerp specializing in oils and those in Mechelen in watercolors. Paintings from both centers were sold through the Exchange *pand*. Artists in the two cities, combined, numbered at least 200 in the mid-16th century, and were capable of producing upwards of 30,000 paintings per year. This was the first instance of a production capability consciously established to export paintings, though export had occurred earlier in a less planned manner, as in Bruges. Production on a large scale, for export, in Mechelen/Malines and in Antwerp – in the latter city via designated, year-round sales venues – were nonetheless natural outgrowths of the popular market created earlier in Bruges and its textiles display halls.
- (5) Venice, too, emerged in the 16th century as a major center of artistic production. However, in contrast to Antwerp, for example, where the rule that only artists should sell paintings was flexibly applied in the sixteenth century, in Venice it was insisted upon. Artists were allowed to make paintings that had not been commissioned, but only artists were allowed to sell – except for paintings to be exported. This posed time-allocation problems for artists, and predictably led to various selling arrangements being made with others – crafts persons in the same guild, friends and associates, and shopkeepers with no links to painting but

eager to move into their marketing. The ban also favored exports, but it was not easy for ordinary artist-dealers to move into the export trade in paintings, which required specialized experience, networks of contacts, and capital. When local demand slowed relative to the export trade, therefore, artists found themselves dependent on traders and specialist dealers, and having to work for them under contract, on unfavorable terms. That was the situation for much of the first half of the eighteenth century. Yet, in 1769, even dealers who had been allowed to join the College of Artists were told to register in future with the guild of furniture painters. The motive behind this is uncertain, but the proscription on selling by non-artists, it seems likely, was born of a desire to maintain control, albeit one that did not benefit painters in the long run. Venice joins Paris among cities whose painters' guilds resisted the marketing of paintings by others, to their ultimate disadvantage.

- (6) The phenomenon of artists working as "in the galley" was not unknown elsewhere, including 17th-century Amsterdam, Paris and Rome. However, in Antwerp, painting for a dealer was also an arrangement within which artists for the first time were able successfully to claim payment for invention. Vertically integrated dealers operating there treated original paintings made for them as means to generate copies and thereby an income stream, out of which payment for invention could readily be made. Viewing the original as capital was the conceptual and practical perspective within which this made sense. Earlier conceptual models may have existed in tapestry cartoons and the engraved printing plate, both of which were considered (and priced as) capital assets. With original paintings viewed as capital, however, even lesser painters had a framework within which they could successfully claim payment for invention. This was a first; artistic ingenuity had been valued within the commission nexus in Florence but was not necessarily paid for, patrons often claiming the originating idea as their own.
- (7) The period 1600 to 1750, the opposition of selected artists' guilds notwithstanding, was the era of the dealer. Quasi-specialist dealers arose in Rome in great numbers, as they had in Antwerp. Whereas in Antwerp the demand of foreign traders was important, however, in Rome demand was swelled by pilgrims and other visitors. The differences between Venice and Rome widened over the course of the seventeenth century. Specialist dealers in Rome employed beginning and newly-arrived artists, established themselves in concentrated areas close to wealthy buyers, and gave their artists such exposure that in many instances these artists could become independent upon serving their contracts – a dynamic not noticeable in Venice. Evidence is also emerging which suggests that even prominent collectors in Rome bought major portions of their collections on the piazza, or from dealers, probably using the great numbers and the physical concentration of artists and dealers to create a bargaining advantage for themselves. Recent research has also begun to reveal details of a low-end market in Rome, where non-specialist dealers also flourished. And it is showing that renting whole

rooms was not uncommon, certainly among those merely seeking decoration but also, for reasons that need to be explored further, among collectors.

- (8) Specialist dealers in a net-exporter market – the Antwerp–Mechelen complex, for example – displayed a wide variety of creative plays to secure market share. Traders based in Antwerp became involved in shipping large quantities of paintings abroad. But Antwerp dealers who chose to stay when local demand showed signs of slowing, found themselves engaging in forays to other cities accessible by land, to increase their share of those markets. Various stratagems were tried, ranging from experimenting with the descending price auction in Amsterdam, to holding post-fair auctions in Lille, though on conditions for transparency laid down by the local magistrates. There were also individual Antwerp-based dealers who drove a sophisticated and far-flung traffic in paintings, using agents, to Augsburg, Vienna, Paris, Lisbon, Bilbao, Seville, Cadiz and the Americas. Such specialist dealers were not merchants with a sideline in paintings. They employed business practices developed by international trading firms and indistinguishable from them, except that as suppliers to foreign markets they were, inevitably, also involved in mediating visual culture across traditions. In a net-importer city such as Amsterdam, by contrast, dealers served more as arbitragers, helping to supply local demand by redistributing surplus supplies originating in other cities. By the 17th century, too, dealers were trading in both new and older – recycled – paintings; in the latter trade they competed with second-hand dealers, who also at times tried to enter the first-sales market.
- (9) If the 17th century was the era of the trader–dealer in many variants, the 18th was the century of the auction. Regular auctions involving paintings were held by the Amsterdam Orphan Chamber in the early decades of the 17th century. However, the first auctions for which we possess printed catalogues with rules, are sales held later in the century, in London. Those sales, nonetheless, were mostly got-up affairs, organized and run by bookseller-auctioneers with no special knowledge of paintings. Paris became the more innovative center in the first half the 18th century. There, extensive, informative catalogues were printed; viewing days were provided; and the sales themselves were leisurely, serving as occasions for arriving at private valuations partly through social exchange. A higher tone was set than in London, where the rules were written by and for the auctioneers. The first prominent dealer and organizer of auctions in Paris, Gersaint, voluntarily adopted the no-friendly-sales rule that magistrates had insisted upon in Lille, and took a responsible attitude towards attribution, at the same time seeking to wean potential buyers away from an undue preoccupation with authorship and subject. He stressed instead that value stems from the pleasure a painting gives to its viewers. Others took up Gersaint’s model – professional dealing by a non-artist – and, by the 1770s, 20–40 sales a year were being held in Paris, surpassing London’s early lead.



## 7. Some developments since 1750

Our coverage ends around 1750, but we can indicate briefly and selectively some more recent developments.

A partially integrated auction circuit developed during the 18th century, incorporating Amsterdam, Brussels, London and Paris. It did not result in common prices for the same artists because tastes differed; moreover, transport costs and the associated risks limited the physical movement of all but high-end paintings. These impediments did not outweigh the arbitrage possibilities for dealers operating between sales centers. The Amsterdam sales in particular served to recycle 17th-century Dutch collections to Germany, while many French collections were dispersed internationally through auctions in the 1790s. Private and public purchases on behalf of Catherine the Great, in the second half of the 18th century, supplied the core of the exceptional collection of the Hermitage in Saint Petersburg. Late 19th- and early 20th-century buying in Western Europe also built the collections of wealthy Americans, most of which, under positive incentives in the US tax code, have since become, or made their way into, public museums. The first great museums in Europe, eventually all open to the public, were established in the 18th century,<sup>135</sup> and have grown subsequently in lockstep with legislation in many countries restricting the export of “national treasures”.

The structure of the international market in paintings quickly acquired the features by which it was characterized in the 20th century. The two great auction houses, Christie’s and Sotheby’s, both date from the mid-18th century, and their rules reflect those preserved for us in the printed catalogues of the late 17th-century sales in London. Transparency, however, has increased, though it is still less than total.

The rules followed by the leading international auction houses are still not universally adopted; moreover, the degree of transparency shown by Sotheby’s and Christie’s is still rare. Until quite recently, too, the international auction market was less than fully integrated, being circumscribed by national restrictions. Thus many countries insist on an export license, which tends to keep international-quality paintings off the market altogether; and in France and Spain, for example, a right of preemption is assumed by the state, while in France until very recently auctions could be conducted only by French auction houses.

Another peculiarity of the international resale market is that the two dominant houses, Sotheby’s and Christie’s, do not constitute a conventional duopoly. Both represent sellers in the first instance, and sell services: they are not producers. They battle, therefore, chiefly to secure collections in markets where stocks are small and declining. This entails making concessions to potential sellers which cause costs to rise. Costs increase, for example, because lesser works (which cannot be refused) often are more difficult to

<sup>135</sup> A short list, supplied by Sophie Raux, includes: the British Museum (1759), the Pinakothek in Munich (1779–1793), the Capodimonte in Naples (1739–1793), the Louvre (1793), the Vatican Museum (1784) and the Imperial Museum in Vienna (1781).

authenticate, and because the seller's premium must be discounted or price guarantees given to a seller on single items and even whole collections. Such concessions and the higher costs they entail led, in 1995, to an agreement between Sotheby's and Christie's to apply a common, non-negotiable sliding scale of seller charges. This was exposed later in the decade as illegal collusion, at which point Phillips auction house was positioned by its new owners, the French consortium LVMH, to topple the leaders. The tactic used was to offer concessions to sellers of the very sort the leaders had concluded they could no longer afford to make. When the art market slowed in 2001, this strategy turned out to be a loser, and the two leaders again find themselves unchallenged at the top, though they are also rumored to have re-started competitive inducements to sellers. Recent reports also suggest that they may be expanding into the territory of dealers, and increasingly engaging in private sales.

The unsustainability of the competitive pressures in select, older markets, such as those in Old Master paintings, drawings and prints, and Impressionist paintings, probably helped induce the proliferation in recent decades of sales in new fields, including, in the case of Christie's Amsterdam, sales of copies. Sotheby's also flirted briefly with online auctions. This has ended, except for relatively low-priced items, but both auction houses retain their lower-end physical auctions in separate locations in London. The market in Contemporary paintings and more generally for low-priced works has continued to expand, aided by new online information services (e.g., Artprice) and by online auctions such as eBay and Lauritz.com.

Gallerists proliferated in the 20th century, especially in the primary market for Modern and Contemporary art. This is a segment of the art trade characterized by the same radical uncertainty as movies – discussed by De Vany in his entry in this Handbook. The idea that the modern gallerist – one who invests in beginning artists' careers – originated in late 19th-century Paris, has been disputed. The true model of this type it seems is a post-World War II New York phenomenon. The modern contract between a beginning artist and a gallerist, however, currently often involving a fifty-fifty split of sales revenues after costs, is to be found in early 20th-century Paris.

As to the status of artists, their right to have the artistic integrity of their work protected came to be widely recognized in the 20th century, as was an artist's right – or that of heirs – to a share of the proceeds of secondary sales. It is not clear that this last supports those who need it most – unknown artists will continue to need willing, risk-loving gallerists – while those who benefit are likely to be relatively few, and the benefits necessarily accrue after the greatest need for income is past. Nonetheless, there is momentum to enlarge the number of countries with an artists' resale right. Copyright law, meanwhile, as shown in the entry on copies in this Handbook by Ginsburgh and Benhamou, continues to evolve, courts tending to recognize that even replicas, provided they embody considerable skill and labor, are not in breach of copyright. The status of digital images, however, remains uncertain.

## Appendix A

### A.1. Number of painters per '000 of population, by city, 1470–1764

The following chart (Table A.1) summarizes estimates of artists per thousand inhabitants of various cities in the early modern period. Where possible more than one observation is provided. Many of the estimates are to be taken as having a large potential range of error, not only because population numbers are uncertain, but also because numbers of artists often have to be inferred. Moreover, it is not always clear from the sources whether “artist” is to be taken in the narrow sense of one trained in design and figurative painting, or has a broader connotation. The value of the exercise lies not in the accuracy of any single estimate but in broad consistency of the ratios presented here with other things that are known about a city. For example, it is known from fairly reliable counts that Amsterdam, even at its peak as a center of production of paintings (c. 1650), continued on balance to be net importer of paintings. This is reflected in the ratios of artists per thousand inhabitants not rising above 1.0. Ratios significantly above

Table A.1

	Florence	Bruges	Valencia	Venice	Antwerp	Mechelen	Haarlem	Paris	Amsterdam	Rome
1470/72	0.8	1.4								
1488		2.8								
1496					0.6					
1513			0.3							
1520		1.7			2.0					
1526					2.4					
1530				0.8						
1542/43					1.5					
1550					1.4					
1566/68					1.5	5.3				
1585/86					1.3					
1588/89					1.8					
1605							0.4			
1611								0.4		
1612					2.2					
1615							0.8			
1619						3.5				
1625							1.1			0.4–1.5
1630/31					1.7				0.9	
1632						3.5				
1635							1.9			
1640				1.1						
1650							1.8		1.0	
1675										1.2
1677								0.5		
1764								1.0		

one, on the other hand, are consistently found for cities such as Bruges, Mechelen and Antwerp which, on other grounds, can safely be dubbed net exporters of paintings. As a rough rule of thumb, therefore, a ratio consistently somewhat above 1.0 inclines us to think that a city was a net exporter of paintings, an inclination always to be checked, however, against collateral historical evidence. In some cases the ratio rises or falls over time indicating, in principle, an alteration of status (e.g., Haarlem, Rome and possibly Paris).

## A.2. *Sources and methods*

*FLORENCE*. Benedetto Dei [1470, for which see Gilbert (1980, p. 183)] listed some 30 painters with their own workshop. However, several on that list were already dead. Wackernagel reports “about thirty” figure painters listed in 1472 as members of the Accademia di San Luca, which artists were supposed to join [Wackernagel (1981, p. 300)]. We have used the number 30. Population in 1469: Herlihy and Klapisch-Zuber (1985, pp. 40, 332) [also in Goldthwaite (1982, p. 33)].

*BRUGES*. In the years 1456–1470, 44 new masters were added to panel and cloth painters, joining 12 members in 1456, for a total of 56 in 1470, assuming zero deaths, retirements, or net immigration of artists. Population in 1449, c. 5000 [Nicholas (1992, p. 371)]. We have adopted this number, amplified by 500 to allow for some net immigration. By 1488 the population might have been 36,000 [Blockmans (1995, Table 2)], while the number of painters – same assumptions as above – had risen to c. 90. By 1520 the number of artists was about 60 while population was still c. 40,000.

*VALENCIA*. A tax record from 1513 lists 19 workshops of figurative painters (one workshop with two masters) [Falomir (1996, p. 211)]. The population early in the 16th century is estimated from the number of houses in 1510 – 9,879 (ibid., 109) – to which we have applied the generous assumption of 6 persons per household, yielding a population of c. 60,000.

*VENICE*. “Figurers”, gilders and miniaturists in 1530 numbered 110, if the same ratio applied to all 230 guild members as to the 77 with a craft listed [Favaro (1975, pp. 137ff., doc. I.1)]. Population in 1530 c. 130,000 [Beltrami (1954, p. 59)]. For artists in 1640, we have used the number 138, based on a guild list for tax purposes of that year [Favaro (1975, pp. 163ff., doc. II.2)], and for population, the figure of 120,300 for 1642 [Pezzolo (2003, p. 151)].

*ANTWERP*. Three absolute counts of painters are known from guild lists: 14 in 1453, 108 in 1585–1586, and 76 in 1588–1589 [the latter two reported by Vermeylen (2003, p. 112, n. 19)]; the number for 1453 and all other estimates arrived at from a list of annual guild inscriptions for master painters [Martens and Peeters (2002)] plus conservative assumptions about length of career (15 years 1453–1467; thereafter 20), zero net emigration 1453 through 1500, thereafter 5 masters per year through 1604, 7 through 1609, 10 through 1617, and 15 through 1625. This results in estimates of the number of painters of 23 in 1496, 102 in 1520, 131 in 1526, 124 in 1542–1543, 140 in 1550, 152 in 1568, 121 in 1612, and 95 in 1630. Population figures: Vermeylen (2003, Table 2, p. 37) and references there cited, plus, for 1520, 1550 and 1630: Van Houtte and Van Buyten, in Wilson and Parker (1977, p. 82).

*MECHELEN/MALINES*. We used the contemporary count of 150 workshops, adjusted downwards by the ratio of known *painters* to total masters, and adding in assistants in the ratio of 27 per 100 master painters, from a 1632 document, to yield 159 painters for 1566, when population might have been 30,000 [Previer and Blockmans (1986, p. 79)]. In 1619, 95 masters protested

against out-of-town dealers. The transcript of a 1632 *Rekenboek van het schildersambacht van Mechelen* recorded 100 workshops [Monballieu (1971, p. 75)]. Known master *painters* plus assistants in those years averaged 69, in a population of perhaps 20,000–21,000 (latter figure from Van Houtte and Van Buyten, in Wilson and Parker (1977, Table 4.3).

**HAARLEM.** Artist and population figures, except for 1650, due to Boers, reported in Sluijter (2000, p. 137, n. 19): 11 artists in a population of 30,000 in 1605, 29 in 35,000 in 1615, 44 in 40,000 in 1625, and 80 in c. 42,000 in 1635. For 1650, Montias estimates 68 artists in a population of c. 38,000 (1990, p. 61). All numbers for painters very conservatively counted.

**PARIS.** Schnapper (2001, pp. 424–425) estimates painters and sculptors in 1611 at 110–115 in a population of c. 300,000. We have applied the 17th-century ratio of 2 : 1, painters to sculptors, to arrive at an estimate of 75 painters in 1611. For 1677, Schnapper estimates painters alone at 185, in a population of c. 450,000. For 1764, he determines that there might have been some 600–700 painters, assuming a ratio of painters to sculptors of roughly 2 : 1, and allowing something for widows and “demoiselles”, who were not counted in the 17th century, in a population of perhaps 600,000–650,000.

**AMSTERDAM.** Montias puts artists in 1630/31 in the range 90–120 [Montias (2002a, p. 117)], in a population of c. 115,000: Frederiks and Frederiks (1890, v.). For 1650, Montias estimates 175 artists in a population of 175,000 [Montias (1990, p. 61)].

**ROME.** The number of artists in 1625 is from a house-by-house census: Lorizzo (2006). Population in 1625 (c. 110,000), from Sonnino (1998). We have estimated the number of artists in 1675 by assuming, first, the same number of specialist dealers as in Venice (average of two observations: tax documents from 1712 and 1719), and using the same ratio of artists to dealers there, namely 7 : 1 [Montecuccoli degli Erri (2003, pp. 145–146)]. We applied this ratio to a fraction, arbitrarily set 0.625, of the number of self-declared dealers in paintings in a 1675 tax list in Rome, namely 33, yielding an estimate of the number of artists of 144 in that year. [For the tax record generating 33 dealers see Lorizzo (2003a, pp. 331–332).] The population in 1675 is estimated to have been about 120,000 [from Sonnino (1998)]. The number 175 is close to that estimated by Bonfait (2003), who found for two parishes in the early decades of the 17th century roughly 1 to 3 artists per thousand population, and takes 1.5 as a reasonable per annum average for the century as a whole. A detailed count of artists who authored public works, using older published sources, has been undertaken by Spear (personal communication). His yearly average total of artists active in Rome falls short of Bonfait’s estimate but provides a solid base from which to make upward adjustments for obvious under-counting (e.g., of works in private collections, and of foreign artists). There is conflict between Bonfait’s 1.5 yearly average number of artists per thousand population and our number for 1625, but we have included both to give a range.

## References

- Akerlof, G. (1970). “The market for ‘lemons’: Quality uncertainty and the market mechanism”. *Quarterly Journal of Economics* 84, 488–500.
- Allerston, P. (2003). “The second-hand trade in the arts in early modern Italy”. In: Fantoni, M., Matthew, L.C., Matthews-Grieco, S.F. (Eds.), *The Art Market in Italy, 15th–17th Centuries*. Panini, Modena, pp. 301–312.
- Beltrami, D. (1954). *Storia della popolazione di Venezia dalla fine del secolo XVI alla caduta della Repubblica*. Cedam, Padua.

- Benedict, P. (1985). "The comparative study of the popular market for art: The ownership of paintings in seventeenth-century Metz". *Past and Present* 109, 100–117.
- Blanchard, I. (1996). "Credit and commerce: From the Mediterranean to the Atlantic in the sixteenth century". In: Diederiks, H., Reeder, D. (Eds.), *Cities of Finance*. North-Holland, Amsterdam, pp. 21–34.
- Blockmans, W. (1995). "The creative environment: Incentives to and functions of Bruges art production". In: Ainsworth, M.A. (Ed.), *Petrus Christus in Renaissance Bruges. An Interdisciplinary Approach*. Brepols/Metropolitan Museum of Art, Turnhout/New York, pp. 11–20.
- Blockmans, W. (1996). "The Burgundian court and the urban milieu as patrons in 15th century Bruges". In: North, M. (Ed.), *Economic History and the Arts*. Böhlau, Cologne, Weimar, Vienna, pp. 15–26.
- Blondé, B. (2002). "Art and economy in seventeenth- and eighteenth-century Antwerp: A view from the demand side". In: Cavaciocchi, S. (Ed.), *Economia e Arte. Atti della "Trentatreesima Settimana di Studi" 30 Aprile–4 Maggio, 2001, Istituto Internazionale di Storia Economica "F. Datini", Prato*. Le Monnier, Florence, pp. 379–391.
- Blume, A.C. (2003). "Boticelli and the cost and value of altarpieces in late fifteenth-century Florence". In: Fantoni, M., Matthew, L.C., Matthews-Grieco, S.F. (Eds.), *The Art Market in Italy*. Panini, Modena, pp. 151–161.
- Boers, M. (2000). "Een nieuwe markt voor kunst. De expansie van de Haarlemse schilderijenmarkt in de eerste helft van de zeventiende eeuw". In: Falkenburg, R. et al. (Eds.), *Art for the Market*. In: *Netherlands Yearbook for the History of Art*, 1999, vol. 50, pp. 195–219.
- Bonfait, O. (2003). "'Molti francesi e fiammenghi che vanno e vengono non li si puol dar regola'. Mobilité géographique, mode d'habitat et innovation artistique dans la Rome de Nicolas Tournier". In: Bertrand, P.-F. (Ed.), *Nicolas Tournier et la peinture caravaggesque en Italia, en France et en Espagne, Actes du colloques des 7–9 Juin, 2001*. CNRS, Université de Toulouse Le-Mirail, Toulouse, pp. 65–75.
- Borean, L. "Ricchezza virtuose. Il collezionismo privato a Venezia nel Seicento (1630–1700)". Ph.D. Dissertation. Università di Udine.
- Campbell, L. (1976). "The art market in the southern Netherlands in the fifteenth century". *Burlington Magazine* 118, 188–198.
- Cavazzini, P. (2004). "La diffusione della pittura nella Roma di prima seicento: Collezionisti ordinario e Mercanti". *Quaderni Storici* 116, 353–374.
- Cecchini, I. (1998). "Il mercato d'arte pittorica a Venezia nel Seicento". Ph.D. Dissertation. Università Bocconi, Milan.
- Cecchini, I. (2000). *Quadri e commercio a Venezia durante il seicento: Uno studio sul mercato dell'arte*. Marsilio, Venice.
- Cecchini, I. (2006). "Troublesome business. Dealing in Venice, 1600–1750". In: De Marchi, N., Van Miegroet, H.J. (Eds.), *Mapping Markets for Paintings in Europe, 1450–1750*. Brepols, Turnhout, pp. 125–132.
- Coen, P. (2004). "Vendere e affittare quadri. Giuseppe Sardi, capomastro muratore e mercante d'arte (Roma, XVIII Sec.)". *Quaderni Storici* 116, 421–448.
- Commanducci, R. (2003). "Produzione seriale e mercato dell'arte a Firenze tra quattro e cinquecento". In: Fantoni, M., Matthew, L.C., Matthews-Grieco, S.F. (Eds.), *The Art Market in Italy*. Panini, Modena, pp. 105–113.
- De Marchi, N. (1995). "The role of Dutch auctions and lotteries in shaping the art market(s) of 17th century Holland". *Journal of Economic Behavior & Organization* 28, 203–221.
- De Marchi, N. (2004). "Auctioning paintings in late seventeenth-century London: Rules, segmentation and prices in an emergent market". In: Ginsburgh, V.A. (Ed.), *Economics of Art and Culture*. Elsevier, Amsterdam, pp. 97–128.
- De Marchi, N., Van Miegroet, H.J. (1994). "Art, value, and market practices in the Netherlands in the seventeenth century". *The Art Bulletin* 76, 451–464.
- De Marchi, N., Van Miegroet, H.J. (1996). "Pricing invention: "Originals," "copies," and their relative value in seventeenth century Netherlandish art markets". In: Ginsburgh, V.A., Menger, P.-M. (Eds.), *Economics of the Arts*. Selected Essays. Elsevier, Amsterdam, pp. 27–70.
- De Marchi, N., Van Miegroet, H.J. (2000a). "Exploring markets for Netherlandish paintings in Spain and Nueva España". In: Falkenburg, R., de Jong, J., Meijers, D., Ramakers, B., Westermann, M. (Eds.), *Kunst*

- voor de Markt/Art for the Market, 1500–1700. In: *Netherlands Yearbook for the History of Art* 1999, vol. 50, pp. 81–111.
- De Marchi, N., Van Miegroet, H.J. (2000b). “Rules versus play in early modern art markets”. *Art Markets* 66, 145–165. Special Issue: Ginsburgh, V. (Ed.), *Recherches économiques de Louvain/Louvain Economic review*.
- De Marchi, N., Van Miegroet, H.J. (2003). “Flemish altarpieces for the Italian market”. Mimeo. Duke University.
- De Marchi, N., Van Miegroet, H.J. (2005a). “Transforming the Paris art market, 1718–1750”. In: De Marchi, N., Van Miegroet, H.J. (Eds.), *Mapping Markets for Paintings in Europe, 1450–1750*. Brepols, Turnhout. In press.
- De Marchi, N., Van Miegroet, H.J. (2005b). “Brueghel in Paris”. *Bulletin van de Koninklijke Musea voor Schone Kunsten van België/Bulletin des Musées Royaux de beaux-Arts de Belgique*. In press.
- De Marchi, N., Van Miegroet, H.J. (2005c). “Invasions of the Lille market (17th C.) by artists and dealers from Antwerp”. In: Vermeylen, F., Vlieghe, H. (Eds.), *Art Auctions and Dealers. The Dissemination of Netherlandish Paintings During the Ancien Régime*. Brepols, Turnhout. In press.
- De Marchi, N., Van Miegroet, H.J. (Eds.) (2006). *Mapping Markets for Paintings in Europe, 1450–1750*. Brepols, Turnhout.
- De Marchi, N., Van Miegroet, H.J., Raiff, M.E. (1998). “Dealer–dealer pricing in the mid-seventeenth century Antwerp to Paris art trade”. In: North, M., Ormrod, D. (Eds.), *Art Markets in Europe, 1400–1800*. Ashgate, Aldershot, pp. 113–130.
- Denucé, J. (Ed.) (1931). *Exportation d’œuvres d’art en 17e siècle a Anvers. La firme Forchondt. Sources pour l’histoire de l’art Flamand. I. De Sikkel*, Antwerp.
- De Vries, J. (1981). *Barges and Capitalism. Passenger Transportation in the Dutch Economy, 1632–1839*. HES Publishers, Utrecht.
- De Vries, J., van der Woude, A. (1997). *The First Modern Economy. Success, Failure, and Perseverance of the Dutch Economy, 1500–1815*. Cambridge University Press, Cambridge.
- Duverger, E. (1969). “Nieuwe gegevens betreffende de kunsthandel van Matthijs Musson en Maria Fourmenois te Antwerpen tussen 1633 en 1681”. Reprint of *Gentse Bijdragen tot de Kunstgeschiedenis en de Oudheidkunde* 21.
- Echeverría, I. (2000). “Notas sobre el comercio artístico entre Sevilla y América en 1586”. *Museo de América, Annales* 8, 103–110.
- Ewing, D. (1990). “Marketing art in Antwerp, 1460–1560: Our lady’s pand”. *The Art Bulletin* 72, 558–584.
- Falomir, M. (1996). *Arte en Valencia, 1472–1522*. Martín, Valencia.
- Falomir, M. (2006). “Artists’ responses to the rise in demand for paintings in Spain, 1575–1625”. In: De Marchi, N., Van Miegroet, H.J. (Eds.), *Mapping Markets for Paintings*. Brepols, Turnhout, pp. 135–161.
- Fantoni, M., Matthew, L.C., Matthews-Grieco, S.F. (Eds.) (2003). *The Art Market in Italy, 15th–17th Centuries*. Panini, Modena.
- Favaro, E. (1975). *L’arte dei pittori in Venezia e I suoi statuti*. Olschki, Florence.
- Frederiks, J.G., Frederiks, P.J. (1890). *Kohier van den Tweehonderdsten Penning Van Amsterdam en Onderhoorige Plaatsen over 1631*. Ten Brink and De Vries, Amsterdam.
- Galassi, M.C. (1997). “A technical approach to a presumed Memling diptych: Original work and some Italian copies”. In: Verougstraete, H., Van Schoute, R., Smeyers, M. (Eds.), *Memling Studies*. Peeters, Leuven, pp. 339–350.
- García Fuentes, P.G. (1988). “El comercio de obras de arte de Sevilla a Hispanoamérica a fines del siglo XVI”. Ph.D. Dissertation. University of Seville.
- Gelderblom, O. (2000). *Zuid-Nederlandse Kooplieden en de opkomst van de Amsterdamse Stapelmarkt (1578–1630)*. Verloren, Hilversum.
- Gilbert, C.E. (1980). *Italian Art 1400–1500. Sources and Documents*. Northwestern University Press, Evanston, IL.
- Glasser, H. (1977). *Artists’ Contracts of the Early Renaissance*. Garland, New York.
- Goldthwaite, R.A. (1982). *The Building of Renaissance Florence. An Economic and Social History*. Johns Hopkins University Press, Baltimore.

- Goldthwaite, R.A. (2003). "Economic parameters of the Italian art market 15th to 17th centuries". In: Fantoni, M., Matthew, L.C., Matthews-Grieco, S.F. (Eds.), *The Art Market in Italy*. Panini, Modena, pp. 423–444.
- Gozzano, N. (2003). "Resa dei conti in Casa Colonna. Contabilità di una collezione romana del Seicento fra committenza e mercato". In: Cappelletti, F. (Ed.), *Decorazione e collezionismo a Roma nel Seicento. Vicende di artisti, committenti e mercanti*. Gangemi, Rome, pp. 175–185.
- Gozzano, N. (2004). *La quadreria di Lorenzo Onofrio Colonna. Prestigio nobiliare e collezionismo nella Roma barocca*. Bulzoni, Rome.
- Greve, A. (2000). "Brokerage and trade in medieval Bruges: Regulation and reality". In: Stabel, P., Blondé, B., Greve, A. (Eds.), *International Trade in the Low Countries (14th–16th Centuries)*. Merchants, Organisation, Infrastructure. Garant, Leuven and Apeldoorn, pp. 37–44.
- Grief, A. (1993). "Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition". *American Economic Review* 83, 525–548.
- Grief, A., Milgrom, P., Weingast, B.R. (1994). "Coordination, commitment, and enforcement: The case of the merchant guild". *Journal of Political Economy* 102, 745–776.
- Haskell, F. (1980). *Patrons and Painters. Art and Society in Baroque Italy*. Yale University Press, New Haven.
- Herlihy, D., Klapisch-Zuber, C. (1985). *Tuscans and Their Families. A Study of the Florentine Catasto of 1427*. Yale University Press, New Haven.
- Hibbert, C. (1985). *Rome. The Biography of a City*. Norton, New York.
- Holmes, M. (2003). "Neri di Bicci and the commodification of artistic values in Florentine painting (1450–1500)". In: Fantoni, M., Matthew, L.C., Matthews-Grieco, S.F. (Eds.), *The Art Market in Italy*. Panini, Modena, pp. 213–223.
- Kinkead, D. (1984). "Juan de Luzón and the Sevillian painting trade with the new world in the second half of the seventeenth century". *The Art Bulletin* 66, 303–310.
- Kubersky-Piredda, S. (2003). "Immagini devozionali nel Rinascimento fiorentino: Produzione, commercio, prezzi". In: Fantoni, M., Matthew, L.C., Matthews-Grieco, S.F. (Eds.), *The Art Market in Italy*. Panini, Modena, pp. 115–125.
- Lancaster, K. (1966). "A new approach to consumer theory". *Journal of Political Economy* 74, 132–157.
- Lorizzo, L. (2003a). "Il mercato dell'arte a Roma nel XVII secolo: "Pittori bottegari" e "rivenditori di quadri" nei documenti dell'Archivio Storico dell'Accademia di San Luca". In: Fantoni, M., Matthew, L.C., Matthews-Grieco, S.F. (Eds.), *The Art Market in Italy*. Panini, Modena, pp. 325–336.
- Lorizzo, L. (2003b). "Documenti inediti sul mercato dell'arte. I testamenti e l'inventario della bottega del Genovese Pellegrino Peri "rivenditore di quadri" a Roma nella seconda metà del Seicento". In: Cappelletti, F. (Ed.), *Decorazione e collezionismo a Roma*. Gangemi, Rome, pp. 159–174.
- Lorizzo, L. (2006). "People and Practices in the Paintings Trade in 17th C. Rome". In: De Marchi, N., Van Miegroet, H.J. (Eds.), *Mapping Markets for Paintings*. Brepols, Turnhout, pp. 343–358.
- Lydecker, J.K. (1987). "The domestic setting of the arts in Renaissance Florence". Ph.D. Dissertation. The Johns Hopkins University.
- Lynch, J. (1969). *Spain under the Hapsburgs, vol. 2, Spain and America 1598–1700*. Blackwell, Oxford.
- Martens, M.P.J. (1998). "Some aspects of the origins of the art market in fifteenth-century Bruges". In: North, M., Ormrod, D. (Eds.), *Art Markets in Europe, 1400–1800*. Ashgate, Aldershot, pp. 19–27.
- Martens, M.P.J. (1999). "The position of the artist in the fifteenth century: Salaries and social mobility". In: Blockmans, W., Janse, A. (Eds.), *Showing Status. Representation of Social Positions in the Late Middle Ages*. Brepols, Turnhout, pp. 387–414.
- Martens, M.P.J., Peeters, N. (2002). "Painters by numbers: Quantifying artists' trades in 16th century Antwerp". Mimeo. University of Groningen.
- Matthew, L.C. (2006). "Painters marketing paintings in 16th C. Venice". In: De Marchi, N., Van Miegroet, H.J. (Eds.), *Mapping Markets for Paintings*. Brepols, Turnhout, pp. 307–325.
- Milgrom, P.R., Weber, R.J. (1982). "A theory of auctions and competitive bidding, II". Mimeo. Stanford University and Northwestern University.
- Monballieu, A. (1971). "Documenten van het Mechels schilders- en beeldsnijders ambacht. II. Het rekwest van 1562 en het probleem van de 51 of 150 ateliers". *Handelingen van de Koninklijke Kring voor Oudheidkunde, Letteren en Kunst van Mechelen* 75, 71–82.



- Montecuccoli degli Erri, F. (2003). "I 'botteggeri da quadri' e I 'poveri pttiroi famelici'. Il mercato dei quadri a Venezia nel Settecento". In: Dal Pozzolo, E.M., Tedoldi, L. (Eds.), *Tra Committenza e Collezionismo. Studi sul mercato dell'arte nell'Italia settentrionale durante l'età moderna*. Terra Ferma, Vicenza, pp. 143–166.
- Montias, J.M. (1982). *Artists and Artisans in Delft. A Socio-Economic Study of the Seventeenth Century*. Princeton University Press, Princeton.
- Montias, J.M. (1985). "Flemish and Dutch trade in works of art in the 17th and 17th centuries". Mimeo. Yale University.
- Montias, J.M. (1986). "Cost and value in seventeenth-century Dutch art". *Art History* X, 455–466.
- Montias, J.M. (1990). "Estimates of the number of Dutch master-painters, their earnings and their output in 1650". *Leidschrift* 6, 59–74.
- Montias, J.M. (1991). "Works of art in seventeenth-century Amsterdam: An analysis of subjects and attributions". In: Freedberg, D., de Vries, J. (Eds.), *Art in History. History in Art. Studies in seventeenth-century Dutch Culture*. Getty Center, Santa Monica, pp. 331–372.
- Montias, J.M. (1996). "Works of art in a random sample of Amsterdam inventories". In: North, M. (Ed.), *Economic History and the Arts*. Böhlau, Cologne, Weimar, Vienna, pp. 67–88.
- Montias, J.M. (2001). "Commentary". In: Ainsworth, M.W. (Ed.), *Early Netherlandish Painting at the Crossroads. A Critical Look at Current Methodologies*. Metropolitan Museum of Art, New York and Yale University Press, New Haven and London, pp. 62–65.
- Montias, J.M. (2002a). "Notes on economic development and the market for paintings in Amsterdam". In: Cavaciocchi, S. (Ed.), *Economia e Arte*. Le Monnier, Florence, pp. 115–130.
- Montias, J.M. (2002b). *Art at Auction in 17th Century Amsterdam*. University of Amsterdam Press, Amsterdam.
- Murray, J.M. (2000). "Of nodes and networks: Bruges and the infrastructure of trade in fourteenth-century Europe". In: Stabel, P., Blondé, B., Greve, A. (Eds.), *International Trade in the Low Countries*. Garant, Lueven and Apeldoorn, pp. 1–14.
- Nicholas, D. (1992). *Medieval Flanders*. Longmans, New York.
- North, D.C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge University Press, Cambridge.
- Nuttall, P. (2004). *From Flanders to Florence. The Impact of Netherlandish Painting, 1400–1500*. Yale University Press, New Haven and London.
- O'Malley, M. (2003). "Commissioning bodies, allocation decisions and price structures for altarpieces in fifteenth- and early sixteenth-century Italy". In: Fantoni, M., Matthew, L.C., Matthews-Grieco, S.F. (Eds.), *The Art Market in Italy*. Panini, Modena, pp. 163–180.
- O'Malley, M. (2005). *The Business of Art: Contracts and the Commissioning Process in Renaissance Italy*. Yale University Press, New Haven and London.
- Páramo, J.P. (1999). "Notes sobre el taller de Zurbarán: Un envío de lienzos a Portobelo y Lima en el año 1636". In: Vadivieso, E. (Ed.), *Zurbarán. IV Centenario del nacimiento*. Museo de Bellas Artes de Badajoz, Badajoz, pp. 17–31.
- Pezzolo, L. (2003). *Il fisco dei Veneziani*. CIERRE, Sommacampagna, Venice.
- Previer, W., Blockmans, W. (Eds.) (1986). *The Burgundian Netherlands*. Cambridge University Press, Cambridge.
- Raux, S. (2002). "Amateurs d'art et de curiosités à Lille au XVIIIe siècle: Un nouveau éclairage à partir des catalogues de ventes". In: Michel, P. (Ed.), *Collections et Marché de l'art en France au XVIIIe siècle*. Les Cahiers du Centre François-Georges Pariset. Université Michel de Montaigne-Bordeaux 3, Bordeaux, pp. 97–109.
- Raux, S. (2004). "La présence des peintres-marchands brabançons sur les marchés flammands: Similitudes et contradictions des modèles de Lille et de Gand au XVIIe siècle". Mimeo. Université Charles de Gaulle-Lille 3.
- Raux, S. (2005). "Éléments sur les ventes publiques de peinture à Lille au XVIIIe siècle, à partir de la dispersion des collections des chanoines". In: *Collectionner dans les Flandres et la France du Nord au XVIIIe*

- Siècle, Actes du colloque des 13 et 14 mars 2003, Université Charles de Gaulle, Lille-3. CeGes, Villeneuve d'Ascq, pp. 165–182.
- Romein, E., Korevaar, G. (2006). “Guilds in 17th C. Holland and the threat of public sales”. In: De Marchi, N., Van Miegroet, H.J. (Eds.), *Mapping Markets for Paintings*. Brepols, Turnhout, pp. 165–184.
- Schnapper, A. (1994). *Curieux du grand Siècle. Collections et collectionneurs dans la France du XVIIIe siècle*. Flammarion, Paris.
- Schnapper, A. (2001). “La population des peintres à Paris au XVIIIe siècle”. In: *Mélanges en hommage à Pierre Rosenberg. Peintures et dessins en France et en Italie XVIIe–XVIIIe siècles*. Réunion des Musées Nationaux, Paris, pp. 422–426.
- Shaw, J.E. (2006). “Institutional Controls and the Retail in Venice”. In: De Marchi, N., Van Miegroet, H.J. (Eds.), *Mapping Markets for Paintings*. Brepols, Turnhout, pp. 107–122.
- Sluijter, E.J. (2000). “Over Brabantse vodden, economische concurrentie, artistieke wedijver en de groei van de markt voor schilderijen in de eerste decennia van de zeventiende eeuw”. In: Falkenburg, R. et al. (Eds.), *Art for the Market*. In: *Netherlands Yearbook for the History of Art*, 1999, vol. 50, pp. 113–133.
- Sonnino, E. (1998). *Popolazione e Società a Roma dal medioevo all'età contemporanea*. Il Calamo, Rome.
- Spear, R.E. (2003). “Scrambling for Scudi: Notes on painters' earnings in early baroque Rome”. *The Art Bulletin* 85, 310–320.
- Stabel, P. (2000). “Marketing cloth in the low countries: Manufacturers, brokers and merchants (14th–16th centuries)”. In: Stabel, P., et al. (Eds.), *International Trade in the Low Countries*. Garant, Leuven and Apeldoorn, pp. 15–36.
- Stabel, P. (2001). “De gewenste vreemdeling. Italiaanse koopleiden en stedelijke maatschappij in het laat-middeleeuws Brugge”. *Jaarboek voor Middeleeuwse Geschiedenis* 4, 189–221.
- Stabel, P. (2006). “Selling paintings in late medieval Bruges. Marketing customs and guild regulation compared”. In: De Marchi, N., Van Miegroet, H.J. (Eds.), *Mapping Markets for Paintings*. Brepols, Turnhout, pp. 89–103.
- Stabel, P., Blondé, B., Greve, A. (Eds.) (2000). *International Trade in the Low Countries (14th–16th Centuries)*. Merchants, Organisation, Infrastructure. Garant, Leuven and Apeldoorn.
- Syson, L., Thornton, D. (2001). *Objects of Virtue. Art in Renaissance Italy*. British Museum Press, London.
- Szanto, M. (2001). “La stratégie de l'artium amatoris. Les banquiers Lumargue et le commerce parisien de l'art dans la première moitié du XVIIe siècle”. *Nouvelles de l'Estampe* 177, 7–18.
- Szanto, M. (2002). “Libertas artibus restitute. La foire Saint-Germain et le commerce des tableaux, des frères Goetkindt à Jean Valdor (1600–1660)”. In: Cavaciocchi, S. (Ed.), *Economia e Arte*. Le Monnier, Florence, pp. 149–185.
- Szanto, M. (2006). “Antwerp and the Paris art market in the pivotal years 1620–1630”. In: De Marchi, N., Van Miegroet, H.J. (Eds.), *Mapping Markets for Paintings*. Brepols, Turnhout, pp. 329–339.
- Thomas, A. (1995). *The Painter's Practice in Renaissance Tuscany*. Cambridge University Press, Cambridge.
- Van Autenboer, E. (1949). “Nota's over de Mechelse Waternverfschilders”. *Mechelese Bijdragen* 11, 33–44.
- Vanden Houte, C. (1913). “Les corporations des peintres de Bruges”. *Cappel-Missiaeu, Bruges-Kortrijk, Registres d'admission: A: 1453–1578*.
- Van der Woude, A. (1991). “The volume and value of paintings in Holland at the time of the Dutch Republic”. In: Freedberg, D., De Vries, J. (Eds.), *Art in History, History in Art*. Getty Center, Santa Monica, pp. 285–329.
- Van Miegroet, H.J. (2005). “Recycling Netherlandish paintings on the Paris market in the early eighteenth century”. In: Raux, S. (Ed.), *Collectionner dans les Flandres et la France du Nord au XVIIIe Siècle*. CeGes, Villeneuve d'Ascq, pp. 251–287.
- Van Zanden, J.L. (1993). *The Rise and Decline of Holland's Economy. Merchant Capitalism and the Labour Market*. Manchester University Press, Manchester.
- Vermeylen, F. (2003). *Painting for the Market. Commercialization of Art in Antwerp's Golden Age*. Brepols, Turnhout.
- Wackernagel, M. (1981[1938]). *The World of the Florentine Renaissance Artist. Projects and Patrons, Workshop and Art Market*. Princeton University Press, Princeton. (Trans. Fuchs, A.).

- Welch, E. (2003). "From retail to resale: Artistic value and the second-hand market in Italy (1400–1550)". In: Fantoni, M., Matthew, L.C., Matthews-Grieco, S.F. (Eds.), *The Art Market in Italy*. Panini, Modena, pp. 283–299.
- Wilson, J. (1983). "The participation of painters in the Bruges 'pandt' market, 1512–1550". *Burlington Magazine* 125, 476–479.
- Wilson, C., Parker, G. (Eds.) (1977). *An Introduction to the Sources of European Economic History, 1500–1800*. Weidenfeld and Nicholson, London.
- Wright, A. (2003). "Between the Patron and the market: Production Strategies in the Pollaiuolo Workshop". In: Fantoni, M., Matthew, L.C., Matthews-Grieco, S.F. (Eds.), *The Art Market in Italy*. Panini, Modena, pp. 225–236.
- Zablotney, S. (1999). "Production and reproduction: Commerce in images in late eighteenth-century London". In: De Marchi, N., Goodwin, C.D.W. (Eds.), *Economic Engagements with Art*. Duke University Press, Durham, NC, pp. 413–422.

## THE EVOLUTION OF MUSIC MARKETS\*

F.M. SCHERER

*Harvard University Emeritus, USA*

### Contents

Abstract	124
Keywords	124
1. Introduction	125
2. The evolution of music performance venues	125
2.1. Church support	126
2.2. Support from noble courts	127
2.3. The rise of private sector support	129
3. A quantitative perspective	130
3.1. Building human capital	133
3.2. National differences	134
4. Music publishing and intellectual property	136
5. The popularization of musical culture	138
References	142

\* This chapter is based in substantial measure on the author's (2004) book. When other references are lacking, it can be assumed that the book is the source of relevant and additional evidence.

**Abstract**

This chapter traces the development of markets for music over the past several centuries. Formally organized music was at first supported primarily by churches and the nobility. Indeed, during the 18th century there was a kind of “arms race” in which noble houses competed for prestige through the quality of their orchestras and operas. Gradually, however, the role of churches and the nobility declined and more market-oriented venues, including vibrant markets for free-lance performance and composition, became dominant. A survey of 646 musicians born between 1650 and 1849 quantifies these trends, along with the increasing tendency for composers to earn their livelihood in part by teaching in conservatories and other music schools. The same data set is tapped to show the various ways in which composers learned their musical skills. When the data are analyzed in still another way, Austria is found to be the most intensive employer of composer–musicians per million population and also to have given birth to the largest number of composers relative to its population. The territory that in 1990 was Czechoslovakia was second in births; Germany was second in employment. One means by which composers derived income, especially with the ascendance of free-lance composition, was publication of their creations. The evolution of publishing technology and music copyright is traced. Gradually, as income per capita grew in the industrialized nations, audiences for musical performances broadened until by the end of the 19th century mass markets came into being. Radio and the phonograph radically transformed the ways music was enjoyed and widened the market for music even more. The electronic media in turn intensified a tendency that was already evident in the early 18th century – the ability of a few “superstars” to achieve spectacular financial success.

**Keywords**

music composition, music performance, free-lance creative activity, feudalism, human capital

*JEL classification:* Z10

## 1. Introduction

Music has been a vital part of human society for as long as there is recorded history. The oldest known musical instrument, a seven-hole Chinese flute, was carbon-dated in 1999 to the year 7000 BC. Not long after the Israelite exodus from Egypt, Joshua is said in the Old Testament to have breached the walls of Jericho by having seven priests blow ram's horn trumpets while circling the city. In the mythology of ancient Greece, the god Apollo playing a stringed lyre is said to have won a competition, judged by the muses, against the flute-playing satyr Marsyas.

One characteristic making music interesting from the perspective of economics is that it is an activity that eventually came to be market-oriented, giving rise to a complex set of vertically-linked enterprises. Consumers derive satisfaction from hearing music and performing it themselves. When individuals or families perform music, they must be trained to do so, often by specialized teachers offering their services privately or through schools. Most of the music they play is composed by other specialists, whose works reach the players through the intermediation of music publication firms. Music is performed on instruments produced by still another specialized industry. And when consumers enjoy music performed by others, the performances are typically organized by patrons, churches, or concert management organizations, which employ professional musicians either regularly or in a series of ad hoc market transactions. Or in more recent times, the music is supplied by electronic recording and/or broadcasting enterprises.

These various industries together entail a substantial amount of economic activity, only parts of which are systematically measured. In the United States during 1998, 183,000 persons were employed as musicians and/or composers [US Bureau of the Census (1999, p. 242)]. Sales of musical recordings in 1996 amounted to \$12.5 billion; instrument sales were \$1.17 billion; and in 1997, music publishers had sales of \$1.36 billion (*Ibid.*, p. 587). For 45 of the estimated 350 professional classical music-performing orchestras in the United States on which systematic data for the 2001–2002 season are available (usually for the more prominent ensembles), annual budgets totaled \$1.14 billion. The largest single budget, at \$197 million, was that of New York's Metropolitan Opera [American Federation of Musicians (2003)].

This chapter focuses on facets of the music “enterprise” that are either explicitly market-oriented or occur within organizational frameworks, with emphasis on the composition and performance of classical music. This by no means denies the fact that in many nations there are long traditions of folk music passed on informally from generation to generation, or that “popular” music now enjoys much wider market appeal than the classical music emphasized here.

## 2. The evolution of music performance venues

Since the earliest times, markets have gradually come to play an increasingly prominent role in the organization and supply of music. Two of Richard Wagner's operas

deal with alternative modes in which musical activity was organized in Europe during the Middle Ages. Under one mode, skilled minstrels wandered from town to town performing their music and enjoying the hospitality (including financial support) of local barons. *Tannhäuser* embellishes on an actual gathering of German minstrels around the year 1210 (documented in the *Codex Manesse*) at the Wartburg castle near Eisenach (much later, the refuge of Martin Luther and the birthplace of Johann Sebastian Bach). Organized by Duke Hermann von Thüringen, it was a competition among the best-known minstrels, including Wolfram von Eschenbach, Walther von der Vogelweide, and Klingsor from far-off Hungary. In the opera version, the prize is the hand of the duke's daughter Elisabeth in marriage. What the actual prize was remains unclear; it must at least have been Olympian prestige and probably also a monetary reward. Wagner's *Meistersinger von Nürnberg* parodies the amateur singing societies or guilds that sprang up in German towns during the middle ages – in the case depicted by Wagner, in the commercial town of Nürnberg during the early 16th century. Strict rules were enacted for the forms in which songs could be sung, and members of the guild performed in periodic competitions with the goal of winning temporary possession of a silver necklace prize awarded to the entrant who committed the fewest singing errors.

### 2.1. Church support

A venue for music performance with even longer historical roots was the church, or more accurately, the assortment of Christian orthodoxies. Plainchant was a significant but growing component of Roman Catholic services from at least the conversion of the Roman emperor Constantine (312–337). By the eighth century, it became known as Gregorian chant. During the middle ages, song during church services began to be accompanied by an organ, first in monasteries and then in lay churches. The music performed in Roman Catholic churches became increasingly elaborate, culminating in the exquisite antiphony and polyphony of Giovanni da Palestrina during the 16th century. In 1562 and 1563 the Council of Trent considered a mandate requiring music during the Catholic mass to retreat toward monophonic forms that “may be clearly understood by all”. Although guidelines were adopted, rigid rules were opposed by Holy Roman Emperor Ferdinand I, and as a result, subsequent popes chose to decentralize musical decisions to local prelates. One hundred flowers were allowed to bloom, leading *inter alia* to the richly orchestrated masses of Johann Sebastian Bach, Joseph and Michael Haydn, Mozart, and Beethoven.

Meanwhile, Martin Luther, who spearheaded the Protestant Reformation with the posting of his 95 *Theses* in 1517, argued that music, as an “excellent gift from God”, should have a central role in religious services – among other things, in the form of chorales sung by the entire congregation. Within the Protestant denominations there were strong differences of opinion. Music flourished in German Lutheran parishes, as it did within the Church of England. In Switzerland, on the other hand, followers of Huldreich Zwingli eliminated singing and organs from their services, and in England under Puritan Oliver Cromwell (between 1649 and 1658) music was banned from church ser-

vices and organs were removed or even destroyed. The Restoration under Charles II among other things returned music to the churches. Thus, both in England and on the European continent, diverse opportunities opened up for the employment of musicians as choir directors, organists, choir members, and (in the better-endowed parishes) orchestra members. Some of the more important parishes also maintained choir schools that trained many generations of fledgling musicians.

As the Middle Ages gave way to the Renaissance, musicians became less itinerant and began to move from amateur status to regular employment. Churches provided one important locus of long-term employment. European towns also hired musicians to be “pipers”, performing regularly on flutes, trumpets, and drums with local militias and at town celebrations. And wealthy, typically noble, households began retaining musicians as regular members of their staffs, rather than relying upon occasional visits by wandering minstrels.

## *2.2. Support from noble courts*

In many parts of Europe, church and state were closely linked. Kings and local feudal lords were responsible for sustaining church services at their residences, and as such, established chapel musical establishments. The chapel musicians were commonly asked to do double duty performing at festive occasions and in some cases regularly during meals. Gradually a division of labor emerged, with some musicians assigned primarily to chapel duties and others assigned to provide entertainment in the royal household. This duality existed, for example, at Versailles during the reign of Louis XIV (1659–1715).

What happened in central and southern Europe was of the utmost significance for the development of music as a professional activity. Between 1618 and 1648 most of Europe was enmeshed in a devastating Thirty Years War. When the war ended with the Treaty of Westphalia in 1648, what is now Germany was divided into several hundred individual kingdoms, principalities, and dukedoms, some Protestant and some Catholic, each required to yield only minimal deference to the Holy Roman Emperor. In Austria, which also controlled Hungary, what later became Czechoslovakia, and parts of northern Italy, political power was more centralized in the Kaiser-Emperor, but feudal lords governed at the local level. The parts of Italy not under Austrian subjugation were also divided into many smaller political units, some under Spanish or French rule, some under papal domination, some surviving as feudal states under local kings or dukes, and some (such as Venice) as independent city states.

Especially in the German-speaking parts of Europe, populations were decimated by casualties, famine, and disease resulting from the Thirty Years War. Most of the arable land was controlled by feudal lords with noble titles, e.g., king, prince, or duke. With fewer mouths to be fed, land rents at first fell. But as populations were restored following the Treaty of Westphalia, rents gradually rose, and noble landlords profited. Somehow – the exact historical dynamics are uncertain – it became fashionable for local sovereigns and lords to establish their own court orchestras and, in the larger dominions, their own



opera houses. Individual local nobles began competing for prestige through this form of conspicuous courtly consumption. See especially *Elias (1969)* and *Raynor (1972)*. The competition escalated in a kind of cultural arms race, so that by the first half of the 18th century, no self-respecting court in what had been the Holy Roman Empire could be without its court orchestra or other group offering regular musical entertainments.

This proliferation of noble courts with musical ensembles is said to have given rise to a golden age of musical creativity. Augmenting a similar hypothesis advanced by sociologist *Norbert Elias (1991)*, *William and Hilda Baumol (1994)* asserted:

Obviously, economic and political conditions cannot create talent, but they certainly can either inhibit it or provide opportunities for its exercise. Our main hypothesis is intended to narrow the pertinent geography – to help account for the striking level of composing activity emanating from Germany and Italy. This hypothesis suggests that the political division of the Holy Roman Empire and the Habsburg possessions into many petty states worked to produce the circumstances (notably substantial demand and a profusion of jobs) that help to explain the profusion of musical productivity (p. 172).

Although the noble courts provided abundant employment opportunities for musicians and, for those with the creative talent to rise from the instrumentalist ranks, for composers, court employment was not without its disadvantages. Once a musician had secured a court position in what was a fairly competitive job market, he could not leave that position without the lord's express permission (or under some contracts, until the lord's death). Johann Sebastian Bach was imprisoned for nearly four weeks in 1717 when he attempted to break his contract with the Duke of Weimar. Mozart was fearful that if he returned in 1783 to visit his father in Salzburg, he would be imprisoned by Prince-Archbishop Colloredo, whose employ he left without permission two years earlier. On the other side of the ledger, a musician could be discharged summarily at the lord's whim – an experience that led Georg Philipp Telemann to remark that “whoever seeks life-long security must settle in a republic” – that is, in a job not dependent upon feudal lords. In addition to addressing their masters in the most abjectly deferential language, musicians were required to wear the livery (uniform) prescribed by the master. Niccolò Paganini was discharged for his failure to do so in Parma. Among other things, masters stringently limited outside dissemination of works composed by their court musicians. When Nicolò Jommelli left the service of Duke Carl Eugen of Württemberg in 1769 during a budgetary crisis, he was not allowed to take with him copies of the operas and other works he had composed. And discrimination in the wages paid musicians could be extreme. The highest salary Carl Philipp Emanuel Bach (son of Johann Sebastian Bach) received as harpsichordist and composer in the court of King Frederick the Great of Prussia was 500 thaler per year (about £80 at the exchange rates prevailing at the time), while Johann Quantz and Karl Heinrich Graun were paid 2000 thaler and leading opera singers employed by the court were paid as much as 6000 thaler.

During the second half of the 18th century and early decades of the 19th century, noble court support of musical ensembles was substantially reduced. There appear to have

been three main reasons for this change. First, major wars, notably, the Seven Years War between 1756 and 1763 and the Napoleonic wars between 1792 and 1815, drained the treasuries of many local and territorial governments. Second, the “arms race” in conspicuous musical consumption raised the costs of maintaining a suitable musical program to the point at which many courts found themselves unable to keep pace. Third, feudal oppression of the peasants employed on feudal lords’ vast landholdings (encompassing a majority of the continental European population in 1750) lost its intellectual support as a consequence of Enlightenment philosophical arguments and because it proved to be much less efficient than agriculture carried out by well-motivated land-owning smallholders.<sup>1</sup> Emancipation of European peasants was accelerated by revolution – notably, the French revolution of 1789 and the brief revolutions of 1848 – and the abolition of many feudal rights by French occupying forces during the Napoleonic wars.

Feudal reform also reduced the resources available to religious establishments, and especially the Catholic Church, for the support of musical activities. In particular, lands owned by clerical establishments were confiscated in Austria under enlightened Emperor Joseph II during the early 1780s and in France following the 1789 revolution.

### *2.3. The rise of private sector support*

Replacing court and church support of musical activity to some extent was the growth of middle-class citizens’ demand for musical performances and training. The industrial revolution originating in England during the 18th century and spreading gradually to the European continent increased both the number and wealth of middle-class merchants, manufacturers, financiers, and barristers with means sufficient to enjoy the consumption of music in its various forms, notably, through both home performance and participation in public concerts. London, vanguard of the industrial revolution, developed a thriving private-sector concert life during the 18th century.<sup>2</sup> Amsterdam and some of the leading mercantile cities of Germany and Italy (notably, Venice) expanded private-sector music venues in parallel. Other European cities gradually followed suit. For musicians, there were increased opportunities to obtain employment with orchestras organized outside feudal courts, to undertake free-lance virtuoso performances, to give music lessons to the children of an expanding middle class, to compose works that would yield revenue through publication or commissions from private orchestras, and to join the faculties of a growing number of music schools and conservatories. Part of this change entailed a shift from one mode of employment to another under less servile conditions. Part represented a shift from earning a living as an employed musician to market-oriented free-lance activity.

In opening up opportunities for free-lance music composition and performance, opera was the leading-edge sector. The prototype for modern opera, Claudio Monteverdi’s

<sup>1</sup> See Blum (1978) and Delon (2001).

<sup>2</sup> See McVeigh (1993).

*Orfeo*, was performed under the sponsorship of the Duke of Mantua, Italy, in 1607. As enthusiasm for opera spread, some noble courts in Italy, Austria, France, Germany, and Russia emulated the pattern established at Mantua. The sovereign or feudal lord built an opera house, created a managerial structure and participated in its decision-making, and provided regular subsidies to sustain continuing performance schedules. This had two disadvantages: given the competition for prestige among various courts, a tendency toward excessive or even ruinous expenditures; and the inflexibility or stifling of creativity that can come from bureaucratic decision-making.

An alternative originated in the free mercantile city of Venice with the opening of the first public opera house, San Cassiano, in 1637. Under the pattern established in Venice, a group of wealthy citizens joined to contribute funds needed to build an opera house, obtaining in exchange for their contributions preferred boxes and the right to nominate a management committee. The management committee retained an impresario, who, with the advice and consent of the committee, laid out a program of performances, chose libretti for new operas, negotiated with composers ad hoc contracts to provide the necessary musical score in collaboration with the librettist, hired singers, musicians, scenery designers, and the like; and attended to the thousand-and-one details required for successful performances. The impresario also accepted some of the financial risks attending the performances he organized, not infrequently sustaining substantial losses and fleeing creditors in the dark of night. Noteworthy also in this organizational schema is the fact that composers, librettists, and leading singers competed on a free-lance basis to make their contributions. The Venetian system was widely emulated in other parts of Italy and eventually in some German cities (notably, those not controlled by feudal lords), London, Vienna (beginning around the onset of the 19th century), and New York, among other locations.<sup>3</sup>

### 3. A quantitative perspective

How these complex changes evolved over time can be seen through a quantitative analysis, drawn from Scherer (2004), of a sample of 646 music composers born between the years 1650 and 1849. The composers were chosen on the basis of two criteria: for creating works with sufficient survival value to be listed in the *Schwann Opus* catalogue of recorded music for Fall 1996; and by leaving a sufficient historical record that their biographies were published in the *New Grove Dictionary of Music & Musicians* [Sadie (1980)]. The number of composers in each of the four half-century birth date cohorts was 141 for 1650–1699, 148 for 1700–1749, 168 for 1749–1799, and 189 for 1800–1849.

Each composer's biography was coded for various attributes, including places of birth, family background, educational background, how the composers earned a living, and the geographic locations in which they worked during their lifetimes. Figure 1

<sup>3</sup> See Bianconi and Pestelli (1998) and Rosselli (1984).

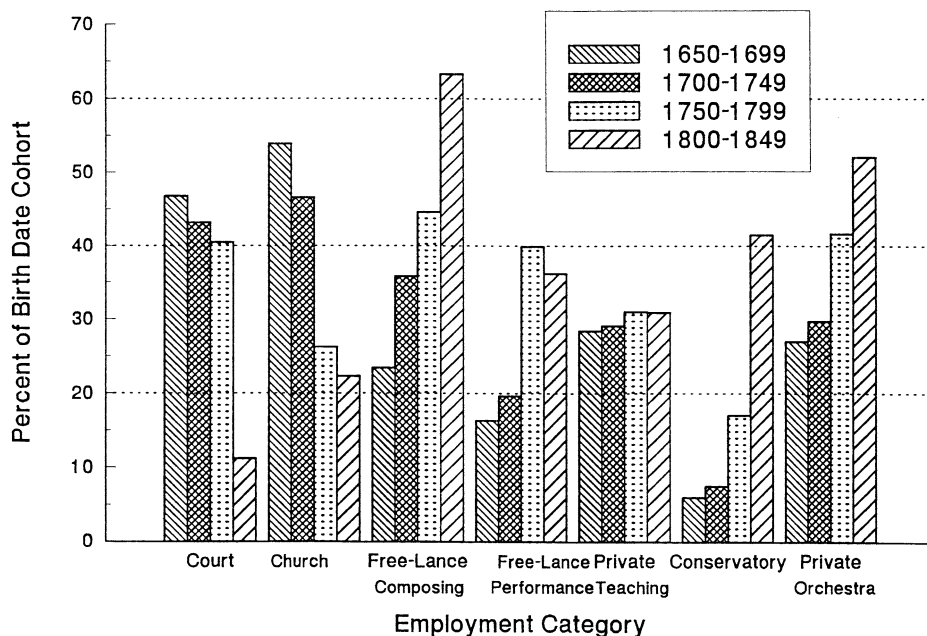


Figure 1. Trends in composer employment over two centuries.

provides broad insight into the ways composers earned their living. Only activities classified as “primary” or “secondary” to a sample member’s occupation are counted. Six categories are traced by 50-year birth date interval:

- (1) employment with noble courts, excluding the receipt of subsidies not conditional upon regular employment duties;
- (2) employment as a church musician or music director, excluding religious duties unconnected with music;
- (3) seeking compensation as a free-lance artist composing music for impresarios, publishers, or performance in self-initiated concerts;
- (4) free-lance performance on self-initiated concert tours or as an ad hoc featured soloist in orchestral concerts;
- (5) teaching in a conservatory or other public music school;
- (6) free-lance teaching (outside the scope of organized music schools and conservatories); and
- (7) participation as either a performer in or director of orchestras unaffiliated with church or nobility.

Needless to say, sample members often pursued more than one of these activities, either at successive career stages (e.g., when Johann Sebastian Bach left his job as Kapellmeister for the Prince of Köthen to become director of music in four Leipzig churches and the Thomasschule), or simultaneously (e.g., when Bach moonlighted

as director of a private concert-giving ensemble, the Collegium Musicum, during his church and school employment tenure in Leipzig). On average, composers born in the 1650–1699 period, as Bach was, were coded as engaging in 2.01 primary or secondary activities, as categorized in Figure 1. For composers born in 1700–1749, the average was 2.11; for 1750–1799, 2.41; and for 1800–1849, 2.57. Thus, composers tended to embrace a wider array of occupational categories with the passing years.

We observe sharp decreases over time in the fraction of sample composers employed in courts and churches, with the most dramatic changes materializing for composers born during the first half of the 19th century for court employment and the second half of the 18th century for church employment. There was a concomitant rise in the fraction engaging in free-lance composing and performance activities. The increase in free-lance composing was fairly steady over time, whereas for free-lance performance, activity peaked for sample members born between 1750 and 1799. Further analysis reveals that most but not all of the 33 composers from the 1650–1699 birth cohort with primary or secondary free-lance composing wrote at least two operas – an activity in which free-lance efforts predominated. There is no clear trend in private tutoring, which ranges from 28.4 to 31.0 percent over the four birth date cohorts. Conservatory teaching rose briskly for composers born between 1750 and 1849. So also did participation in private-sector orchestras.

To extend the sample into later periods, a much smaller and less representative sample of ten classical music composers born after 1850 was drawn. With birth and death dates in parentheses, it includes Edward Elgar (1857–1934), Gustav Mahler (1860–1911), Arnold Schoenberg (1874–1951), Ottorino Respighi (1879–1936), Béla Bartók (1881–1945), Paul Hindemith (1895–1963), Aaron Copland (1900–1990), Kurt Weill (1900–1950), Dimitry Shostakovich (1906–1975), and John Cage (1912–1992). The composers in this “modern” sample are uniformly more prominent than the median composer in the earlier, larger sample of 646. When the biographies were coded using criteria matching as closely as possible those applied to the larger sample, only one composer is found to have been employed by “nobility” – Shostakovich, supported by agencies of the Soviet Union such as the Union of Soviet Composers. None had significant church employment. All engaged in significant free-lance composition. Four were free-lance performers; six performed in and/or directed private-sector (including municipal) orchestras; at least five gave private lessons; and eight taught for substantial periods of time in conservatories or their modern equivalents, the music departments of universities. Confirmed with particular strength are the tendencies toward declining noble and church employment and the increase in free-lance and conservatory teaching activity.

To sum up, over a period of two-plus centuries, there was a transition from support of music composers – the elite of professional musicians – by churches and the nobility to support through more market-oriented institutions, and especially free-lance activities. The changes were gradual and evolutionary, and not linked closely to particular composers such as Mozart and Beethoven, as some students of musical history have suggested.

### 3.1. Building human capital

From the sample of 646 composers born between 1650 and 1849, insights can also be gleaned on how the sample members learned their trade. For 532 composers, there was sufficient historical information to identify principal learning modes. In many cases several forms of learning were pursued by a single individual; thus, young people might initially be tutored by a family member, then move to instruction from an unrelated professional musician, and then, at least in later periods, receive formal conservatory training. Altogether, 1.83 codes were assigned on average per sample member.

Figure 2 summarizes the results. Within-the-family tutoring was evident in nearly half of all cases, and, because of gaps in the historical materials, its incidence is undoubtedly underestimated. By far the most ubiquitous form of human capital building was private instruction with a professional musician. Although no attempt was made to assess the quality of named teachers, the high incidence of well-known names was striking. Talent is attracted to talent, which in turn leads to professional success. As conservatories were established throughout Europe, there was a sharp rise in formal conservatory training for musicians born between 1800 and 1849. Church choirs and affiliated choir schools provided training for more than a quarter of the sample members born during the second half of the 17th century, but their role faded as the years advanced. Especially in Czechoslovakia and Germany, as defined according to 1992 borders, the primary and secondary school curricula included substantial musical education. The self-teaching

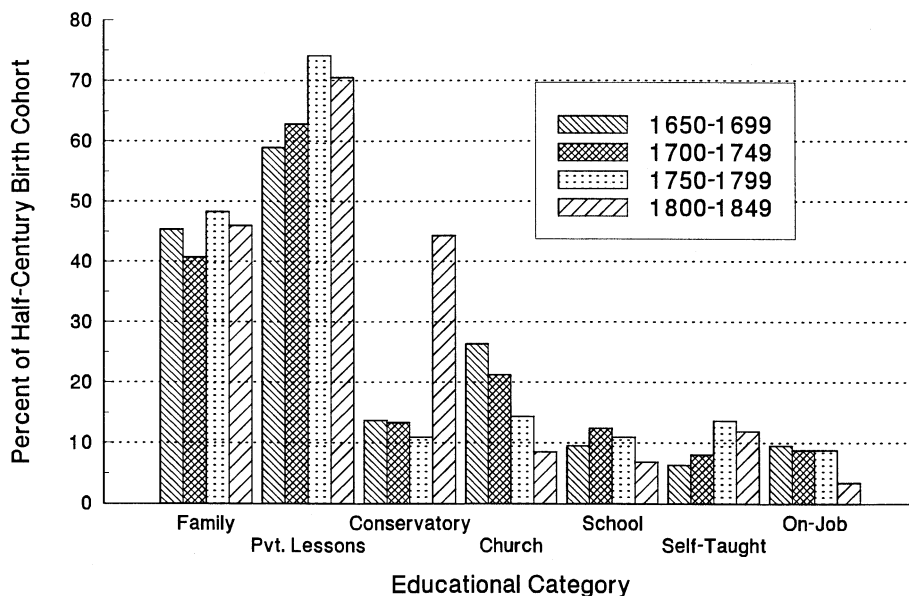


Figure 2. Percent of half-century cohorts with diverse career preparation modes.

and on-the-job-training categories were coded conservatively, counting only cases in which no other training was evident at a formative period or in which a youth joined a court orchestra long before reaching maturity.

### 3.2. *National differences*

Classical music composition during the 18th and 19th centuries was preponderantly a European enterprise. Among the sample of 646 composers born between 1650 and 1849, only 17 were born outside Europe, notably, in North and South America. Figure 3 shows for the principal European nations and national groups, as defined by borders existing in 1992, the number of composers born and employed per million population. The raw counts are divided by national population estimates at the time of birth for birth data and at age 35 for the employment counts. The employment counts are for “working life equivalents”. That is, for any given sample member, the number assigned to a nation in which the composer worked for two years or more was the fraction of the composer’s life from age 20 on spent in that nation.

Austria – home to Mozart, Haydn, Schubert, and (in his mature years) Beethoven, among others – led other nations in both composer births and working-life-equivalent years of residence relative to its (modest) population. The other remnants of the Holy Roman Empire – Germany, Italy, and Czechoslovakia – were the next most prolific producers of composers relative to their population. Russia, with a huge serf population and a particularly small class of noble and middle-class music consumers, trails far behind on both dimensions. Germany, with the richest proliferation of independent noble courts, was second in employment of composers per million residents.

Dividing the employment counts into the birth counts, one can see which nations were the most vigorous importers of composer–musicians and which were exporters. Austria leads the list of importers, with 1.66 equivalent composer working lives per birth; England, with wealthy London as a magnet to immigrant musicians, was second; France with Paris as a magnet was third. Czechoslovakia was in relative terms the leading exporter, followed by Italy. Czechoslovakia’s inability to retain its numerous home-grown composers was the result of both demand and supply-side forces. On the demand side, its leading city, Prague, was small relative to most other European nations’ capitals, and many Bohemian nobles spent the most fashionable seasons, when music performance was at its peak, at their town houses in Vienna, the imperial capital. On the supply side, there is probably truth in Hector Berlioz’ observation that “I must say – since it is a matter of public notoriety – that the Bohemians are the best musicians in Europe, and that the love and feeling for music are universal in all classes of society” [Berlioz (1966, p. 414)].

Apart from this cultural factor, which cannot readily be measured, three main hypotheses contend to explain the intensity with which diverse nations or national groups attracted composers into employment. More employment per million population is expected:

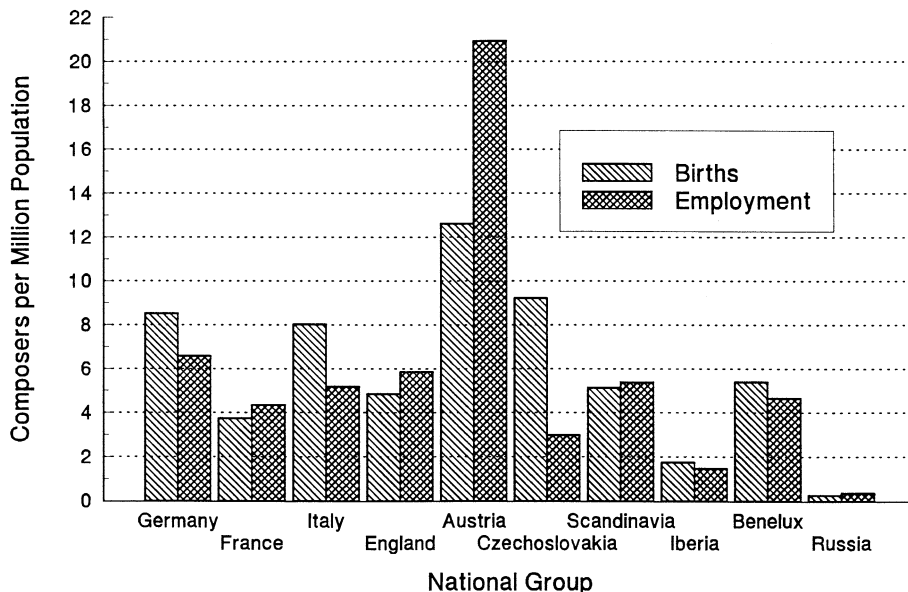


Figure 3. Births and employment of composers per million population.

- (1) in the four remnant nations of the Holy Roman Empire, especially before the decline of feudal courts as the 19th century dawned;
- (2) in nations such as England, France, and Austria with rich magnet cities; and
- (3) in richer nations, as measured by gross domestic product per capita.

Where WORKLIVES is the measure of equivalent working lives per million population plotted in Figure 3; GDP is (crudely estimated) gross domestic product per capita, measured in 1990 dollars; COURT is a zero-one dummy variable with unit values for Germany, Austria, Czechoslovakia, and Italy for each of the first three half-century periods; and MAGNET is a zero-one dummy variable for nations with magnet cities (with Russia added only for the fourth half-century interval); the resulting regression equation was as follows:

$$\begin{aligned}
 \text{WORKLIVES} = & -0.29 + 0.00080 \text{ GDP} + 1.127 \text{ MAGNET} + 2.497 \text{ COURT}; \\
 & (0.41) \quad (1.41) \qquad (2.70) \qquad (5.58) \\
 R^2 = & 0.505; \quad N = 40.
 \end{aligned}$$

T-ratios are given in subscripted parentheses. All three variables have the expected signs and help explain variations in composer employment per million population across four half-century intervals in ten nations or national groupings.



Russia began to absorb western European musical culture on a significant scale during the reign of Catherine the Great (1762–1796). St. Petersburg and Moscow attracted numerous traveling musicians in the first part of the 19th century, during which time a Russian school of composition also grew under the leadership of Mikhail Glinka.

The United States was an even later arrival, but arrive it did. By 1850, New York City had a larger population, excluding suburbs isolated by water barriers, than any European city excepting London and Paris. The New York Philharmonic Society began a regular concert series in 1842. An opera series was established at the New York Academy of Music in 1854. The Academy was displaced following the success of the Metropolitan Opera, whose creation was precipitated in 1883 by a dispute between old and new wealthy factions over the allocation of boxes.<sup>4</sup> Many prominent European music performers, composers, and directors were attracted through the high salaries offered by New York organizations and the orchestras of other leading American cities. At the beginning of the 20th century, as the United States moved ahead of England in having the world's highest average gross domestic product per capita, most of the musicians in the Metropolitan Opera orchestra were from Europe, and rehearsals were conducted in a combination of German and English.<sup>5</sup> But as formal musical training expanded with the founding of the Boston Conservatory in 1867, the Michigan University School of Music in 1892, and the forerunner of the Juilliard School in 1905, among others, the United States began to generate its own indigenous supply of musicians and moved into the forefront of world musical culture.

#### 4. Music publishing and intellectual property

One way music composers could earn free-lance income was by selling their works to publishers or, when the music publishing industry was still in its infancy, engaging in self-publication. In the late middle ages, church music was printed using fixed wooden blocks into which staves, notes, and other symbols were carved. Adapting Johannes Gutenberg's 15th century movable type invention to music printing was difficult because music notation is much more complex dimensionally than simple letter text. Improvements in printing music with movable type accumulated over a period of three centuries. During the 18th century, printing music engraved onto copper or pewter plates was the dominant technology, and there was a proliferation of music publishing houses, a few of which (e.g., Breitkopf & Härtel, Schott, and Ricordi) survived into the 21st century. In the 19th century, lithography appeared as a viable lower-cost alternative to typeset and engraved printing. High-speed rotary presses and computerized note-setting have reduced relative costs even more.

Through much of this formative period, hand copyists provided important competition to music publishers. Opera houses retained copyists to produce parts for the various

<sup>4</sup> See Fiedler (2001).

<sup>5</sup> See again Fiedler (2001, p. 16).

ensemble members, and composers employed copyists to make copies for distribution to patrons and potential publishers as well as for the musicians performing new compositions. Under the traditions prevailing in Italy during the 18th century, the impresario for a new opera had the right to the composer's original score and the copyist to the first copy. The copyists then made additional copies which they sold to provincial and foreign opera houses. When composers hired copyists, it was not uncommon for the copyist surreptitiously to make an additional copy, from which they then derived additional copies for sale to a network of music lovers. Noble courts made copies of their composers' works for exchange with other courts in order to expand their performance repertoire, and once a copy reached some other domain, it was difficult to prevent further copying and circulation. The copies that proliferated in this way were often riddled with errors, but the business thrived.

Music publishers often obtained musical manuscripts through these indirect channels and published a printed version without the composer's permission. Absent effective copyright protection, publishers also pirated for distribution in their home market the works published by their peers in other territories. Rampant piracy limited the revenues composers could derive through authorized publication of their creations.

Music composers were not without some rights to the works they created. Plagiarism, i.e., claiming another person's creation as one's own, was widely condemned. In the 16th through 18th centuries, the principal opportunity for obtaining more secure intellectual property rights in one's musical creation came through the granting of "privileges" by national or local governments. A composer could appeal to the sovereign for an exclusive privilege to publish works within the relevant domain, or publishers could obtain exclusive privileges to publish certain classes of musical compositions. To be sure, political connections and/or bribes might be necessary to obtain a privilege. However, Hansjörg Pohlmann (1962, p. 186 ff.), the leading student of the history of musical property rights, argues that in some parts of Europe, the privilege system was sufficiently open and transparent that it resembled a modern copyright system. Yet even when privileges could be obtained, they were difficult to enforce. And when a geographic area was fragmented into many independent political entities, as Germany was following the disintegration of the Holy Roman Empire, a privilege issued, say, in Mainz had little or no jurisdictional reach to prevent the piracy of a musical work in Leipzig, Munich, or Hamburg, let alone in non-German-speaking lands.

England was the first nation to enact a modern copyright law, the Law of Anne, in 1709. Its application to musical works was first clarified in 1777 through a law suit brought by Johann Christian Bach, the son of Johann Sebastian Bach. France replaced its privilege system – probably the most effectively functioning one in Europe – with a copyright law following the revolution of 1789. The United States passed a copyright law for the protection of domestic residents in 1790; under it, the works of foreign authors and composers could still be pirated legally. In Germany, Austria, Czechoslovakia, and Italy – as we have seen, the most intensive centers of musical composition – copyright law was slower to come. In 1825 a petition drafted by composer Johann Nepomuk Hummel, urged on among others by Ludwig van Beethoven, asked the assembly

of German-speaking nations to overcome their differences and enact uniform copyright laws to prevent publishers from “getting fat by robbing without penalty their neighbors’ property”. The petition is reproduced in Benyovszky (1934, pp. 131–133). It spurred the formation of a committee to study the problem, and eventually, in the late 1830s, most of the German states passed copyright laws protecting *inter alia* musical works. The Austrian version covered Austrian possessions in northern Italy as well as Czechoslovakia, and in 1840 Austria entered into a mutual copyright accord with the Kingdom of Sardinia (including Savoy), so that the opportunity to secure musical copyright extended to much of Italy, excluding the Kingdom of Naples and the Papal states. The Berne Convention of 1887 allowed composers (and authors) to obtain copyright in any signatory nation, so in effect, the copyright system became world-wide. In the United States, a relatively late signatory, the number of musical copyright applications rose from 9132 in 1890 to 29,151 in 1920 and 52,309 in 1950. In 1950, there were nearly as many copyright applications for musical works as there were for periodical prose works.<sup>6</sup>

Giuseppe Verdi was the first composer of note to derive substantial advantage from the Austrian–Sardinian copyright regime. With his publisher Giovanni Ricordi, he began collecting opera performance fees from the numerous provincial opera houses, at first on a flat fee basis and then, when Ricordi recognized that price discrimination would yield higher revenues, on a graduated “ability to pay” basis. They also aggressively published “reductions” of individual opera arias for voice and piano, solo piano, violin, flute, string quartet, and many other combinations, selling them to the countless opera-loving families of Italy. Through these activities Verdi became quite rich, and as his wealth grew, he reduced the pace of his opera-writing efforts, from 14 operas in the decade of the 1840s to seven in the 1850s, two in the 1860s, and one each in the succeeding three decades. From Verdi’s letters emerges introspective evidence that his was in effect a backward-bending labor supply curve.<sup>7</sup> It cannot be ruled out, however, that the fortune Verdi accumulated by selling his works had a demonstration effect, inspiring others to try their hand at composing.

## 5. The popularization of musical culture

Up to the onset of the 18th century, paying to hear music performed was a pastime enjoyed mainly by the nobility and the wealthiest members of a relatively small middle class. Gradually, however, increasing prosperity plus technological and organizational innovations broadened audiences and ultimately engendered nearly universal public access to professionally performed music.

One event that signaled a change in audiences for music was the London premiere in 1728 of John Gay’s *Beggar’s Opera*, an irreverent pastiche of already-popular and

<sup>6</sup> See US Bureau of the Census (1960, p. 606).

<sup>7</sup> See Scherer (2004, Chapters 4 and 7).

new English-language songs that enjoyed 62 performances during its first season and was repeated in every London season for the next seven decades. (The plot was later adapted successfully in the Brecht–Weill *Threepenny Opera*.) Several public gardens were opened beginning in the 1730s, presenting open-air concerts during the summer and charging admission sufficiently modest that, according to a contemporary 1786 account:

There were last night above 6000 persons present, among them some of the first people in the kingdom, but as is always the case at Vauxhall, it was a *melange*; the cit and the courtier jostled each other with the usual familiarity; the half guinea was no repellant to the middling order; John Bull loves to shoulder his superiors in rank . . . and where he pays as much for admission, he never considers them to be more than his equals [quoted in McVeigh (1993, p. 41)].

In Vienna during the early decades of the 19th century a counter-part to Vauxhall Gardens emerged through ball rooms in which members of the middle classes waltzed to the music of Joseph Lanner and Johann Strauss Sr. At the peak of Strauss' popularity, these venues were expanded so that they could accommodate 50,000 participants simultaneously. By 1830, Strauss had 300 musicians under contract, whom he deployed in groups of 25 to play in the various ball rooms.<sup>8</sup>

In Paris another means of providing music to large numbers of listeners, “the monster concert”, was pioneered by Hector Berlioz. Berlioz' first such concert, in 1844, was held in an industrial exhibition shed, attracting an audience of 8000 to hear music performed by a thousand instrumentalists and singers.<sup>9</sup> A later monster concert organized by Berlioz in 1855 attracted an audience estimated at 40,000. The pinnacle of 19th century monster concerts was probably reached in the Boston (Massachusetts) Peace Jubilee of 1872. For it, Johann Strauss Jr. was paid \$100,000 plus traveling expenses to direct an orchestra of 2000 and a chorus of 20,000 performing Strauss compositions before audiences estimated at 100,000 persons [Fantel (1971, pp. 194–195)].

The performance of music at home provided a market for instruments and sheet music. Julia Moore (1987, pp. 50–64) punctures the myth that in music-loving Vienna during the time of Beethoven, there was a piano in every house, showing that the original source claimed a piano in the home of every educated family – a major difference in a time when higher education was rare. However, as education spread and prosperity increased, so also did the presence of pianos in homes. Using estimates by Ehrlich (1990, p. 222) and then US Census of Manufactures data, Figure 4 traces trends in new piano production in the United States from 1850 through 1930. Solid circles mark the years for which production estimates are available; the connecting lines are interpolated. Piano production did grow rapidly during the first seven decades, at an average rate of 4.3 percent per year. Scherer (2004, Chapter 2) estimates an income elasticity of piano

<sup>8</sup> See Fantel (1971, pp. 36 and 194–195).

<sup>9</sup> See Holoman (1989, pp. 308–311 and 476–477).

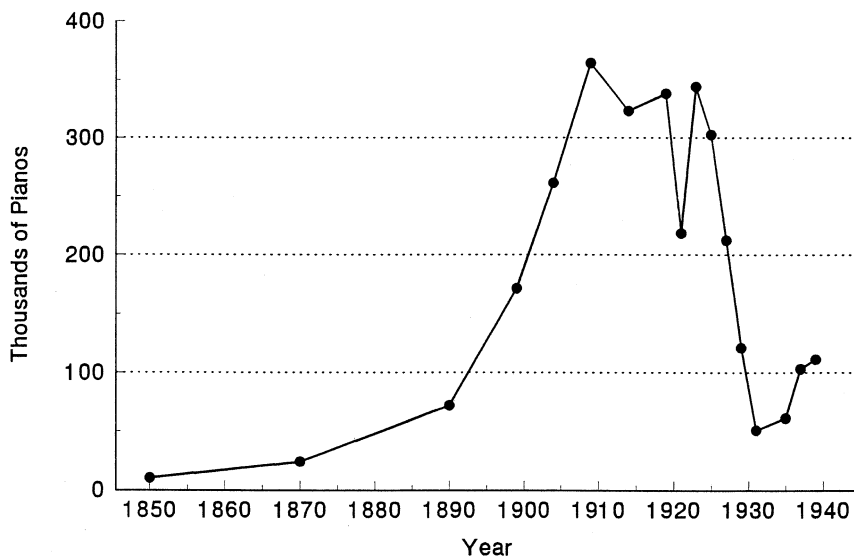


Figure 4. New piano production in the United States, 1850–1939.

demand over this period in the range of 3.16 to 4.32, depending upon the econometric specification used. Transforming the production data into cumulative (but depreciating) stocks, it is estimated that in 1850, the maximum home ownership rate (assuming no ownership by schools, churches, and taverns) was 5 percent. By 1923, it peaked at approximately 23 percent.

The steep decline in piano production during the mid-1920s, a time of extraordinary prosperity, cries out for explanation. (The earlier decline in 1921 was attributable to a sharp but brief recession.) The answer is simple: home radios entered the market in the early 1920s, followed by the introduction in 1924 of electrical phonographs with fidelity far superior to that of the acoustical phonographs previously in use. By 1927, 10 million American homes had radios, and a roughly equal number owned phonographs. A new different way of consuming music (and much else) in the home had appeared, changing radically the character of American family life and making music available every day to families that otherwise would have attended public concerts only rarely.

Electronics also changed the criteria for success as a musical performer. For vocalists in particular, power in an amplifier substituted for the previously essential vocal cord power.

During the 1920s it was believed that radio would allow wider dissemination of classical music, educating and upgrading public tastes. There was in fact an abundance of classical music programming. But especially in the United States, where radio transmitters were more dependent upon advertising than were the publicly-owned stations of Europe, sponsors soon found that broadcasting popular music was more effective in attracting the mass audiences they sought. The weekly Guy Lombardo show was

an early success beginning in 1928. By the late 1930s, the Lucky Strike Hit Parade, which offered renditions of the week's ten best-selling tunes, drew one of the largest radio audiences. As additional channels became available with the introduction of FM broadcasting following World War II, radio stations began to specialize in the content provided. Among the nearly five thousand US stations specializing in music broadcasts during 1994, the largest number (2642) emphasized country music – many, to be sure, in remote rural areas with small populations. There were 44 classical and fine arts specialists [*World Almanac* (1999, p. 186)].

Phonograph records and later compact disks also tapped mass markets for all kinds of music. The first year in which more than one billion musical recordings (including a relatively new genre, music video tapes) were sold in the United States was 1994. In that year, rock and roll music, which achieved popularity in the 1950s, led the market with 35.1 percent of unit sales, country music was second at 16.3 percent, and “pop” music placed third at 10.3 percent. Classical music records had a 3.7 percent share [*World Almanac* (1999, pp. 187–188)]. The share of classical record sales declined over the years, e.g., from 25 percent in 1950 and 5 percent in 1980, in part because each new record must compete with the established repertoire accumulated over three centuries, much of which had already found its way into record buyers' libraries.<sup>10</sup> In the more popular forms of music, the life span of the typical new composition is short, but there is a constant proliferation of new varieties seeking to differentiate themselves from previously recorded tunes.

The decline in classical music market shares parallels another more ominous alleged trend – the “graying” (i.e., aging) of audiences for live classical music performances, combined with a tendency for music directors to force upon audiences *avant garde* music appreciated by few.<sup>11</sup> However, a 1997 survey yielded optimistic signs [*US Bureau of the Census* (1999, p. 275)]. On average, 16 percent of the surveyed Americans had attended at least one classical music performance during the preceding year. The highest attendance rate by age category – 20 percent – was for persons 45 to 54 years old. Respondents in the 18–24 year old category, however, had the same 16 percent rate as the entire population, suggesting a potential for future audience-building. The lowest (11 percent) rate was for persons 25 to 34 years old. Not surprisingly, attendance at live classical music performances rises with income per capita and education, reaching maxima of 35 percent for those with annual incomes above \$100,000 and 45 percent for persons who have attended graduate school.

The ability of musical composers and performers conveniently to reach a market encompassing the entire world intensified a phenomenon already evident three centuries earlier: superstardom.<sup>12</sup> At the outset of the 18th century, the leading musical superstars were castrati singers, who were in demand throughout western Europe and who

<sup>10</sup> See Arnold (1997).

<sup>11</sup> See, e.g., Holland (2003).

<sup>12</sup> See especially Rosen (1981) and Hamlen (1991).

traveled to wherever the rewards were highest. Before he retired to Spain and became chamber musician and confidential advisor on matters of state to Kings Philip V and Ferdinand VI, castrato Farinelli (whose given name was Carlo Broschi) earned £5000 during the 1735–1736 opera season in London, the most lucrative market, at a time when English building craftsmen averaged £30 per year.<sup>13</sup> A century later Niccolò Paganini, considered the most brilliant violinist of all time, netted £10,200 from six concerts in London after a spectacularly successful tour through Continental Europe. The most successful classical music superstars of the late 20th century were the Three Tenors – José Carreras, Plácido Domingo, and Luciano Pavarotti. Their concert during the 1994 World Cup football championship games in Los Angeles attracted a “live” audience of 56,000 persons, was broadcast throughout the world on television, and yielded record sales of some 9 million units. For a follow-on concert at the 1998 World Cup playoffs in Paris, the Three Tenors received an advance payment against anticipated broadcast and record sales royalties of \$18 million. Their receipts paled, however, in comparison to those of the most successful popular music stars. The best-selling musical record of all time is said to be Michael Jackson’s album, “Thriller”, which from its launch date in 1992 achieved sales of 46 million copies. In 2002, Jackson’s accumulated financial net worth was estimated to be in the range of \$250 to 350 million. But by that time Jackson’s reception from the public was fading visibly. Similar popularity declines befell most superstars during the heyday of classical music composition – that is, in the 18th and 19th centuries. But Jackson, unlike his golden-throated predecessor Farinelli, is unlikely to find a second career as advisor to national sovereigns. In that sense too, conditions change.

## References

- American Federation of Musicians of the United States and Canada (2003). *Wage Scales and Conditions in the Symphony Orchestra*. New York.
- Arnold, D. (1997). PolyGram Classics, Case Study 9-598-074. Harvard Business School, Boston.
- Barbier, P. (1996). *The World of the Castrati*. Souvenir Press, London.
- Baumol, W.J., Baumol, H. (1994). “On the economics of musical composition in Mozart’s Vienna”. *Journal of Cultural Economics* 18, 171–198.
- Benyovszky, K. (1934). *J.N. Hummel: Der Mensch und Künstler*. Eos, Bratislava.
- Berlioz, H. (1966). *Memoirs of Hector Berlioz from 1803 to 1865*. Dover, New York. Edited by E. Newmann; translated by R. Holmes and E. Holmes.
- Bianconi, L., Pestelli, G. (Eds.) (1998). *Opera Production and Its Resources*. University of Chicago Press, Chicago.
- Blum, J. (1978). *The End of the Old Order in Rural Europe*. Princeton University Press, Princeton.
- Delon, M. (Ed.) (2001). *Encyclopedia of the Enlightenment*. Fitzroy Dearborn, Chicago.
- Ehrlich, C. (1990). *The Piano: A History*, rev. ed. Clarendon Press, Oxford.
- Elias, N. (1969). *Die höfische Gesellschaft*. Luchterhand, Darmstadt.
- Elias, N. (1991). *Mozart: Zur Soziologie eines Genies*. Suhrkamp, Frankfurt.

<sup>13</sup> See Barbier (1996).

- Fantel, H. (1971). *Johann Strauss: Father and Son and Their Era*. David & Charles, Newton Abbott, UK.
- Fiedler, J. (2001). *Molto Agitato: The Mayhem behind the Music at the Metropolitan Opera*. Doubleday, New York.
- Hamlen, W.A. (1991). "Superstardom in popular music: Empirical evidence". *Review of Economics and Statistics* 73, 729–733.
- Haloman, D.K. (1989). *Berlioz: A Musical Biography of the Creative Genius of the Romantic Age*. Harvard University Press, Cambridge.
- Holland, B. (2003). "How to kill orchestras". *New York Times*, June 28, 2003, Section 2, pp. 1 and 33.
- McVeigh, S. (1993). *Concert Life in London from Mozart to Haydn*. Cambridge University Press, Cambridge.
- Moore, J.V. (1987). *Beethoven and Musical Economics*. Ph.D. Dissertation. University of Illinois, Urbana-Champaign.
- Pohlmann, H. (1962). *Die Frühgeschichte des musikalischen Urheberrechts*. Bärenreiter, Kassel.
- Raynor, H. (1972). *A Social History of Music from the Middle Ages to Beethoven*. Schocken, New York.
- Rosen, S. (1981). "The economics of superstars". *Journal of Political Economy* 71, 845–857.
- Rosselli, J. (1984). *The Opera Industry in Italy from Cimarosa to Verdi*. Cambridge University Press, Cambridge.
- Sadie, S. (Ed.) (1980). *The New Grove Dictionary of Music & Musicians*, 20 volumes. Macmillan, New York.
- Scherer, F.M. (2004). *Quarter Notes and Bank Notes: The Economics of Music Composition in the 18th and 19th Centuries*. Princeton University Press, Princeton.
- US Bureau of the Census (1960). *Historical Statistics of the United States: Colonial Times to 1957*. US Government Printing Office, Washington.
- US Bureau of the Census (1999). *Statistical Abstract of the United States*. US Government Printing Office, Washington.
- World Almanac and Book of Facts: 2000 (1999). World Almanac Books, Mahwah, NJ.



## DEFINING CULTURAL AND ARTISTIC GOODS

ROGER MCCAIN

*Drexel University, Philadelphia, USA*

### Contents

Abstract	148
Keywords	148
1. Introduction	149
2. Definitional issues	150
2.1. Culture	150
2.2. Economic and non-economic values	150
2.3. Cultural value	152
2.4. Provenance	154
2.5. Artistic value	155
3. Creativity	156
3.1. Simon on creativity	156
3.2. Generative linguistics	157
3.3. Creativity in consumption	160
4. Implications for consumers of art	161
4.1. Learning-by-consuming	161
4.2. Artistic goods as stimulus goods	163
4.3. Intellectual property	164
5. Conclusion	164
References	165

**Abstract**

This essay addresses the definition of artistic and cultural goods by the commonsense and pragmatic assertion that they are respectively goods that carry artistic and cultural (non-economic) values. However, these categories of non-economic value are themselves highly contested and require clarification for the definitions to be viable. The essay suggests an interpretation of the distinction of economic and non-economic values, and of non-economic cultural value, by drawing on the ideas of Nozick. For artistic value, the essay argues that creativity is a central concept, and explores the learning on creativity to be found in cognitive science. Cognitive science sees creativity as requiring both novelty and appropriateness to a particular cultural community. These insights underline some of the characteristics of markets for creative goods, including artistic goods, that cultural economists have stressed in their work independently of cognitive science. For artistic goods in particular, consumers join in the creativity of the artists, and this joint creativity enables us once again to invoke Nozick's concepts of intrinsic value and identify artistic value as a particular instance of non-economic value. The definitions are defended as coherent in the light of their consistency with economic thinking on stimulus goods and learning-by-consuming. Intellectual property is seen as overlapping but not necessary to artistic and cultural goods. Nothing in this essay should be taken as final, but, rather, as a possible starting point for a substantive discussion of non-economic values in these fields.

**Keywords**

cultural goods, non-economic value, creativity, learning-by-consuming

*JEL classification:* H11, H31, Z11

... we should define [terms], so as to agree with the sense in which they are understood in [the] ordinary use of them. When the sanction of this authority is not attainable, on account of further distinctions being required, the next best authority is that of some of the most celebrated writers in the science, ... It is acknowledged, however, that a change may sometimes be necessary; and when it is, the natural rules to be attended to seem to be, ... That the alteration proposed ... on the whole be obviously more useful in facilitating the explanation and improvement of the science. [Malthus, *Definitions in Political Economy* (1827)]

## 1. Introduction

Cultural or artistic goods bring to mind the famous comment about pornography. The appropriate definition of “cultural goods” or “artistic goods” is far from obvious. On the one hand, there may be common underlying, or interconnecting phenomena that lend themselves to definition of a range of commodities reasonably called “cultural” or “artistic”. On the other hand, not everything can be stuffed into the definition. Some associations have to be left for empirical demonstration<sup>1</sup> and some may simply be fallacious. It may be that common usage is inconsistent, so that no definition can account for all of it, or a definition intended to correspond to common usage might be so broad (though consistent) and consequently so vague as to be useless. A definition must be definite. It seems fair to say that we have no definitions for cultural or artistic goods that are authoritative in either of Malthus’ first two senses, so a definition of these categories of goods will have to be justified by “facilitating the explanation and improvement of the science”.

For the “improvement of the science”, empirical reliability is one important criterion of success, but coherence is also important. Like value, coherence is “not just some vague laudatory term”.<sup>2</sup> As I use the term, a discourse is coherent if it is consistent with (though not necessarily deducible from) a compact set of fundamental propositions. One way to show that a discourse is coherent is to display the set of fundamental propositions and show that other propositions in the discourse are deducible from the fundamental set, with a few auxiliary assumptions that may vary from case to case. Neoclassical economics is highly coherent in that sense, and it seems that economists place a high value on coherence. In any case, one objective in defining terms is to assure that the terms admit of a coherent discourse, and a major objective of this essay will be to argue that the definitions offered here do that.

<sup>1</sup> This reflects the methodological stance known as operationalism. Bridgeman (1952) observes that, in order for empirical tests to be performed, there must be at least two distinct sets of operations that may or may not be correlated. To attempt to include all possible associations in the definition would prejudice the empirical issues.

<sup>2</sup> The quotation is from Nozick (1989); see further below.

## 2. Definitional issues

### 2.1. Culture

We might begin by adopting a consensus definition of “culture”. However, the word “culture” presents yet another complex of difficulties of definition. In the broadest anthropological sense, culture is more than “high culture” and “popular culture” taken together. Rather, culture in this sense encompasses everything that people derive from their tradition and heritage, including folklore and kinship patterns, “material culture”, religion, and so on. In what follows, the word “culture”, otherwise unqualified, will be used in this anthropological sense, although it is an open question whether there is a definition of culture that is applicable across cultures. Kroeber and Kluckhohn (1963) devote a large book – and the power of their knowledge of anthropology, the human science that focuses on the study of culture – to the search for a viable definition of culture in the anthropological sense. Economists might well hesitate to retrace their steps.

How then may we distinguish artistic and cultural goods and services from other goods and services? It would be convenient if we could say that cultural goods are goods that carry cultural values, and that artistic goods are goods that carry artistic or aesthetic values. These conceptions of values are, if anything, even more contested. Nevertheless, that will be the strategy of this essay. Because of the contested nature of all the relevant issues, the essay is unavoidably rather speculative, but is offered with the hope that it will provide some “fit”. While that “fit” cannot be finally demonstrated, an attempt will be made to show that some of the ideas and concerns with long histories in cultural economics are unified by the proposed definitions.

### 2.2. Economic and non-economic values

“Economic values” as discussed in neoclassical economics are derived from the preferences of individuals. The claim that there are non-economic values distinct from economic values implies that the non-economic values are intrinsic or objective in the sense that they are independent of individual preferences. Critics sometimes assert that neoclassical economists neglect non-economic values, focusing only on economic values. This is a misunderstanding. In a passage many neoclassical economists regard as definitional, Lord Robbins (1952, p. 30) writes “Aesthetics is concerned with certain kinds of ends. The beautiful is an end which offers itself for choice in competition, so to speak, with others. Economics is not concerned at all with any ends as such”. From this point of view, there is no difficulty with non-economic values; economics has to do not with any particular realm of values, but with the balancing of different values and different realms of value.

But this view is an obstacle to communication between economists and others interested in cultural and other forms of public policy. It does seem that the issue here is

pragmatic:<sup>3</sup> advocates of active cultural policy pose their proposals in terms of a distinction between economic and cultural values. If economists cannot speak to such a distinction, then economists cannot contribute to the solution of the problems posed. The difficulty, then, is to construct a scheme in which one can meaningfully distinguish economic from non-economic values.

As a tentative basis for a distinction, consider the ideas of Robert Nozick, a social philosopher known for his free-market views and his affinity to neoclassical economics. Nozick (1981, pp. 162–164) writes:

The notion of value is not just some vague laudatory term. Some things have value only as a means to something else that has value. And some things have a value of their own, an intrinsic value.<sup>4</sup> This notion of intrinsic value is the basic one; other kinds of value exist by their relation to intrinsic value. . . . Let us consider things frequently said to be valuable in themselves. We begin with works of art. Recall what happens in art appreciation classes. You are shown how the different parts and components of a painting are interrelated . . . You are shown how the painting is a unity . . . A painting has aesthetic value, theorists have held, when it manages to integrate a diversity of material into a tight unity often in new and striking ways. Such a “unity in diversity” was termed an *organic unity* . . . The greater the diversity that gets unified, the greater the organic unity, and also the tighter the unity into which the diversity is brought, the greater the organic unity . . . Its organic unity *is* its value.<sup>5</sup>

Nozick’s concept of value is not special to art, even though it begins with an aesthetic example. Indeed, Nozick’s argument is that the same unity-in-diversity is the basis of all intrinsic value, and Nozick promptly extends it to scientific value (and specifically, by implication, coherence: p. 163), to the value of an ecological system (p. 164), and to the value of the mind-body unity of a thinking being (p. 165). However, it might

<sup>3</sup> Like everything else of concern to this chapter, pragmatism has its own problems of definition. For a useful survey of pragmatism, see Haack (2004). Haack stresses the differences among pragmatists, especially between the founders and more recent philosophers who have taken that name. She concludes “ask, rather, what we can borrow from the riches of the classical pragmatist tradition, and what we can salvage from the intellectual shipwreck of radical contemporary neo- and neo-neo-pragmatisms” (p. 34). However, the centrality of *the problem* is expressed in the editorial note to the first issue of the journal *Contemporary Pragmatism*, which precedes Haack’s survey: “A scholarly journal, like any intellectual enterprise, has meaning only insofar as it contributes to efforts to resolve some practical problem” [Shook and Ghirdelli (2004, p. 2)]. This understanding of the meaning of an intellectual enterprise is characteristic of pragmatism.

<sup>4</sup> Utilitarians would criticize the notion of intrinsic value as confused thinking, along the following lines: to say that something is good is to say that it does someone some good, so that intrinsic value – something that is good regardless of whether it does anyone any good or not – is simply confusion. Nozick does not attempt to refute this critique, nor shall we. For a related argument see Baron (1994) and my comments on it [McCain (1994)].

<sup>5</sup> Emphases in the original. Of course, the discussion of intrinsic value has a history of thousands of years, but contains various concepts, many of them theological. In the interest of coherence, this essay relies *specifically* on Nozick’s conceptualization.

not even be a complete account of aesthetic value. It strongly suggests an Apollonian, not a Dionysian beauty. Nietzsche (1871) writes “that in the Greek world there exists a huge contrast, in origins and purposes, between visual (plastic) arts, the Apollonian, and the non-visual art of music, the Dionysian”. For him the highest art arises from the unification of these radical opposites, the Apollonian and Dionysian, unity in diversity even in aesthetic principles. On the other side, value in Nozick’s sense needs not be additive nor partible, even if it is measurable – and Nozick disavows any attempt to measure it. Is something that is possibly unmeasurable, not additive and not partible, really value? Partisans of the labor theory of value will have their doubts!

It should be observed that Nozick does not represent this as the last word on intrinsic value (nor does this essay). On the contrary, Nozick’s objective, as explained in *Philosophical Explanation* (1981) and *Invariances* (2001), “is to open possibilities for consideration, not to close them . . . the reasons . . . to support them are meant to launch them for exploration, not to demonstrate that they are correct” [Nozick (2001, p. 3)]. The discussion of value in this essay is presented in the same exploratory spirit. On this basis, what can we say about *economic value*? A human being is a remarkable unity of a diversity of parts, tiny material parts and spiritual and mental parts, a thing of great intrinsic value. And since the capacity to choose between alternatives is one important function of such a creature, her choices derive a certain instrumental value as realizations of herself. This, then, is the basis of economic value, measurable by willingness to pay (perhaps with adjustment for ability to pay), and as such, additive and partible. The apparatus of economic theory, including preference systems and Pareto-optimality, is well adapted to considering the choices of individual human beings without giving distinction more to one than another; that is, to the analysis of economic value as it is constructed here. But this would be only one sort of value, indicating the truth in the idea that neoclassical economics is focused on economic values.

### 2.3. *Cultural value*

Human groups may also have their distinct value, that is, unity-in-diversity. Nozick tentatively (p. 73) allows a loving couple as a novel unity with its own intrinsic value. Larger groups with distinctive cultures, such as tribes and nations, probably have less unity but much greater diversity, and so some degree of intrinsic value. Nozick (pp. 166–168) relates meaning to value: “Value involves something’s being integrated within its own boundaries, while meaning involves its having some connection beyond those boundaries . . . Meaning can be gained by linking with something of value”. Thus symbols and practices that link the cultural group have meaning to that group, and by enhancing its unity, enhance the intrinsic value of the group.

Klamer (1997) relates how windmills came to be identified with Dutch national culture, and how this change in the anthropological or popular culture of Dutch people was associated with the rise of Dutch nationalism. He writes (1997, p. 82): “Whether a windmill is exemplary of Dutch cultural heritage depends on the common values that people share, on their culture” [Klamer (1997, p. 75)]. Accordingly, it seems appropriate

to consider windmills as “cultural” goods, in some sense we have yet to define. Many of the same things can be said about covered bridges in North America.<sup>6</sup>

Some cultural economists and others have argued that no attempt should be made to justify cultural policy by an appeal to “economic value” alone [Throsby (2001, pp. 31–34 et seq.; 2003, p. 282)]. Throsby draws on concepts from economic development and argues that investment in “cultural capital” may be necessary for the sustainability of a culture [Throsby (2001, Chapter 3)]. But is this feasible? Ideas like investment and the rate of return to cultural capital presuppose partibility and additivity, which, as we have seen, cannot be assumed for intrinsic value.<sup>7</sup> Klamer (2003) adopts a pragmatic view, drawing on the ideas of Dewey<sup>8</sup> to indicate that there may be distinct realms of value. But pragmatism *per se* takes the problems as being given, and so provides us with no guidance as how to distinguish these different realms. By offering some such guidance, and raising important questions about the applicability of arithmetic methods to cultural value, Nozick’s intrinsic value conception may provide an important complement to the pragmatist’s problem-orientation, and thus a basis for formal definition.

On this view, then, objects of art may have intrinsic value (as they successfully realize a novel diversity-in-unity), cultural value (as symbols of some cultural unity), and economic value (in that some individual is willing to pay for them). Objects that carry cultural heritage may have cultural value and meaning as such. Cultural value may be correlated with economic value only if some people prefer objects with cultural value to objects that lack it. But having granted as much, what progress have we made toward a

<sup>6</sup> Popular attitudes toward artifacts such as windmills and covered bridges can often be documented by picture postcards; see McCain (2002b).

<sup>7</sup> Throsby concedes that “a single measure is very difficult to specify, given the multi-dimensional nature of cultural value” (2001, p. 84), but regards such a measure as theoretically possible and suggests that the measure might be used in cost–benefit analysis (p. 83). The implication suggested here is that such a measure, and such an exercise in cultural cost–benefit analysis, is impossible even in principle. Non-partibility means that an object of cultural value cannot be treated as an aggregate of parts, so that marginal concepts are inapplicable. This goes beyond indivisibility. Rose Tree Tavern is a historically important building near my home, which has recently been moved a few hundred feet to allow a road-widening project. But the historic significance of the building is not simply a sum of the structure and its site, so we may say that only the site has been sacrificed. Rather, arguably, the building at its original site is simply a different cultural object, with an *incommensurably* different cultural value, than the building at its new location. Similarly, that intrinsic values are non-additive means both that the value of the Rose Tree Tavern cannot be considered as a sum of the value of the structure and that of the site, and that future cultural benefits from its preservation cannot be accumulated and discounted to present value to be considered as an offset against the cultural loss of moving it from its original site (rather than demolishing it). On the other hand, recognition of the non-partibility and non-additivity of cultural and some other intrinsic values gives us an answer to a neoclassical criticism. The neoclassical criticism is that non-economic values must be reducible to economic values, since people do in fact make choices that trade off one kind of value against another, and these tradeoffs give us the terms of equivalence for the reduction. But this criticism too presumes that both kinds of values are partible and additive and fails when we realize that cultural values do not have these arithmetic properties.

<sup>8</sup> While pragmatism is also the philosophical standpoint of American Institutionalism, Klamer does not cite the works of economists and philosophers such as Veblen, Ayers and Dugger who write in that tradition.

definition of cultural goods? If the definition is to be helpful, it should allow us to make some connection to recognized aspects of cultural goods. One widely recognized aspect of cultural goods is the economic value of provenance.

#### 2.4. Provenance

Consider the following conundrum. We suppose that two pictures are put on sale at the same time: an authentic Rembrandt and a perfect duplicate of it. We suppose that the duplicate has been created by means of the newly invented “Molecular Duplicating Device” (MDD) which places each molecule of the duplicate in the exact relative position of a corresponding molecule of the original, so that the duplicate cannot be distinguished from the original by any ordinary means. Common sense tells us that the market value of the authentic picture will be much greater than that of the duplicate. Indeed, something of the kind can occur in real life when the attribution of a work of art to a famous artist is rediscovered. A Raphael, worth £8000 when it was thought to be a copy, commanded an offer of £35,000,000 after its attribution to Raphael was rediscovered and clarified [Spanier (2003); Jeromack (2002)]. A painting that had not seemed worth selling when attributed to Jan van den Hoecke was estimated at about £6,000,000, but sold for £49.5 million when reattributed to Rubens [Moncrieff (2002a, 2002b)]. Similarly, a Paul Revere silver porringer might sell for ten times the price of a contemporary porringer of equally fine workmanship, because of “the patriot factor”. What does the authentic picture or porringer have that the duplicate lacks? It is easy to give a name to it: provenance.

People are willing to pay for items with provenance. In antiques markets, for example, a weapon known to have been used in war will generally get a higher price than an otherwise identical weapon that was never used in violence. This association of the weapon with war is an objective historical fact, as is the association of the authentic Rembrandt with the historic van Rijn. For a broad category of goods, including (but not limited to) goods that we think of as cultural or artistic goods, *as a matter of fact* the willingness of people to pay for them is dependent on their history. The Revere porringer’s provenance makes it valuable partly because it is symbolic of the American struggle for independence; covered bridges are considered worthy of preservation because they are symbolic of the idealized rural North America of the 19th century; and windmills symbolic of a Netherlands in some similar idealized time. The porringer may also have intrinsic value as a fine item of craft, but, as we have seen, the value of provenance goes beyond that.

Tentatively, then, we define “cultural goods” as goods that carry cultural value, in that they derive their meaning from the unity-in-diversity of some specific cultural group and from the fact that they symbolize, through their provenance, the unity and distinctness of that group.



## 2.5. *Artistic value*

We have speculated with Nozick that artistic value is yet another distinct realm of intrinsic value. However, an art historian<sup>9</sup> writes, “I don’t think ‘unity in diversity’ is the essence of value in art. For one thing, there are many things, like airplanes and computers, which are marvelous examples of unity in diversity, but they are not art, and even if they have aesthetic value, that is not the same as the value that is in art”. This is a useful but not a decisive counterargument. It has not been asserted that value as unity in diversity is special to art, but the contrary, that it is the basis of intrinsic value in non-art as well as artistic cases. That, however, leads to the question: what particular unity-in-diversity may we associate with artistic value?

In fine craft, of course, unity of “form and function” may itself be a value. Similarly, the unity-in-diversity of an airplane has much to do with its function as a flying machine, and is largely obvious on the basis of that function, while the unity-in-diversity of the Greek Tragedy (on Nietzsche’s account) has little to do utilitarian function,<sup>10</sup> and is far from obvious on any basis. Mossetto (1993, p. 19) goes still further “The absence of ‘interest’ or ‘aim’ is a permanent feature of aesthetic judgments . . .” As an interim conclusion we may say that while Nozick’s conceptualization of value allows us coherently to assert a distinction between cultural and economic (and other) values, it does not tell us under what circumstances unity-in-diversity defines artistic rather than other sorts of values.

What can we now judge about the relation of artistic and cultural goods? In discussions consistent with UNESCO’S official position, Throsby (2001, p. 4) conceptualizes artistic goods as a subset of cultural goods and defines cultural goods by three characteristics: that creativity is (especially) involved in their production, that “they are concerned with the generation and communication of symbolic meaning”, and that they are potentially intellectual property. However, in the perspective of this essay, this seems to go too far. Goods that carry artistic value may also carry cultural value, so that the categories can have an important overlap, but it is clearly logically possible that goods may carry one sort of value and not the other. Not all cultural goods are artistic: Klamer’s windmills (and McCain’s covered bridges) may have involved no more creativity in their production than does any engineering work, have acquired but did not generate and are not “concerned with . . . communication” of symbolic values, and are perfectly ordinary property but not potentially intellectual property. Even some art works such as sculptures can hardly be identified as potential intellectual property, although they can be quite ordinary property. The Revere porringer may have been a highly creative product at its origin, since Revere was a fine silversmith and a creative man, but its cultural value has nothing to do with that, and it is ordinary, not intellectual property. A sword

<sup>9</sup> Charles Morscheck, Drexel University, personal communication, 2003.

<sup>10</sup> On the other hand, the social distinction of art from craft seems to encourage “artists” to avoid any interest in either form or function. The late romantic idea that this devalues “pure” art relative to fine craft expresses a particular aesthetic position consistent with, but not entailed by, Nozick’s conception of intrinsic value.

that happens to have been used in combat in the American Civil War almost certainly does not satisfy the first or the third of Throsby's list and bears the same relation to symbolic value as do the windmills and covered bridges. To an American it may be a very important cultural good. On the other side, artistic goods are not of cultural value in all cultures. Not all cultures value creativity as modern European-derived cultures do. In some such cultures, High Culture may be identified with athletic prowess, or with the faithful repetition of the tradition, rather than with creativity. In the face of these examples, it seems that the relationship of cultural and artistic goods is better represented by a ballentine than a Venn diagram. Nevertheless, Throsby's triad provides us with a key point to refine our understanding of artistic value both in itself and in its overlap with cultural value. For this we need a clear concept of creativity. That is the next task of the essay.

### 3. Creativity

#### 3.1. *Simon on creativity*

Creativity is two-sided. In *Simon's* (2001) words "We judge thought to be creative when it produces something that is both novel and interesting or valuable" (p. 208). In the words of *Amabile and Tighe* (1993, p. 9), "... [M]ost researchers accept a conceptual definition that includes two elements: novelty and appropriateness". For *Radford* (2004, p. 56), "In order to be creative it may be argued that it is insufficient to draw attention to a dissonance or controversy, but rather the artist must offer a way of addressing it. Jerome Bruner has said that the creative act should bring about a 'shock of recognition', a sense that the act tells us something that we already half knew". *Bruner's* (1979) "shock of recognition" points up the tension between these two aspects of creativity, even at the definitional level.

The novelty and seeming freedom of creativity suggests that it is spontaneous, independent of preparation. But that seems to be something of a romantic fallacy. As *Cohen* (2002) writes: "If we survey the work of any major artist, we get the distinct impression of someone who knew exactly what he or she was doing and knew exactly where he or she was going. Mozart always sounds like Mozart. Matisse did not produce a Picasso on Monday and a Miró on Thursday, he produced Matisses every day". Moreover, "Mozart began to compose at age four or five; until he was at least seventeen he composed no music that, however remarkable for a youngster, could be regarded as world class" [*Simon* (2001)]. Furthermore, "Most creators do not function in isolation from other creators, but rather their creativity takes place within a particular artistic, scientific, or intellectual discipline" [*Simonton* (2001)].

These facts are stressed by Simon in support of his interpretation of creativity. For Simon, creativity is nothing more than expert problem solving. "In fact, we can describe in considerable detail the fundamental processes that creative activities – in common with other human mental activities – employ ..." (p. 205). "The memory of the expert,

then, is like an indexed encyclopedia, where the index entries are recognizable patterns, each of which points to a body of organized relevant knowledge . . . The empirical evidence suggests that a world-class expert has a quarter million or half million patterns, with associated information, about the expert domain” (p. 207). On a widely-held alternative view, [Ansburg and Hill \(2003, p. 1143\)](#) argue that “Creative thinking produces innovative solutions, whereas analytic thinking evaluates and tests existing ideas”.

### 3.2. Generative linguistics

There is another formal model that has been perhaps even more influential in the writing on creativity. This is generative linguistics.<sup>11</sup> The formal modeling of this sort of creativity begins with the work of [Chomsky \(1972\)](#). In Chomskian linguistics, “. . . a speaker’s linguistic ability must be characterized in terms of a generative grammar, a system of combinatorial rules or principles that ‘generate’ or ‘derive’ an infinite set of possible sentences from a finite vocabulary” [[Jackendoff \(1999, p. 7\)](#)]. Thus, in generative linguistics, we have an approach comprised of formal rules that are creative in that they can generate an unlimited array of novel, grammatically valid expressions.

In accepting this view of linguistic creativity, we need not accept the idea that linguistic expression is always creative. [MacKenzie \(2000, p. 173\)](#) is critical of Chomsky’s minimalism: “Not unlike the oral-traditional poets of Homer’s day, we routinely rely on a vast store of fixed, prepatterned phrases, which we use more often than we generate locutions entirely from scratch”. Nevertheless, he goes on (p. 174): “Obviously, no one would deny that there *is* [emphasis in the original] a generative element to language acquisition, as it is this that enables us to be wholly novel and innovative, or more frequently, to produce slightly novel and unexpected variations on familiar usage”. On this synthetic view, expert behavior *à la* Simon facilitates, but is not sufficient for, linguistic creativity, and the same point may be applied to creativity in general.

In recent years, formal generative linguistic models have been used in studies of creativity in musical composition [[Pearce and Wiggins \(2002\)](#)], visual composition [[Stebbing \(2004\)](#)], and creative computer programs [[Cohen \(2002\)](#)]. However, the more established literature on creativity uses ideas from generative linguistics informally, with “generative rules” as a key concept. Here, the work of Margaret Boden is widely recognized as central. She writes (1991, p. 40): “A merely novel idea is one which can be described and/or produced by the same set of generative rules as are other, familiar ideas. A genuinely original, or creative, idea is one which cannot”. Thus, creative expressions are expressions that change the rules (in more Chomskian language, that change the surface structure)<sup>12</sup> of the system. This key insight leads on to important

<sup>11</sup> Yet another no less scientific approach is found in the broadly evolutionary view of Campbell and Simonton; see [Simonton \(2001\)](#). While this view is valuable especially for its insights about the way that novelty arises, it is less helpful in understanding novelty *per se* and thus is not used further in this discussion.

<sup>12</sup> In the words of [Jackendoff \(1999, p. 8\)](#) “The fact that the same semantic relation can be expressed by different syntactic means (e.g., active and passive sentences) is a consequence of moving constituents in the

conclusions for economics, especially for the study of the economics of the arts and innovation.

While "... few theorists hold the strong position that a creative idea must be completely *unique*",<sup>13</sup> we may "... contrast between 'little C' creativity – the sort which all of us evince in our daily lives – and 'big C' creativity – the kind of breakthrough that occurs only occasionally" [Gardner (1993, p. 29)]. In the case of "big C" creativity, novelty or innovation is particularly apparent. At the same time, the creative product must be appropriate or adequate according to some external criterion. In music, for example, there is "the need to compose music whose structure may be perceived by the listener: there must be some kind of alignment between the compositional mechanisms of the composer and the perceptual mechanisms of the listener" [Pearce and Wiggins (2002, p. 4)]. "To a large degree such criteria are culturally determined; we do agree on most issues, and it is rare for the work of a single individual to enforce a major shift of criteria for the entire culture. At the same time, creative behavior must involve an increasing differentiation, in whatever degree, of the individual's criteria from those of the culture at large" [Cohen (2002, p. 61)].

All the same, novelty and appropriateness are alike a matter of subjective judgment, and the judgment is specific to a domain. "... creativity occurs in domains. That is, a creative individual is never creative across the board. Rather, creative individuals display their creativity in disciplines or crafts; usually one domain, though occasionally, as in the case of Leonardo da Vinci, in two or even three domains"<sup>14</sup> [Gardner (1993, p. 32)]. Thus, domain-relevant skill is first among the skill components required for creativity [Amabile and Tighe (1993, pp. 14–15)]. Gardner goes further: "... a creative individual is one who regularly solves problems, fashions products, and/or poses new questions in a domain in a way which is initially considered novel but which is ultimately accepted in at least one cultural setting" (p. 32).

According to Boden, creative ideas are surprising. Gardner's distinction of "big C" and "little c" creativity corresponds roughly to Boden's "P-creative" (psychologically creative) and "H-creative" (historically creative) ideas. P-creative ideas "are fundamentally novel with respect to *the individual mind* which had the idea. If Mary Smith has an

course of a syntactic derivation, creating a disparity between "Deep" and "Surface" structure ... " This is part of Chomsky's original view, and Jackendoff questions it, but some such hierarchy of rules is essential for creativity studies, as we shall see.

<sup>13</sup> Amabile and Tighe (1993, p. 7). They add that for their theory, a third characteristic is required: "the task must have been heuristic for the individual, rather than algorithmic. That is, the task as presented must have been somehow open-ended, with no clear and straightforward path to a single solution". But this seems to be another aspect of innovation.

<sup>14</sup> Indeed one of the *prima facie* exceptions, Herbert Simon, himself disavowed that he worked in multiple realms. Citing a personal communication, Dasgupta (2003, p. 686) quotes Simon as follows: "... the 'Renaissance Mind' is not broader than other intelligent minds but happens to cover a narrow swathe across the multi-dimensional space of knowledge that happens to cut across many disciplines which have divided up the space in other ways. My own narrow swathe happens to be the process of human problem solving and decision making, and almost everything I have done lies in that quite narrow band".

idea she could not have had before, her idea is P-creative – no matter how many people have had the same idea already. The historical sense applies to ideas that are fundamentally novel with respect to *the whole of human history*. Mary Smith’s surprising idea is H-creative only if nobody ever had it before” [Boden (1991, p. 32)]. Both categories of ideas are surprising in a deep sense. A parallel, but distinct, dichotomy distinguishes exploratory from transformational creativity. “The process of generating ‘merely novel’ artifacts is described as exploratory creativity, while the generation of ‘genuinely original’ artifacts is described as transformational creativity” [Pearce and Wiggins (2002)]. Once again, the latter category is surprising. “Where creativity is concerned, we have to do with expectations not about probabilities, but about possibilities. Our surprise at a creative idea recognizes that the world has turned out differently, not just from the way we thought it *would*, but even from the way it *could*” [Boden (1991, pp. 30–31; emphasis in the original)].

However, there seems to be an inconsistency here. In generative linguistics, the novelities that occur are not themselves generative rules. Rather, the generative rules generate novel utterances. Conversely, the generative rules themselves are universal, and so unchanging – at least on any timeframe less than that of human evolution. Thus, Boden’s formulation seems to contain an ambiguity: on the one hand, the generative rules make creativity possible, but on the other hand, the creative product cannot be a result of the previously existing generative rules. We might resolve it this way: the ambiguity is between the production system (if–then rules as envisioned by Simon and MacKenzie) and the underlying generative rules. Objects of art (and technology and science) are indeed outputs of a production or expert system of if–then rules. But this system of if–then rules is not itself given for all time. Rather, it is a product of a system of generative rules which is capable of generating an unlimited number of potential if–then rules. When Boden (1991, p. 40) says “A genuinely original, or creative, idea . . . cannot . . . be described and/or produced by the same set of generative rules as are other, familiar ideas”, she might instead have said that a genuinely original, or creative, idea cannot be produced by the same set of if–then production system rules as are other, familiar ideas. Instead, it is the product of a new set of if–then rules, which are themselves novel products of the unchanged underlying generative system. In any case, that will be the position of this essay. Indeed, it is an extension to creative activity in general of MacKenzie’s understanding of creativity in language.

The situation is probably still a little more complex. In addition to the fundamental and unchanging system that generates novel if–then rules, and the if–then rules an expert uses directly to create a new work of art or econometric model, there will be a middle stratum of rules that are products and are also open as generative rules are, capable of producing an unlimited range of if–then rules to handle specific circumstances. Most likely the scientific method, theories of aesthetics, and engineering disciplines belong to this middle stratum. When Boden speaks of generative rules, we should (I suggest) interpret her comments as applicable not to the fundamental generative rules that do not change, but to this middle stratum. In what follows I will use the broader term

“subjective knowledge structures”<sup>15</sup> to encompass all three. In any case, we will need both (unchanging) generative rules and (changing and expanding) production systems to understand creativity.

Boden’s formulation can serve to pull some of the threads of this discussion together. The domain-relevant skill (Amabile and Tighe) or cultural setting (Gardner) required to judge creativity supply the knowledge of pre-existing subjective knowledge structures. The subjective knowledge structures themselves define the discipline or craft (Gardner) in which creativity takes place. The necessary knowledge of the craft and its subjective knowledge structures define the domain-relevant skills required of the creative person herself (Amabile and Tighe). In some fields, such as the simpler technologies, the appropriateness of the creative product may be obvious – measurable, perhaps, by a change in labor productivity [McCain (1981b)] – but in other cases judgments of appropriateness are themselves products of the subjective knowledge structures shared by the community skilled in the pre-existing domain. Indeed we may say that judgments of appropriateness are themselves products of the rules.<sup>16</sup> Note that the creative production is *impossible* on the basis of the pre-existing subjective knowledge structures, so it can only be that the creative act has given rise to new rules of production or of judging appropriateness, or, perhaps in most cases, both. Thus, even in fields such as art, knowledge is accumulative, and for this reason the H-creative act will influence subsequent work.

Indeed it is H-creative *because* it influences further work. As McCain (1981b) observes, tradition and innovation are interdependent: we recognize the innovation in that it contributes to the tradition. It is this historic association that gives the great work of art its special provenance. As McCain also points out, there are many failed innovations for every successful one, and there is no possibility of predicting which innovations will succeed, since the subjective knowledge structures necessary to judge them do not exist until after the successful innovation has been produced and recognized as such. This leads naturally to the “nobody knows property” and the crucial role of intermediaries and buffs with particular skill in the domain in markets for art [Caves (2000, p. 3 et seq., p. 185)]. The knowledge of the intermediaries and buffs cannot eliminate, but does reduce, uncertainty about the historic quality of an innovation.

### 3.3. Creativity in consumption

Turning now to consumers, we note that new works especially will not be successful without interpretation. In Moseotto’s words, “Interpretation is characterized by the same features of creativity . . .” (1993, p. 73). For works of art, then, it is not only the artist whose creativity is engaged. Being a spectator or consumer (interpreter) of art calls on

<sup>15</sup> This precise phrase is due to Boulding (1956), but the term “knowledge structures” is widely used in cognitive science; see, e.g., Mandler (1985), McCain (1992).

<sup>16</sup> Formally, a rule for judging appropriateness takes a product as its argument and generates a judgment as its dependent variable. This is the form of a linguistic generative rule.

the skills and mental processes characteristic of the artist herself, though, to be sure, in lesser ways. The creativity of the art consumer is P-creativity, not H-creativity, in Boden's terminology. Nevertheless, the creation and consumption of art links the artist and the consumer in a unity of interrelated creative action. This sort of linkage through a shared experience does not depend on just what this joint creativity of artist and interpreter creates. The "what is created" may be an emotional mood, an arousal of the senses [Shanahan et al. (1978, p. 16)], an image, a story, or all of those and more. It is the linkage through shared creativity that counts.

This provides an answer to Bentham's (1843) famous view that the game of push-pin is of equal value with the arts and sciences of music and poetry. Art is about ideas, doctrines, poetry, rhetoric and wisdom. It is about expression and affecting the emotions of the viewer, reader or hearer.<sup>17</sup> But many things that are not art are also about those things. What makes the emotional mood or the wisdom art rather than something else is that its full appreciation engages the consumer's creativity in a joint act of creation with the artist. This joint creativity defines a new unity-in-diversity, a new intrinsic value, which I submit is the distinctive intrinsic value of art.<sup>18</sup>

Thus, it seems that creativity is (*pace* Caves) the key aspect of artistic goods and services. Further, cognitive-scientific study of creativity can both rationalize many aspects of art markets and can explain the connection of important art products to provenance, that is, the tendency of buyers to assign value to the item on the basis of its historical associations. Tentatively, then, we define "artistic goods" as goods that carry non-economic value, in that they create a novel unity-in-diversity by calling forth a common creativity of the artist and the consumer of art.

#### 4. Implications for consumers of art

Once again, if the proposed definitions are to meet Malthus' criterion, they will underscore the connections we expect to find with other phenomena in markets for art. In this section we consider learning-by-consuming and artistic goods as stimulus goods and (in a negative sense) intellectual property.

##### 4.1. *Learning-by-consuming*

The domain-specific knowledge and skills necessary for creative consumption of art are together known as "taste", and are something not given but acquired. One cannot be a productive consumer of creative work without at least some knowledge of the generative

<sup>17</sup> Charles Morscheck, personal communication, cited above (Footnote 9).

<sup>18</sup> Bryant and Throsby (2006) (Chapter 16 of this volume) focus mainly on the conflict of interest between the artist, who is assumed to value creativity, and the marketer. The conflict of interest would follow from the "nobody knows" property as more experimental work is likely to fail and, even if successful, to appeal in the marketplace to a narrower expert market. Thus their view would be consistent with the view sketched here.



rules that produced it and determine its evaluation, and that knowledge is gained in part by the experience of consuming the artistic product. Caves (2000, pp. 175–178) observes that creative industries rely on tastes that are to some extent cultivated, and cultivation of taste has been modeled in economics as “rational addiction”. As he notes, the basic tools for such a model come from Stigler and Becker (1977), but as McCain has argued, the Stigler–Becker model does not account for the observed multimodal distribution of cultivation of taste without non-linearity and some other element, such as bounded rationality, hysteresis, or some game-theoretic coordination failure.<sup>19</sup>

This idea lends itself to modeling in nearly conventional terms, if the cognitive details are kept ruthlessly in the background. Let  $\{x, y\}$  be the quantities respectively of an artistic and an ordinary good consumed in some particular market. Then let  $z = kx$  be the “sensation” derived from consuming  $x$  units of the artistic good, and  $U = f(kx, y)$  the utility function of a representative consumer. Then we let  $\dot{k} = g(x, k)$ , with  $g$  an increasing function over at least part of its range and a decreasing function of  $k$  over at least some part of its range. Positive values of  $g(x, k)$  correspond to ongoing cultivation of taste.

This schema lends itself to the formulation of a dynamic optimization problem. The variables  $x$ ,  $y$  and  $k$  will then be functions of time and  $x$ ,  $y$  are subject to an income constraint. Let  $m$  be income, a function of time, and  $p_x$  and  $p_y$  the prices of  $x$  and  $y$ , respectively, which may be constants or functions of time. Then the problem is

$$\text{Max} \int_0^T f(kx, y) dt \quad \text{subject to } \dot{k} = g(x, k)$$

and

$$\int_0^T (p_x x + p_y y) dt \leq \int_0^T m dt.$$

The solution to this problem would give a path of “rational addiction” to the arts. Over this path, the rational agent trades off a rate of expenditure on the arts now (which might not maximize instantaneous utility) against increased pleasure from the arts, at the same cost, in the future [McCain (1981a)]. There could be multiple equilibria and initialization effects. For some agents who begin with relatively large “arts capital”  $k$ , perhaps because of exposure to the arts in childhood, it might be optimal to become highly cultivated, whereas for others who begin with lesser  $k$ , the optimal path would be one of rational philistinism and minimal cultivation. This multiplicity of equilibria, depending on initial conditions, could explain a bimodal distribution of degrees of cultivation of taste, as agents tend (depending on their initial conditions) to cluster in the neighborhoods of two or more distinct “basins of attraction” (to use the jargon of mathematical dynamics). The welfare economic implications of such a schema do not seem to have been investigated.

<sup>19</sup> References for McCain are: (1) non-linearity [McCain (1979)]; (2) hysteresis [McCain (1981a)]; (3) bounded rationality [McCain (1995a, 1995b)]; (4) game theory [McCain (1986)].



Another possibility, however, is that people are shortsighted, in that at each moment of time they

$$\text{Max} \int_0^T f(kx, y) dt \quad \text{subject to} \quad \int_0^T p_x x + p_y y dt \leq \int_0^T m dt$$

ignoring the future benefits of cultivated tastes. In such a case cultivation of taste enters the model as an externality as would be the case in a more conventional welfare economic model [McCain (1979, 1982, 2002a); Stevens (1985)]. Moreover, this may not be a failure of rationality, since there is a coordination problem among the agents. Unless a sufficiently large proportion of them cultivate tastes for the arts, there will not be enough demand for the arts to pay the fixed costs of an artistic infrastructure, and in that case, those who do cultivate a taste for the arts may be worse off rather than better off [McCain (1986)]. Moreover, real people are boundedly rational learners, and computer simulations suggest that in a population of agents who learn by experience with bounded rationality, only a portion of the population becomes cultivated even though the cultivated are unambiguously better off than the uncultivated. Furthermore, the proportion and other details are highly sensitive to the initial conditions and random disturbances of the simulation [McCain (1995a, 1995b)].

By keeping the value issues, creativity issues, and cognitive issues off the table, this approach enables us to model the demand for arts in ways that are consistent both with the importance of learned, domain-specific consumer knowledge (indicated by the creativity approach) and with the economics of social behavior. Nevertheless these reflections do not contribute toward a definition of cultural or artistic goods, because the role of cultivated taste is not limited to the economics of the arts. Indeed, McCain's empirical demonstration of it was a study of the demand for wine [McCain (1979)] and he observes that Scotch whiskey is no less subject to cultivation of taste than wine or grand opera. The same comments could be applied to sports and a wide range of hobbies. If we were to define artistic or cultural goods simply as goods whose productive consumption requires cultivation of taste, then we would be admitting whiskey and soccer as forms of high culture! That will not do. However, our discussion of cultivation of taste does at least extend our understanding of the role of creativity in markets.

#### 4.2. Artistic goods as stimulus goods

Scitovsky (1972, 1976) has proposed an alternative economic psychology in which there is an innate human need for stimulus. He suggests that individual decisions may be biased away from stimulus goods, and artistic goods are seen as stimulus goods that may be underconsumed. Scitovsky reasons that artistic goods are pattern-complex, so that they are able to hold our attention on repeated exposure; but this same pattern-complexity requires effort (and perhaps cultivation) that may deter individuals from making the commitment necessary to meet their (hypothetical) need for stimulus. These ideas are consistent with the literature on creativity, where domain-specific knowledge of complicated generative rules (and, indeed, the consumer's own creativity) is seen

as necessary for full *appreciation* of the creative product. We can accept the idea that artistic products are stimulus goods whether or not we endorse Scitovsky's hypotheses of an innate need for stimulus. It is also possible that Scitovsky's approach can support a model of market failure through asymmetric information<sup>20</sup> [McCain (1980)]. Like cultivation of taste, however, stimulus and engaging complexity are not unique to artistic goods. Here again, sports and such games as bridge and chess share the fascination Scitovsky attributes to art, and for many people politics, religion, and fighting are the stimulus goods of choice. As with cultivation of taste, the discussion of stimulus goods contributes to the definition of artistic goods in so far as it extends our understanding of the implications of creativity.

#### 4.3. *Intellectual property*

Many artistic goods may be copyrighted. However, many works may be copyrighted that we would not ordinarily think of as artistic. Consider, for example, a textbook of economics. Such a work is eligible for copyright if it is original in expression. I can say from experience that some P-creativity is necessary to write a text; but if the text were H-creative it would fail as a textbook. Indeed, the standard for eligibility for copyright seems explicitly to allow work that is not H-creative, by contrast with the requirement for patentability, which specifies that the invention to be patented must be an advance on products already in existence. Artistic creativity then is at most a sufficient but not a necessary condition for the work to be eligible for copyright. Like cultivation of taste and stimulus, the importance of copyright in the artistic world follows from the key role of creativity, but is not by itself definitional of artistic goods.

### 5. Conclusion

The ultimate justification for distinguishing artistic and cultural goods from other goods is pragmatic. Problems of the artistic industries or of cultural sustainability may not be adequately understood, and therefore not solved nor ameliorated, if artistic and cultural goods cannot be adequately defined. But appeals to pragmatism are not conclusive. Problem recognition is itself an aspect of creativity, dependent on domain-specific knowledge, but reflective also of preferences and prejudices. Put otherwise, one person's problems may not be problems from the point of view of another. This chapter has suggested a broader conception of value, that allows distinction among economic, cultural, artistic and aesthetic values, and has proposed definitions of cultural and artistic goods and services in terms of those categories of values. These concepts of value allow

<sup>20</sup> But McCain's (1980) model fails empirically in that it predicts a chronic shortage of artists, whereas the evidence, as reported by Alper and Wassall (2006) in Chapter 23 of this volume, indicates that artists are disproportionately affected by structural unemployment.

us to recognize cultural value in artifacts that are symbolic of the diversity-in-unity of particular cultural groups, and to recognize artistic value in the experience of shared creativity on the part of the consumer and producer of art. This recognition leads us to incorporate learning on creativity from cognitive science in cultural economics. This chapter has reviewed the learning on creativity and given examples of its use as a tool for the economic understanding of those industries. The hope that motivates this exercise is the hope that reason may yet arbitrate our differences of opinion about culture, art and public policy.

## References

- Alper, N., Wassall, G. (2006). "Artists' careers and their labor markets". In: Ginsburgh, V.A., Throsby, D. (Eds.), *Handbook of the Economics of Art and Culture*, vol. 1. Elsevier. **Chapter 23** in this volume.
- Amabile, T., Tighe, E. (1993). "Questions of creativity". In: Brockman, J. (Ed.), *Creativity*. Simon and Schuster Touchstone Books, New York.
- Ansborg, P.I., Hill, K. (2003). "Creative and analytic thinkers differ in their use of attentional resources". *Personality and Individual Differences* 34, 1141–1153.
- Baron, J. (1994). "Nonconsequentialist decisions (with commentary and reply)". *Behavioral and Brain Sciences* 17, 1–42.
- Bentham, J. (1843). "The rationale of reward". *Jeremy's Labyrinth*, University of Texas. Available at <http://www.la.utexas.edu/labyrinth/rr/rr.b03.c01.html>, as of October 7, 2002.
- Boden, M.A. (1991). *The Creative Mind, Myths and Mechanisms*. Basic Books, New York.
- Boulding, K.E. (1956). *The Image*. University of Michigan Press, Ann Arbor.
- Bridgeman, P.W. (1952). *The Nature of Some of Our Physical Concepts*. Philosophical Library, New York.
- Bruner, J. (1979). *On Knowing: Essays for the Left Hand*. Harvard University Press, Cambridge, MA.
- Bryant, T., Throsby, D. (2006). "Creativity and the behaviour of artists". In: Ginsburgh, V.A., Throsby, D. (Eds.), *Handbook of the Economics of Art and Culture*, vol. 1. Elsevier. **Chapter 16** in this volume.
- Caves, R.E. (2000). *The Creative Industries*. Harvard University Press, Cambridge, MA.
- Chomsky, N. (1972). *Language and Mind*. Harcourt Brace Jovanovich, New York.
- Cohen, H. (2002). "A self-defining game for one player: On the nature of creativity and the possibility of creative computer programs". *Leonardo* 35 (1), 59–64.
- Dasgupta, S. (2003). "Multidisciplinary creativity: The case of Herbert A. Simon". *Cognitive Science* 27, 683–707.
- Gardner, H. (1993). "Seven creators of the modern era". In: Brockman, J. (Ed.), *Creativity*. Simon and Schuster Touchstone Books, New York.
- Haack, S. (2004). "Pragmatism, old and new". *Contemporary Pragmatism* 1 (1), 3–41.
- Jackendoff, R. (1999). "Reexamining the foundations of generative linguistics". *Proceedings der 4. Fachtagung der Gesellschaft für Kognitionswissenschaft*. University of Bielefeld.
- Jeromack, P. (2002). "Getty buys Northumberland Raphael – Met nets Lorenzetti". *The Art Newspaper*. Available at <http://www.theartnewspaper.com/news/article.asp?idart=10234>, as of December 27, originally posted on 10/31/2002.
- Klamer, A. (1997). "The value of cultural heritage". In: Hutter, M., Rizzo, I. (Eds.), *Economic Perspectives on Cultural Heritage*. St. Martin's, New York, pp. 74–87.
- Klamer, A. (2003). "A pragmatic view on values in economics". *Journal of Economic Methodology* 10 (2), 191–212.
- Kroeber, A.L., Kluckhohn, C. (1963). *Culture: A Critical Review of Concepts and Definitions*. Vintage, New York.
- MacKenzie, I. (2000). "Improvisation, creativity and formulaic language". *Journal of Aesthetics and Art Criticism* 58 (2), 173–179.

- Malthus, T.R. (1827). *Definitions in Political Economy*. John Murray, London.
- Mandler, G. (1985). *Cognitive Psychology: An Essay in Cognitive Science*. Lawrence Erlbaum Associates Publishers, Hillsdale, NJ.
- McCain, R.A. (1979). "Reflections on the cultivation of taste". *Journal of Cultural Economics* 3 (1), 30–52.
- McCain, R.A. (1980). "Markets for works of art and markets for lemons". In: Hendon, W. (Ed.), *Economic Policy for the Arts*. Abt, Boston.
- McCain, R.A. (1981a). "Cultivation of taste, catastrophe theory, and the demand for works of art". *American Economic Review* 71 (2), 332–334.
- McCain, R.A. (1981b). "Tradition and innovation: Some economics of the creative arts, science, scholarship, and technical development". In: Galatin, M., Leiter, R. (Eds.), *Economics of Information*. Nijhoff, Boston, pp. 173–204.
- McCain, R.A. (1982). "Optimal subsidies to the arts in a shortsighted world". *Journal of Cultural Economics* 6 (1), 15–32.
- McCain, R.A. (1986). "Game theory and the cultivation of taste". *Journal of Cultural Economics* 10 (1), 1–15.
- McCain, R.A. (1992). *A Framework for Cognitive Economics*. Praeger Publishers, New York.
- McCain, R.A. (1994). "Consequentialism in haste". *Behavioral and Brain Science* 17 (1), 23–24.
- McCain, R.A. (1995a). "Cultivation of taste and bounded rationality: Some computer simulations". *Journal of Cultural Economics* 19 (1), 1–15.
- McCain, R.A. (1995b). "Genetic algorithms, teleological conservatism, and the emergence of optimal demand relations: The case of learning-by-consuming". In: Gilbert, N., Conte, R. (Eds.), *Artificial Societies (Proceedings of Second Conference on Simulating Societies, Siena, Italy, July 1993)*. UCL Press, London.
- McCain, R.A. (2002a). "Taste formation". In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham, pp. 682–688.
- McCain, R.A. (2002b). "A sentimental time machine: Old picture postcards and old covered bridges". Available at <http://william-king.www.drexel.edu/top/bridge/CBcardz.html>, as of October 8, 2002.
- Moncrieff, E. (2002a). "Rediscovered Rubens could make £6 million. The early masterpiece had been mis-attributed for 220 years". *The Art Newspaper*. Available at <http://www.theartnewspaper.com/news/article.asp?idart=8994>, as of December 27, originally posted on 4/8/2002.
- Moncrieff, E. (2002b). "Recently discovered Rubens becomes most expensive Old Master painting ever sold. Unidentified private buyer pays US \$76.73 million". *The Art Newspaper*. Available at <http://www.theartnewspaper.com/news/article.asp?idart=9761>, as of December 27, originally posted on 7/12/2002.
- Mossetto, G. (1993). *Aesthetics and Economics*. Kluwer, Dordrecht.
- Nietzsche, F. (1871). "The Birth of Tragedy". Malaspina University College, Nanaimo, BC, Canada. Translated by Johnson, I.C. Available at [http://www.mala.bc.ca/~johnstoi/Nietzsche/tragedy\\_all.htm](http://www.mala.bc.ca/~johnstoi/Nietzsche/tragedy_all.htm), as of November 6, originally posted on 12/1/2000, last modified on 6/1/2003.
- Nozick, R. (1981). *Philosophical Explanation*. Belknap Press, Cambridge, MA.
- Nozick, R. (1989). *The Examined Life*. Simon and Schuster Touchstone Books, New York.
- Nozick, R. (2001). *Invariances*. Belknap Press, Cambridge, MA.
- Pearce, M., Wiggins, G.A. (2002). "Aspects of a cognitive theory of creativity in musical composition". 2nd International Workshop on Creative Systems. European Conference on Artificial Intelligence. Lyon, France.
- Radford, M. (2004). "Emotion and creativity". *Journal of Aesthetic Education* 38 (1), 53–64.
- Robbins, L. (1952). *An Essay on the Nature and Significance of Economic Science*, second ed. MacMillan, London.
- Scitovsky, T. (1972). "Arts in the affluent society: What's wrong with the arts is what's wrong with society". *American Economic Review* 62 (2), 58–69.
- Scitovsky, T. (1976). *The Joyless Economy*. Oxford University Press, New York.
- Shanahan, J., et al. (1978). "The consumption of music: Integrating aesthetics and economics". *Journal of Cultural Economics* 2 (2), 13–41.
- Shook, J.R., Ghirdelli, P. (2004). "Editorial note". *Contemporary Pragmatism* 1 (1), 2.
- Simon, H. (2001). "Creativity in the arts and the sciences". *The Kenyon Review* 23 (2), 203–221.

- Simonton, D.K. (2001). "Creativity as cognitive selection: The blind-variation and selective-retention model". *Behavioral and Brain Sciences* 24, 554–556.
- Spanier, S. (2003). "Heritage Lottery Fund makes huge award to keep Raphael in UK. But the vendor may not accept the present offer". *The Art Newspaper*. Available at <http://www.theartnewspaper.com/news/article.asp?idart=11260>, as of December 27, originally posted on 7/25/2003.
- Stebbing, P.D. (2004). "A Universal Grammar for Visual Composition?" *Leonardo* 37 (1), 63–70.
- Stevens, D. (1985). "The social efficiency of arts addiction". In: Hendon, W., et al. (Eds.), *Proceedings of the 3rd International Conference on Cultural Economics and Planning: Governments and Culture*. Association for Cultural Economics, pp. 43–53.
- Stigler, G., Becker, G. (1977). "De gustibus non est disputandum". *American Economic Review* 67 (2), 67–90.
- Throsby, D. (2001). *Economics and Culture*. Cambridge University Press, Cambridge.
- Throsby, D. (2003). "Determining the value of cultural goods: How much (or how little) does contingent valuation tell us?" *Journal of Cultural Economics* 27, 275–285.

## VALUE AND THE VALUATION OF ART IN ECONOMIC AND AESTHETIC THEORY

MICHAEL HUTTER

*Universität Witten/Herdecke, Witten, Germany*

RICHARD SHUSTERMAN

*Florida Atlantic University, Boca Raton, Florida, USA*

### Contents

Abstract	170
Keywords	170
1. Premodern theories	171
2. Art in early subjectivist theories	174
3. Art in 19th century economics and aesthetics	183
4. Art in economic theory until the 1970s	187
5. Art in aesthetics until the 1990s	189
6. Varieties of artistic external effects in contemporary economics	193
7. Varieties of artistic value in contemporary aesthetics, and their economic valuation	197
8. The logic of aesthetic and economic evaluation	201
References	205

**Abstract**

Artistic value played a minor role in classical philosophy, but moved to center stage in 18th century aesthetic theory and also played a role in moral philosophy. The value of Art and the process of its valuation has remained an indispensable subject of modern aesthetics, while economists have excluded these topics from consideration. Recent attention to “external effects” has opened new ways of interpreting artistic value in a manner consistent with economic theory. Sections 1–5 narrate historical positions in both disciplines, many of which have left their imprint on current analysis. Sections 6–8 focus on the contemporary discussion of artistic values and their logic of evaluation in economics and in aesthetics.

**Keywords**

artistic valuation, consumption skills, tastes, use value

*JEL classification:* Z11

## 1. Premodern theories

Both economic and aesthetic theory, in the most strict and technical sense, are a product of modernity. Both theories have been very powerfully shaped by Greek philosophy, particularly by the works of Plato and Aristotle. Therefore we must begin our study with ancient theories of economics and aesthetics ‘*avant la lettre*’.

Plato (427?–347 BC) [see Plato (1997)] provides what is generally regarded as the first substantial theory of art. Though the value of art and beauty are usually closely related, Plato treats them very differently. Beauty plays an extremely positive role in his philosophy, serving as an exemplar of the very highest level of the ideal Forms and associated with truth and the good. Moreover, beauty is seen as the inspiration and goal of philosophy itself. Plato’s *Phaedrus* characterizes beauty as the clearest, most understandable Form and his *Symposium* describes the philosophical quest as an ascent from the love of beautiful bodies to the love of beautiful deeds, discourses and thoughts, and finally to a vision of Beauty itself from which the philosopher can give birth to the beautiful.

In contrast, art – in our modern sense of the fine Arts – fares miserably. Plato’s *Republic* defines such arts in terms of “mimesis” (typically translated as imitation though sometimes also as representation), while the general Greek term for art (*techne*) had a much wider meaning, denoting any systematic skill or form of knowledge. Plato denounces the mimetic arts, such as poetry, drama, painting and sculpture, as an imperfect imitation of the forms of the phenomenal world, which for him are themselves but a distorted imitation of the ideal rational Forms that constitute true reality. Art is thus condemned both ontologically and epistemologically as an imitation of an imitation that distorts the truth it pretends to present. Plato further condemns mimetic art on psychological, ethical, and political grounds. By appealing to the lower, emotional part of the soul and inciting it with passions, art disrupts the rational psychological order that should prevail and thus corrupts character and leads to improper behavior. Since political order and justice are intimately interdependent with the order of proper moral psychology, mimetic art – at least the popular genres criticized by Plato – represents a grave political danger; the vivid depiction of war’s horrors and of love’s delights could, for example, sway soldiers from their duties.

Plato never really considered art’s value on aesthetic grounds for to do so would establish criteria that might give it more autonomy, and art’s autonomy and social prestige were exactly what Plato wanted to undermine in order to establish the hegemony of philosophy. This deprecatory strategy was useful because the nascent and still fragile discipline of philosophy, in order to establish its authority, needed to struggle against the cultural prestige of the artists, particularly the poets, who were recognized as repositories of ancient wisdom. Defining art as an imperfect imitation not only helped to demean art but also to conceal the fact that Plato’s philosophy itself imitated many aspects of art – the concern for rational form and coherence, the satisfactions of imagination and contemplation of form, and the interpretation of the meaning of experience and events [see Dewey (1987); Shusterman (1992)].



Plato's political philosophy includes a similarly detailed discussion of economic theory and practice. The art of managing a household or the economic affairs of a city-state ranks among the more highly valued theoretical arts.<sup>1</sup> Among the lower practical arts, there are the skills of producing or building objects, the skills of applying them, and the skills of acquiring them, either by conquest or by exchange. While all these skills are necessary for the maintenance of the city-state, they do not contain much of the quality of the ideal, divine Form. But, at least, they do not disrupt the political natural order.

Aristotle (384–322 BC), who had been Plato's disciple for two decades, maintained the Platonic notion of an ideal, divinely inspired state of human behavior and political organization. With his works, the proportion between the different arts which make up the city-state gained in prominence. All human action contributes to the realization of that state of the community which corresponds to the divine and, thus, to natural order. Ideal wisdom may be perfect self-reflection. But humans need a balance between theoretical (gnostiké) and practical (praktiké) wisdom. "Oikonomía", the art of managing a public household, qualifies as a theoretical art. Among the practical arts, there is a distinction between actions and occupations that generate some external result (poietiké) and actions which employ these results. Only the latter are truly practical and ethical because they have their end and value in themselves.

Aristotle recognized the Platonic distinction between the employment and the acquisition of material objects. He transformed it in a way which remained paradigmatic for the next two millennia: the "natural" use of objects is their immediate employment. The "unnatural" use of objects is their exchange. Unnatural uses are legitimate for a community since resources are distributed unevenly throughout a territory. But since they do not follow natural proportion, they are limitless, lead to excess and thus to disorder. "Value in exchange", then, is interpreted as a limit to the natural "values in use". Beyond that limit, the accumulation of treasure, measured in a money commodity, becomes an end in itself. The lending of money sums against interest is an example of economic activity with negative value.

If practical arts are exercised with virtue (areté), i.e. in accordance with the divine laws of ethical and political order, they contribute to the well-being of the city-state. Yet, they remain far from those activities which lead to immediate experiences of "eudaimonía", i.e. the ability of humans to make their souls the receptacle of divine wisdom. In the case of the Arts (in the modern sense), the judgment is more favorable. In general, the "kátharsis" induced by them is valuable for both the individual and society, because it allows such negative emotions as pity and fear to be stirred up and then expurgated within the protected context of art's experience rather than having them spill over into real life where they could wreak psychological, ethical, social, and political havoc. If art's most obvious, general, and traditionally affirmed values can be summed up under the categories of pleasure and use, Plato recognized the pleasures but deemed them

<sup>1</sup> The term "techné" is used for any skill, be it technical, artistic or theoretical. The arts as discussed above – tragedy, music, sculpture – are dealt with as phenomena *sui generis*.

base and corruptively dangerous, just as he argued that art had negative utility in the cognitive, psychological, ethical, and socio-political spheres, while Aristotle defended the legitimacy of art's pleasures and their positive value.

Beyond catharsis, Aristotle argued for both the cognitive value of mimesis and the psychological, ethical, and social value of art's emotional arousal. Claiming that imitation was a natural and primary means of human learning and also a natural source of human delight, he further argued that art had important cognitive value because it imitated the essential and universal rather than mere contingent superficialities. That is why he described poetry as being more philosophical than history (*Poetics*, 1448b, 1151b) [see Aristotle (1947)]. He singles out music as a core discipline of education. Music, moreover, generates images (*homíoma*) that come close to "true nature" and thus succeeds in moving the soul (*Politics*, 1340a). Therefore, music is an activity which is able to combine banal pleasure with blissful happiness, a feeling of a higher order.

Aristotle's explicitly introduced formalistic or compositional principles for the analysis and evaluation of artworks, most notably for works of tragedy. More than merely advocating a general idea of organic unity (the need for a satisfying whole to "present a certain order in its arrangements of parts" and be of an appropriate "magnitude"), Aristotle's compositional principles refer to the various elements that form the artwork, classifying them in terms of whether these elements pertain to the object represented (plot, character, thought), the means of representation (diction and melody), or the manner of representation (spectacle and narrative versus dramatic role playing) (*Poetics*, 1447a–1450b). These elements are differentially valued. For example, in tragedy, the element of plot is clearly asserted as the most important. The different elements can be used to evaluate different genres. Tragedy is valued as nobler than comedy since tragedy represents nobler plots and characters and has consequently less vulgar diction as well. Aristotle evaluated tragedy as a higher mimetic form than the epic, which also had nobility of action, plot, and diction, because it had more positive elements, such as music and spectacle, as well as more unity through its narrower focus or scope. More important than any specific list of formal elements and their evaluative import was, in the long run, Aristotle's suggestion that there exist criteria for evaluating art in terms of its formal composition and that they are not reducible to ontology, epistemology, psychology, morality, or politics. The belief in such criteria – significantly linked to properties of form, expression, and quality – played an important role in the modern theories of the aesthetic that began to be formulated in the late 17th century.<sup>2</sup>

Despite the recognition of the values of poetry, music and tragedy, Aristotle continued the Platonic strategy of subordinating art to philosophy. Although art affords pleasure through cognition, philosophy's truths and pleasures are clearly asserted as superior and

<sup>2</sup> A prominent example for a frequent practice in the period is Roger de Piles (1635–1709) who suggested in his *Cours de Peinture par Principes* (1709) to decompose paintings into four fundamental characteristics and rate them each on a scale between zero and twenty [de Piles (1743)]. See Ginsburgh and Weyers (2003) and De Marchi (2006).

more rewarding, its contemplative activity of “theoria” being the “summum bonum” of human life. Just as the theory of catharsis emphasized that art’s passions are aroused in a special artworld context so that they can be harmlessly purged without harming real-life character or society, so Aristotle’s continued to interpret art as “poiesis” as contrasted with “praxis”. Thus, he further isolated art from the sphere of ethics and social and political practice. Art as “poiesis” means external making, the creation of objects outside the self. The end and value of the making is in the objects made. In contrast, praxis or ethical action have their ends and values in themselves and in their agents. They both derive from the agent’s character and reciprocally help shape it (*Nichomachean Ethics*, 1140a1–1140b25).

## 2. Art in early subjectivist theories

Two thousand years later, the terminology of philosophical thought was essentially the same. But the interpretation of the basic assumptions had changed radically. In ancient philosophical theories, properties of value like beauty or utility were conceived as objectively inhering in the objects of which they were predicated. Beauty and utility were considered real properties of things rather than a transactional product that essentially depended on the subjective experience of the beholder. This so-called “realist” view of value continued to be used in medieval and Renaissance philosophies of beauty and of states craft. Earlier works were based on Aristotelian thought, while later texts emphasize abstract measurement, following the rediscovery of Plato’s treatises in Western Europe. Since about 1500, the “nominalist” counter current of Humanist thought began to question the objective existence of ideal Forms, culminating in the strictly empiricist epistemology of John Locke (1632–1704). According to Locke, the mind is a clean slate which is filled with the sensory impressions of the outside world. This is in contrast to Descartes (1596–1650) who, inspired by the scientific advances of Galileo and others, reinterpreted the material world and its objects in essentially physical terms of mathematically measurable extension as could be mathematically measured. As such measurable physical properties became the paradigm of the real, sensory properties such as color, taste, and texture came to be regarded as secondary, less objective properties, while aesthetic properties seemed even more subjective.

Such was the intellectual setting in which a specific theory of taste as an explanation of aesthetic value and a specific theory of self-interest as an explanation of economic value emerged.

For aesthetic theory, Anthony Ashley Cooper, 3rd Earl of Shaftesbury (1671–1713), a student of Locke, was a pivotal figure. As part of his generally Platonic stance, he held a realist theory of beauty. But his Lockean convictions about cognition led him to advance the idea of a special faculty of taste that made moral judgments and aesthetic judgments by respectively discerning the Forms of Good and Beauty. This special mental attitude allows man to properly grasp and appreciate beauty in a disinterested

manner, without the desire to possess or control it. We should not, for example, in contemplating the beauty of a human form allow erotic interest to intervene, since such “desires, wishes, and hopes . . . are . . . no way suitable . . . to your rational and refined contemplation of beauty” [Shaftesbury (1711/1964, p. 126)]. It follows that a connoisseur desirous of acquiring a beautiful painting does not have the right attitude to properly judge its beauty. Shaftesbury’s notion of disinterestedness, in the versions developed by Kant and Schopenhauer, became one of the most influential ideas of modern aesthetics.

The concept of taste had also undergone a switch from external, realist meaning – as in the taste or manner of an artist or a style-period – to the internal, subjectivist meaning of taste as a sensory or perceptual competence. As an individual skill, taste could be conceived as something that is brought to excellence by one class, namely Shaftesbury’s own class of landed gentry. The members of this class had the means to develop their sense of taste for the beautiful. More importantly, they fulfilled one condition which practitioners of all trades and other commercial occupations did not fulfill: they could afford disinterested contemplation. Shaftesbury’s firm Platonic assertion of art’s value, and beauty in general, continued to resist the increasingly empiricist tendencies of British thought.<sup>3</sup>

Early modern economic thought was already deeply suspicious of such hierarchical constructions. Economic pamphlets were published in immediate response to Shaftesbury’s aesthetic theory. Prime example is Bernard Mandeville’s (1670–1733) *Fable of the Bees: Or, Private Vices, Publick Benefits* [Mandeville (1714/1988)], where the model of a state in which everyone consumes in appropriate measure, thus increasing the total volume of blissful happiness experienced in contemplating “real” beauty or moral action is rejected and replaced with a model in which all individuals seek to maximize their pleasures without regard to virtue, and yet the state prospers and grows, as does a bee-hive, filled with creatures that lack the slightest idea of virtue. Nineteen years later, Mandeville (1670–1733) published a second volume called *Fable of the Bees*, in which he continued his attack on Shaftesburian theory, this time in immediate aesthetic terms. The volume contains six conversations between three spectators of paintings of

<sup>3</sup> Shaftesbury was also responsible for introducing to British aesthetic theory the evaluative property of sublimity, first developed by an ancient unknown author (thought to have lived in the first century A.D.) who is known as Longinus [Longinus (1992)]. Shaftesbury regards the sublime as a kind of beauty, but later theorists, treated the sublime as an alternative value to beauty and in some ways more powerful or higher than it. Burke (1998) regards the sublime as “the strongest emotion which the mind is capable of feeling” because it is based on the “passions which concern self-preservation” (pp. 86, 97), while Kant describes it as “arising from a higher intellectual feeling” than the beautiful (p. 33).

Edmund Burke (1729–1797), who rejects the doctrine of a distinct internal faculty of taste, explains our judgments of taste through our ordinary sensory and mental capacities and in terms of our experiences of pleasure and pain. Burke (1757/1998) distinguishes between positive pleasure which engenders the feeling of beauty and the pleasure of delight (deriving from the removal of pain or danger, i.e., the threat of pain) which inspires the experience of the sublime. In contrast to Shaftesbury and later idealist aestheticians, Burke has a distinctly embodied approach to aesthetic value.

Christ's birth. They discuss the merits of works done in Italian and in Dutch style, and they do so in a way which discredits the Shaftesburian valuation of features that cater to artificial taste and which favors the valuation of features that render sensory impressions and thus cater to common sense.<sup>4</sup>

Mandeville was able to make contributions to an emerging political economy as well as to an emerging aesthetic philosophy, but that accomplishment is outdistanced by the works of David Hume (1711–1776). Hume's *Political Discourses* [Hume (1752)] contain essays with significant contributions to theoretical economics, most notably *Of Money*, *Of Interest*, *Of Commerce* and *Of the Balance of Trade*. At the same time, Hume's *Of the Standard of Taste* [Hume (1757/1963)] provides the most important British text on the evaluation of art.

Hume aimed to determine an objective standard for what he regards as the clearly subjective judgment of taste, which as a judgment of "sentiment" rather than "fact" admits of no objectively "true and decisive standard", even if there are objective properties in artworks that tend to elicit taste's sentiments. "Beauty is no quality in things themselves; it exists merely in the mind which contemplates them" (p. 234). While Hume recognized the diversity of evaluative judgments, he insisted, on the other hand, that some judgments (e.g., Milton's superiority to Ogilby) are undeniably true and that some individuals are better than others at evaluating art. Hume's strategy for defining a standard of taste is to link judgments of taste to judgments which do have a determinate standard "in real existence and matter of fact", and his crucial device for this linkage is the consensus of good critics – "arbiters acknowledged by universal assent to have a preference above others" [Hume (1963, p. 248)]. The standard of taste for establishing the value of an artwork is thus determined by the consensus of sentiment of good critics regarding that work, and, for Hume, the questions of who these critics are and what qualities they must have "are questions of fact, not of sentiment".

The five requisites Hume lists for good critics are "delicacy of imagination" (essentially a matter of perceptual acuity and sensibility to fine discriminations), "practice in appreciating good artworks", "experience in their comparative assessment", a "mind free from all prejudice", and "good sense" (pp. 239–246). Hume's evaluative theory reflects the liberal dilemma of wanting to guarantee both freedom of taste and an authoritative standard to ensure cultural coherence and stability [Shusterman (2002, pp. 93–107)]. Personal freedom of sentiment in evaluation is at least preserved in one's free decision to submit one's judgment to the authoritative standard set by those recognized as superior judges, the good critics. The parallel of this solution to the political solution of representative democracy with only a partial franchisement of the electorate, which was the political system of Hume's Britain, should be obvious.

Given the empiricist and subjectivist premises of Hume's philosophy, a remarkable similarity between aesthetic and economic valuation comes to light. As Schumpeter (1883–1953) has noted, the aesthetic theory in question can be seen to explain the objective fact that a work of art is considered as "beautiful" by the subjective valuations

<sup>4</sup> See Solkin (1993, pp. 13–19).

of the members of a given social group, much as the “analogous economic theory” explains the fact of market prices by subjective valuations of the individuals participating in a market: “In both cases subjective valuation creates the objective value” [Schumpeter (1954, p. 127)].

A third figure of transition is Francis Hutcheson (1694–1746). His early fame was based on *An Inquiry into the Origin of our Ideas of Beauty and Virtue* [Hutcheson (1725/1971)]. His main work, *A System of Moral Philosophy*, appeared posthumously in Hutcheson (1755/2005). Hutcheson picked up where Shaftesbury had left off.<sup>5</sup> Shaftesbury had been able to turn the ancient Platonic notion harmony and equilibrium of the soul into a modern, Christian notion of benevolent love. The principal virtue in man’s dealing with others is no longer justice and temperance but benevolence. Thus, “the motive of benevolence becomes the key to goodness” [Taylor (1989, p. 258)].

Hutcheson proposed that society is held together by two “moral principles”, namely “moral sense” and “self-love”. Two sentiments, benevolence towards others and self-interest, correspond to the two principles. They compete with one another and constitute an equilibrium. Furthermore, they form a hierarchy: moral sense generates more intensive pleasure than self-love. Hutcheson’s construct of “moral principles” bridges the gap between the two widely distinct sensations of blissful happiness and common pleasure. Benevolence is motivated through religion and thus linked to God. Through benevolence, “we participate in god’s plan through re-engagement” [Taylor (1989, p. 265)]. Such participation can take the traditional form of philosophical contemplation, but also the form of aesthetic contemplation.

Benevolence competes with self-love as a source of pleasure. Thus, in the final analysis, self-love does not differ from benevolence in its basic ability to contribute to pleasure (or happiness). To exemplify the operation of the moral sentiment of pleasure, the valuation of art is invoked: beauty is an expression of divine order. Beauty, just as virtue, triggers a particular sensation of joy or bliss, different from the everyday pleasure reached by the satisfaction of self-love. Thus, the subjectivist notion of taste connects with the sentiment of benevolence.

Hutcheson deploys the idea of taste, which he ascribes to a specific internal sense of beauty, in a thoroughly empiricist fashion. Beauty is not inherently in the object through its participation in a Platonic Form, it instead resides in the empirical experience of a subject’s mind: it is an “idea rais’d in us” [Hutcheson (1726/1971, p. 7)], an experienced feeling of pleasure that is caused by properties of a contemplated object and that arises in an essentially passive, automatic reaction just as our external senses automatically generate ideas in us from the properties they perceive. The appreciation of beauty thus requires no specially acquired knowledge or attitude, and, since it is grounded wholly in our shared internal sense of beauty and shared sensory faculties, our judgments of beauty should be shared. This means that though beauty is subjective in the sense of

<sup>5</sup> In fact, the subtitle to his *Inquiry* reads: “. . . in which the Principles of the late Earl of Shaftesbury are Explain’d and Defended, against the Author of the *Fable of the Bees*”.

existing in subjective experience, it can be objective in the sense of being widely shared rather than individualistic. For Hutcheson, the valuable triggering property of our experience of beauty is unity in variety, and he explains different judgments of taste in terms of differences or defects in people's sensory acuity and in terms of association of ideas that can distract an individual from what the object actually presents to his or her internal sense of beauty.

Self-love does not differ from benevolence in its basic ability to contribute to pleasure. In Hutcheson's *A System of Moral Philosophy* [Hutcheson (1755/2005)], the "public good" of society is attained as the sum of the pleasures attained by the members of the society. Given these conditions, it is the duty of a government to promote the common good with the aim of creating the greatest happiness for the greatest number.<sup>6</sup>

Individual ethical action takes a surprising turn in Hutcheson's interpretation: costly and beautiful commodities lay the grounds for a sense of community by stimulating the same feelings in others. By proposing aesthetic delight as the central force responsible for social cohesion, Hutcheson found a way of defending luxury, or any other kind of conspicuous consumption: it is not vice but the Ideas "of Friendship, of Love, of communicating Pleasure to others" which motivate such expenses and thus contribute to the public good [Solkin (1993, p. 83)].

The next generation of authors was not engaged in both strands of theory anymore. The "separation at birth" [Guillory (1993, p. 303)] had taken place. Both economic and aesthetic theory continued to assume a strictly subjectivist epistemology, but economic theory focused on self-interested action while aesthetic theory focused in its complement, disinterested contemplation. The process of separation can be observed particularly well on the side of economic theory. We have the unusual case of two theories, written by one author within the time span of two decades.

Adam Smith (1723–1790), student of Hutcheson and successor to his chair in Glasgow, published his *Theory of Moral Sentiments* [Smith (1759/1982)] in 1759. Following Hutcheson as well as Hume, two basic human propensities are assumed: "fellow-feeling" and "self-love". Man's disposition for sympathy comes under strict subjectivist scrutiny. Humans are limited to their own impressions and imaginations in feeling the distress or the happiness of their fellows. Mutual sympathy is a result of mutual inaccessibility. Humans also need to gain sympathy from others. The rules of propriety grow out of the need for mutual respect. Moreover, the need to secure an adequate measure of fellow-feeling in others is interpreted as a major motive of individual action [Agnew (1986, pp. 177–181)].

Self-love leads to pleasure in activities that are immediately useful for oneself. Smith identifies – as he believes, for the first time – a kind of pleasure that is distinct from mere physical satisfaction. Such pleasure is associated with the principles of beauty

<sup>6</sup> Hutcheson proposes mathematical formulae with which to calculate this quantity, but omits these in the fourth edition of *Inquiry* [Hutcheson (1738/1971)] because, as he states in the preface, "they have proved to be useless".

and elegance. Beauty and elegance, in turn, are gained through form and color on the level of perception, and through variety, fitness and imitation on a higher level of interpretation. Fitness, for instance, leads to an aesthetic appeal in objects of ingenuity and utility. The source of additional value lies in the degree of an object's fitness for the divine plan which is manifest in every natural and social event. Fitness reflects "the regular and harmonious movement of the system, the machine or oeconomy by which it is produced" (p. 183). Objects that display such values of form or design "strike the imagination as something grand and beautiful and noble" (p. 183), and that is what makes them precious. Imitation generates the pleasure of beauty in a similar manner. Smith notes that we experience amazement in seeing an object of one kind represent an object of a different kind. Such imitation requires artifice, ingenuity and imagination. He cites the example of a Dutch still life which is valued more highly than the carpet which is represented in the painting.<sup>7</sup>

The additional value thus identified plays a central role in explaining the force which drives the economic process: according to Smith, the rich "select from the heap what is most precious and agreeable" (p.184). In order to attain the grand, beautiful and noble objects which will secure them the attention and approbation of their fellows, they not only employ those who labor to provide the means for such purchases, they also make improvements in the production processes in order to increase their own buying power. The increase in productivity leads to further growth. The quest for beauty is turned into an explanation of wealth [Guillory (1993, p. 312)].

Seventeen years later, Smith published his second model of explanation, *An Inquiry into the Nature and Causes of the Wealth of Nations* [Smith (1776/1976)]. In the meantime, he had been exposed to another explanation for fueling the machine of commerce: Physiocratic French authors like Francois Quesnay (1694–1774) and Anne Robert Jacques Turgot (1727–1781) assumed that the produce of nature, standardizable in units of "wheat", constitutes the primary, external source of an economy's wealth. Smith adopted the approach of using the measurability of production input factors to establish a firm link with market value. But he extended the source of value beyond the "produce of land" and ultimately shifted it to human labor. Labor is a factor whose external nature remains ambiguous. It consists in the observable exertion of human force for periods of time, but it also consists in the mental opposite of pleasure, namely pain. Labor thus fulfills four requirements: it is, at least in principle, measurable, it legitimizes property, its "toil and trouble" mirror the pleasure enjoyed in consumption, and it contains a moral quality because work is a virtue. Smith has not abandoned his older belief in the propriety of a society founded on sympathy, but he has identified a separate realm, the realm of commercial society. Commercial society is organized around the valuations of the market. The quantity of labor is an "analogue of the standard of taste, a standard which is both transcendent and immanent" [Caygill (1989, p. 95)]. This absolute standard is exposed to the fluctuations of market valuation. The notions of "real"

<sup>7</sup> See Berg (2002) and De Marchi and Van Mignoet (1999).



and “nominal” price correspond to the distinction between the impartial spectator and the sum of empirical spectators already proposed in the *Theory of Moral Sentiments*. The gravity of “natural price” corresponds to Hume’s “mark of conformity”.<sup>8</sup> The separation of the aesthetic and the economic realm is completed, but the mode of analyzing the latter is still shaped by the mode developed for the former.

The literature of classical English economic theory that followed Smith did not retain the ambivalent connotations of Smithian labor, and the references to a society connected by mutual attention and attraction. The idea of differently valued kinds of pleasure is abandoned. New value is generated on the production side by harvesting a life-form, while the demand side of the market is assumed to remain basically passive. David Ricardo’s (1772–1823) *Principles of Political Economy and Taxation* [Ricardo (1817)] completed the formulation of a theory that was based on the absolute value of human labor. Ricardo explicitly excluded objects like “rare statues and pictures, scarce books and coins, wines of a peculiar quality” whose value depends solely on “the varying wealth and inclination of those who are desirous to possess them” (1973, p. 6).

In aesthetic philosophy, the positions of Hume and Immanuel Kant (1724–1804) mark the shift from empirical subjectivism towards analytical idealism. Kant responded to Hume’s empiricist skepticism that held there are no objective truths beyond the consensus of one’s fellows and that most values are held not rationally but because of habit or custom. He split the perception of phenomena into two levels: on a secondary level, sensations caused by the object are perceived. On a primary level, our subjective apparatus, the mind, is capable of ordering the perceived phenomenon in certain relations. Space and time are such “forms of intuition” (Formen der Anschauung), and so are 12 priori concepts. These forms cannot exist apart from human experience because they take place in the minds of the individual in a society. But, under favorable conditions, they can be recognized among members in the society, leading to “subjective universality”. Judgments of taste provide a particularly good example of universal consent because they take place under conditions of disinterestedness, unperturbed by the distortions of profit and desire.

Kant’s aesthetic theory is worthy of especially extended attention, not only because it has been historically so influential but also because it remains the dominant orientation (for better and for worse) of most contemporary philosophy of art. Kant’s pivotal position derives in part from the way he builds on insights from both the empiricist British philosophers and the rationalist continental philosophers that preceded him. He develops ideas of disinterestedness, sublimity, and correctness of taste that Shaftesbury, Burke, and Hume brought to the fore, but he also builds on the rationalist tradition of Leibniz (1616–1716), as applied to aesthetics by Alexander Baumgarten (1714–1762) who coined the term “aesthetics” and essentially founded it as a distinct philosophical “subdiscipline”.

Alexander Baumgarten (1714–1762), a disciple of Leibniz and Wolff (1679–1754), defined it in very broad terms as a science of sensory perception that would parallel

<sup>8</sup> See Caygill (1989, pp. 85–97).

logic's science of conceptual thought: "Aesthetics (as the theory of the liberal arts, the science of lower cognition, the art of beautiful thinking, and the art of analogical thought) is the science of sensory cognition". "The end of aesthetics", he continues, "is the perfection of sensory cognition as such, this implying beauty" [Baumgarten (1750–1758, §§1, 14)]. Artworks, as objects purposefully and carefully crafted to achieve beauty – which was associated at that time with harmonious, rationally proportioned form – seem paradigmatically suitable for sensory cognition. This general meaning of aesthetics as the theory of sensible cognition as well as the narrower study of beauty, sublimity, and arts remains saliently present in Kant. Much later, aesthetics became restricted to its contemporary meaning as the philosophy of art, beauty and related aesthetic concepts.

Kant's *Theory of Aesthetic Judgment* [Kant (1790/1986)] provides a solution to the evaluative problem of holding that aesthetic judgments are essentially subjective yet nonetheless can command, in a necessary way, universal assent. "The judgment of taste . . . denotes nothing in the object, but is a feeling [of pleasure or displeasure] which the Subject has of itself and of the manner in which it is affected by the representation" of the judged object (1986, pp. 41–42). Yet such judgment displays "the necessity" of "universal assent like an objective principle" (pp. 41–42, 84–85). Rather than rely on a group of good critics, Kant argues that anyone can in principle perform an accurate pure aesthetic judgment by exercising the proper aesthetic attitude. This attitude requires disinterestedness, which Kant (p. 43) describes as "indifference" or lack of concern for "the real existence of the thing" judged. For example, we should not care whether we are observing a real landscape or a mere illusory appearance of one. Aesthetic pleasure is thus distinguished from pleasure in the good and in the agreeable because these latter involve interest and desire. A pure aesthetic judgment also requires refraining from the use of concepts and functionality as determining grounds for one's evaluation. Instead one's attention should be directed exclusively to the form of the aesthetic object in terms of its presentation of finality, without regard to any function.

Pure aesthetic judgments are always particular and cannot be rule governed. This is why Kant treats aesthetics as critique rather than science: a science needs to introduce general concepts, while a critique leaves the determining ground of the judgment to the pleasure of the subject who contemplates – with the proper attitude – the object of taste. Diverse concepts and interests can lead to disagreement in judgments of taste, but if we detach our perception from concepts and interests, we can affirm that pure aesthetic judgments claim a necessary "subjective universality" since they rely on the "mere nature" of "the Subject's faculties" that is shared by all humans (pp. 51, 212–223). Kant argues that when properly contemplating the form of a good aesthetic object without concepts or interests and without any regard to "charm or emotion" (p. 64), any person should necessarily get pleasure from the enjoyable "free play" (pp. 86, 88) of the cognitive faculties, in which one's imaginative experience freely con-forms (through form) to the rationality of the understanding without being "forced" to conform in terms of a specific prescribed concept of understanding.

Kant's model objects for pure aesthetic judgment belong to nature whose beauty and sublimity he valued higher than art. Judgments of taste regarding art cannot be absolutely pure and free from concepts, since they always involve the concept of art (pp. 34, 166–167). To properly judge a landscape painting we need to consider it as an artwork and not simply as the appearance of landscape. The introduction of conceptual knowledge for art means that we can also no longer expect universal convergence, since some individuals might lack the requisite knowledge. Kant held the appreciation of nature to be also morally more beneficial, because art, unless “brought into combination with moral ideas”, tends to degenerate into mere “diversion” that “renders the soul dull and the mind dissatisfied with itself” (p. 191).

Kant also ranked the different arts, unequivocally giving poetry the highest place by all important criteria. When considered with respect to “charm and mental stimulation”, music would rank next. But Kant goes on to argue that music's value is much diminished when “we estimate the fine arts by the culture they supply to the mind”. Here, “since it plays merely with sensations”, music has the lowest place among the fine arts (pp. 193–195). The formative or plastic arts, with painting judged foremost among them because of its formal and ideational power, thus can ultimately be ranked higher than music because they not only please but promote “the urbanity of the higher powers of cognition”. Moreover, “music has a certain lack of urbanity” because “its intrusive loudness forces itself on free subjects who would prefer not to hear it” (pp. 194–196).

Though Kant seemed to give primacy to the aesthetic experience of nature – both beautiful and sublime – over art, he nonetheless had a very high regard for fine art, whose creation, he insisted, “needs genius”, an ability to create something original and exemplary rather than merely producing something mechanically according to a given rule (p. 172). A true work of fine art, he explains, requires “Soul (Geist) in an aesthetical sense, [which] signifies the animating principle in the mind [. . . and] this principle is nothing else than the faculty of presenting aesthetic ideas”. Kant defines an aesthetic idea as a “representation of the imagination which induces much thought, yet without the possibility of any definite thought whatever, i.e. “concept”, being adequate to it, and which language, consequently, can never get quite on level terms with or render completely intelligible” (pp. 175–176). Kant's account of art as requiring the creativity of genius whose aesthetic ideas resisted conceptual formulation yet stirred up deep and fruitful thought was very influential to the romantic movement and contributed to elevating the status of the artist and his work. Art and artist were regarded as belonging to a realm of genius and value that transcended conceptual definition and could not be reduced numerical reckonings.

The Kantian world of aesthetic philosophy is already far away from the Smithian world of political economy. Kant perceived order in the immutable intuitions and concepts that govern the perceptions and thus the actions of all members of society. Smith perceived an additional kind of order in commercial society, where the consensus of evaluating participants leads to the measurable result of prices. Kant bypassed the transitory and accidental level of commercial activity to reach the level of common understanding, Smith added the organizing power of the market to explain the coordi-

nation of those who follow their separate and antagonistic interests. Both of them treated beauty as a special source of pleasure, but Smith accorded little attention to artworks in *Wealth of Nations*, and Kant considered natural beauty and sublimity in important ways superior.

Since the beginning of the 19th century, aesthetic philosophy and political economy have moved along different paths. The separation of the narratives in the following sections reflects that separation.

### 3. Art in 19th century economics and aesthetics

The dominating current in political economy continued to be British in the 19th century. The influence of empirical subjectivism in general and of Hutcheson's maximization of collective pleasure in particular remained strong. Thus, we find a remarkable split in the contributions to political economy: one strand, exemplified by Ricardo, developed a theory of production cost, ultimately based on human labor, but principally concerned with objectively measurable cost. The other strand, exemplified by Jeremy Bentham (1748–1832), James Mill (1773–1836) and John Stuart Mill (1806–1873), developed a normative theory of maximizing the “summum bonum”, modeled as the measurable sum of the pleasures experienced by the members of a society. This Utilitarian credo has a reformist and educational dimension: every person can better his capacity for receiving pleasure, and the state can promote the appropriate institutions and initiate actions that raise aggregate well-being. It also has, at least in Mill's version, an evolutionist dimension, as an endstate of society which is compatible with the stationary equilibrium of society's productive, value-generating factors.

The consumption-oriented strand of theory moved to the center of the literature after the second half of the 19th century, when William S. Jevons (1835–1882) proposed a way to measure the utility contained in objects. The measure relies on new scientific ways of registering physiological sensations, and on the fundamental observation that such sensations decrease in intensity with their time of duration. In consequence, there is no need to measure all of the potential pleasure or pain that flows from an object. It is sufficient to measure the intensity at the margin, i.e. at the point where a specific quantity of that object is exchanged for a sum of money or for a quantity of another object rendering more pleasure. Given such a measure, states of consumption equilibrium can be calculated with techniques of constrained maximization, just as in problems of energy conservation and transformation [Mirowski (1989)], and such states are in fact calculated in markets. There is no need to rely on estimates of cost, or even on metaphysical ideals, because the subjective, physiological effects of pleasure can be measured directly.

Jevons applied his measurement of commodity utility through human physical sensations to the “lowest rank of feelings”. He was interested in connecting the traditional utilitarian discourse around pleasure and pain with a modern reading of utility as an analogue of energy, the fundamental force in the world of nature. He was quite aware

of the effects of aesthetic experiences, but he positioned them in the realm of “sympathy”, where the peculiar conditions of commercial society do not hold. In a manuscript “On the Functions of Music”, he described his own response to music as similar to “the contemplation of subjects of Interest, Beauty or Sublimity . . . a general removal of the mind from its ordinary course of duties and frailties, and its continual mixture of slight pleasures and pains”.<sup>9</sup> He believed that the fine arts are capable of enriching the lives of the members of the working class, but this experience cannot be anticipated and would therefore not sell well. There is a vague link between base utilitarian pleasure and sublime aesthetic enjoyment because social progress might lead to an upgrading from satisfactions of physical need to satisfactions “derived from the beauties of nature and art”.

Jevons’ admiration for the Arts is typical for the English academic tradition of late 19th and early 20th century. Alfred Marshall (1842–1924), for instance, explicitly recommended in his *Principles of Economics* (1890) that one should increase the beauty of things in one’s possession, once the necessities of life are provided: “an improvement in the artistic character of furniture and clothing trains the higher faculties of those who make them and is a course of higher happiness to those who use them” (p. 113). But neither Jevons nor any of the major authors of the following generations gave artistic value a special role in the use or utility value of the consumed set of commodities.

The increasing admiration of the Arts was a result of their increasing relevance in aesthetic philosophy, which will be reported below. That change in attitude toward the Arts was even more pronounced on the Continent. It is therefore instructive to contrast the treatment of marginal utility value in English Political Economy with that in the other two centers of the emerging paradigm of Economics, Vienna and Lausanne.

Carl Menger (1840–1925) contributed a long chapter in *Grundzüge der Volkswirtschaftslehre* [Menger (1871/1968)] to commodity value (Güterwerth). In his construction, the source of value is squarely placed in a psychological dimension. Pleasure value is determined along a subjective ordinal scale of relevance. At the limit, that measure is precise enough to determine the exchange value in the market for a given commodity quantity. He clearly interprets desires (Bedürfnisse) as a purely mental activity. The value of commodities lies in their relative ability to satisfaction a desire (Bedürfnisbefriedigung). The satisfaction of commodities is determined by their position on a scale of importance from vital to trivial, from necessities like eating bread to fancies like tobacco, hunting castles and artificial duck ponds (1968, p. 111). The value of the “commodities of lower order” (Güter niedrigerer Ordnung) determines the value of “commodities of higher order” (Güter höherer Ordnung). But “higher” simply means “earlier”: the final consumption value determines the value of the inputs used in earlier stages of the process leading up to the purchase of the end product. Mental consumption value determines material production value. Art plays no particular role among the examples its objects clearly rank among those of lesser importance, although their rarity might secure them a somewhat higher market price.

<sup>9</sup> See Goodwin, Chapter 2 in this volume.

Léon Walras (1834–1910) taught in Lausanne, but was strongly shaped by French philosophical and economic tradition. In his youth, he showed a strong interest in the Arts. He wrote a novel and an essay on *Philosophie de l'Art*. In this essay, he reduces aesthetic phenomena to ontological causes, namely substance, matter and force. In his *Eléments d'économie politique pure* [Walras (1874)], the problem of value origin recedes into the background. Some kind of sensation motivates consumers to demand limited quantities of commodities at certain prices, but the focus is on the mechanics of states of multiple market equilibria. Markets, equipped with imaginary auctioneers, provide a stable and unique solution to the problem of adoption to changes in productive and consumptive conditions. Artworks may be among the commodities that motivate acts of consumption, but they play no special role in the theory.

As we now turn to the development in aesthetic philosophy, we can easily see how the neglect for the sphere of Beauty and its claims of higher satisfaction is reciprocated with neglect for the sphere of commercial activity.

The Idealist strand of philosophical thought found its most successful continuation in the works of Friedrich Hegel (1770–1831). With respect to aesthetics, his contribution marks a turning point because art supplants nature as the paradigm object of aesthetics. As an expression – along with religion and philosophy – of absolute spirit, art, unlike natural beauty, “is capable of truth”. Therefore, it admits of clearer criteria for erecting aesthetics into a science, which is Hegel’s aim in limiting aesthetics to the philosophy of fine art (Schöne Künste).

Fine art, he argues, proves its “worthiness” for scientific study by distinguishing itself as “free” in contrast to arts “that serve the ends of pleasure and entertainment”. Hegel grounds art’s value in other ends: “its highest task . . . is revealing to consciousness and bringing to utterance the Divine Nature, the deepest interests of humanity, and the most comprehensive truths of the mind. It is in works of art that nations have deposited the profoundest intuitions and ideas of their hearts” [Hegel (1835/1993, p. 9)]. Fine art is thus valued and comparatively ranked in terms of the success of its sensuous representation of worthy collective ideas, both according to the quality of the representation and to the idea represented. The idea has greater importance, since the clarity and well-formed character of the idea is a condition for a clear, well-structured representation.

Hegel has a complex ranking of artistic forms and genres. The lowest form of the hierarchy, which he calls symbolic art, is exemplified by “the primitive artistic pantheism of the East”. In such art, the idea still exists in too much “indistinctness and obscurity” for it to have a fittingly determinate form. Hence it is typically rendered in objects that “exaggerate the natural shapes and phenomena of reality into indefiniteness and disproportion”. The next stage in the historical process is “the classical form of art” exemplified by Greek anthropomorphic sculpture whose sensuous human forms express and fully coincide with the rationality of human mind. Their perfect balance of idea and sensuous representation generates the greatest beauty. But though classical art “attained the highest excellence, of which the sensuous embodiment of art is capable”, Hegel points to a more recent and still higher “romantic form of art”. Romantic art shows the inability of the sensuous to fully capture the Idea in its ideal form which is beyond the

realm of the sensuous. Christian art exemplifies this form in which sensuous images are used to point to a realm of spirit. Hegel, thus ultimately found art's highest value in its promotion of the spiritual truth of the Idea rather than in the mere experience of beauty (pp. 82–87).

Hegel also ranks the genres of art in terms of their potential to serve the Idea and spiritually transcend materiality. Architecture lies at the bottom, followed by sculpture, painting, and music in ascending rank, with poetry at the very top. "Poetry is the universal art of the mind which has become free in its own nature, and which is not tied to find its realization in external sensuous matter, but expatiates exclusively in the inner space and inner time of the ideas and feelings. Yet just in this its highest phase art ends by transcending itself, in as much as it abandons the medium of a harmonious embodiment of mind in sensuous form, and passes from the poetry of imagination into the prose of thought" (p. 96). This is one expression of Hegel's famous thesis of the end of art. In earlier times, man needed art to advance spiritual expression since thought was not advanced enough to express the spiritual without the sensuous. But in the more philosophical "reflective culture" of modernity, art "has lost for us its genuine truth and life" and serves largely for "our immediate enjoyment". Therefore Hegel thinks a science of aesthetics necessary to continue to link art to truth and thus save art from having its value reduced to entertainment (pp. 12–13).

Arthur Schopenhauer (1788–1860) develops another variation of Kantian thought. He rejected Hegel's view of art's historically passing truth, but he introduced human will as a subjective force behind cognition. Materiality is the appearance while will is the fundamental reality, objectified into Ideas. Art's supreme value is in revealing human Will. The Ideas that art expresses, Schopenhauer argues in *Die Welt als Wille und Vorstellung* [Schopenhauer (1819/1966)], are not historical concepts but eternal Ideas, even if art's expression is naïve and provides but a fleeting image, not a permanent universal knowledge for which philosophical reflection is ultimately needed. Aesthetic experience offers a special penetration into reality, because its disinterested, "will-less" contemplation allows art's Ideas to shine forth in themselves rather than being distorted by the practical interests that normally guide our perception. The various arts are ranked on levels of Ideas: architecture ranks lower than the plastic arts of sculpture and painting, and they rank lower than poetry. The highest rank goes to music, "Because music does not, like all the other arts, exhibit the Ideas or grades of the will's objectification, but directly the will itself, . . . it is the most powerful of all the arts" (1966, p. 448).

Sharing Schopenhauer's extremely high valorization of art, Friedrich Nietzsche (1844–1900) strongly contested his predecessors's thesis of art's will-less disinterested contemplation, mordantly mocking the dogma of disinterestedness as an expression of philosophers' prudishness, innocence, and their second-hand, spectators' view of art which he contrasts to the creative experience of the artist. The power of art and beauty, Nietzsche argues in *The Birth of Tragedy* [Nietzsche (1872/1956, p. 239–240)], derives not from disinterest but rather from "the excitement of the will, of 'interest' ". "When our estheticians tirelessly rehearse, in support of Kant's view, that the spell of beauty en-

ables us to view even nude female statues ‘disinterestedly’ we may be allowed to laugh a little at their expense. The experiences of artists in this delicate matter are rather more ‘interesting’; certainly Pygmalion was not entirely devoid of esthetic feeling.” Art’s great value, for Nietzsche, is in its service to life, not in the sense of menial practical utility but as providing the heights of beauty, meaning, and pleasure that justifies existence. Though it relies on appearance, true art celebrates through its “esthetic delight” the principle of “eternal life . . . beyond all appearance and in spite of destruction”. It is through art that “this world can be justified only as an esthetic phenomenon” (pp. 55–58, 101–102, 143). Contesting Schopenhauer’s view that art reveals truth in the form of Platonic Ideas, Nietzsche argues that art provides not only beautiful life-serving illusions – the Apollonian dream-world of clear and perfect forms – but also a penetrating glimpse into a deeper Dionysian reality of frenzied will and flux that defies our principles of order and individuation. Art, “that sorceress expert in healing”, enables us to face and recover from such terrifying visions: “the spirit of the sublime . . . subjugates terror by means of art” (p. 52). But art also offers escape from the distressing or hideous truth. “Truth is ugly”, Nietzsche concludes, “We possess art lest we perish of the truth” [Nietzsche (1901/1968, p. 822)].

Hegelian philosophy was radically reinterpreted by Karl Marx (1818–1883). Marx adopts the notion of an inevitable, scientifically provable progress of society, but he replaces historical spirit with matter, or rather man’s relation to matter, as the driving force. Thus he blends Hegelian historicity with the mechanical inevitability of British political economy. In consequence, all “institutions” of human culture, including the arts, are determined by a society’s production relations. Production determines the creation of value and it shapes the intellectual superstructure to which the Arts belong. New production relations will lead to new forms of art. Therefore, a separate recognition of artistic value is not necessary. Yet, Marx’ influence on continental aesthetic philosophers, for instance, Adorno and Benjamin, was considerable.

#### 4. Art in economic theory until the 1970s

Vilfredo Pareto (1848–1923), Walras’ successor in Lausanne, made a number of contributions that helped to give greater analytical precision to economic models. In *Manuel d’Economie politique* (1909), Pareto pitches the obstacles of production against the tastes (goûts) of consumption [Pareto (1909)]. Taste, in this interpretation, is not a faculty to be developed and improved. It is any kind of predilection the user of a commodity might have and which he or she is able to rank in their order of preference. The notion of preference proved helpful for the development of English economic theory which took a formal analytical turn after the First World War.

London, the center of global finance, had become merely another European capital. The shock led to a boom in philosophical discourse. Epistemological positions were radicalized, and these positions were soon to characterize the most influential works in economic science. Intellectual circles included both economists and philosophers. The



philosophical tenor of all these circles was one of a complete separation from metaphysics. Philosophy turned towards itself – towards the observation of the way in which observing statements are made. In London, intellectual discourse liberated itself from the dominance of the “Cambridge Circles” with their strongly literary and philosophical flavor.<sup>10</sup> The circle of economists formed by Lionel Robbins (1898–2004) at the London School of Economics was, compared to Cambridge, less bound by tradition, more cosmopolitan in outlook, and more interested in applying the newly found laws of logic to one’s own methods. Such were the conditions under which John Hicks (1904–1989) presented *Value and Capital* (1939), a model of the economy where consumption value is the undisputed driver of economic action.

Following Pareto, Hicks makes tastes and preference orderings part of the formal foundation upon which the edifice of general equilibrium theory is erected. Along ordinal scales of preferences, the points of consumption equilibrium between various commodities at given budgets are determined. Without having to measure utility in cardinal units, the “value equilibrium” of every individual can be measured in money units “with respect to a system of market prices” [Hicks (1939, p. 20)], and that determination is deemed fully sufficient.

It was a small step from Hicks’ version of value determination to the version published by Gérard Debreu (1921–2004) [Debreu (1959)]. Debreu calls his contribution *Theory of Value*. In his model the need for the term is eliminated. “Value” is used synonymously with “market price times commodity volume”. Use or consumption value is still assumed to drive the economy. But it has been reduced to the subjective and impermeable preferences/tastes of individual agents. “Value . . . became whatever the individual globule of desire made it out to be, a gravitational attraction ‘sui generis’, and therefore was not something over which one should have a rational dispute . . .” [Mirowski (1989, p. 25)]. A neat division of research is thus established: economics deals with the properties of interdependent markets, while the formation of taste for art, as any other process leading to a change of preferences, is the business of other disciplines, like psychology, art history or aesthetic philosophy.

When the stability of tastes or preferences as a basis for economic modeling came under attack, George Stigler and Gary Becker developed a variation of the human capital approach, with explicit reference to artistic appreciation. In a paper titled *De gustibus non est Disputandum* (1977), Stigler and Becker do recognize that there are cases where additional exposure leads to a growth in consumption of the particular commodity. Heroin consumption is cited as an example for harmful, music consumption as an example for beneficial “addictive” effects. Rather than to postulate a change in taste, the authors suggest that the effect can be more fruitfully explained through changes in the shadow prices which govern the household production function: in order to consume

<sup>10</sup> An example is the Conversation Society of the “Apostles”, whose membership included Bertrand Russell, J.M. Keynes, E.M. Forster and Lytton Strachey. The latter were also members of the “Cranium Club”, founded in 1924, which was one of the London offshoots of pre-war Bloomsbury. See Skidelski (1992, p. 13).

music, not only market goods are needed, but also time and “music capital”. Accumulated knowledge and skill reduce the cost of future consumption and thus account for the observed increase in marginal utility.

## 5. Art in aesthetics until the 1990s

Aesthetic theory in the twentieth century displays considerable diversity, including some skepticism with respect to the possibility of universal theories of aesthetics because of the worry that “art” named a historically constructed concept of rather ambiguous, contested, and shifting boundaries rather than signifying a natural kind with a common essence. This skepticism extends to criteria for artistic valuation. At the turn of the century already we find such anti-essentialist positions.

Leo Tolstoy (1828–1910) was a vehement advocate for art’s communicative value. He defined art as the communication or contagion of feelings. “Art is a human activity, consisting in this, that one man consciously by means of external signs, hands on to others feelings he has lived through, and that others are infected by these feelings and also experience them” [Tolstoy (1896/1997, p. 681)]. From this definition of art, he derived evaluative criteria of two kinds. First in terms of efficacy of communication, “The stronger the infection the better is the art, as art, speaking of it now apart from its subject matter – that is not considering the value of the feelings it transmits”. The degree of the infectiousness depends on three conditions: “the individuality of the feeling transmitted”, the “clearness” of transmission, and “the sincerity of the artist” understood in terms of “the force with which the artist himself feels the emotion he transmits” (1997, p. 685). The condition of sincerity, Tolstoy argued, is the most important and in fact includes the others. With respect to subject matter, Tolstoy adopted a religiously Christian and democratic perspective, arguing that good art should express feelings that “unite all men” and construing this as comprising only two kinds of feeling: “first, feelings flowing from a perception of our sonship to God and of the brotherhood of man; and next, the simple feelings of common life accessible to everyone without exception, such as feelings of merriment, of pity, of cheerfulness, of tranquility, and so forth” (p. 689). One obvious (and awkward) consequence of Tolstoy’s theory is his devaluation of partisan, patriotic, difficult, or elitist art, which includes a shocking condemnation of Beethoven’s *Ninth Symphony*.

Benedetto Croce (1860–1952) offered a theory of art as intuition-expression that contained some striking views on evaluation. A true or successful artistic intuition, he argued, implied its active expression, though such intuition-expression did not need to be externalized in a physical object. Technical skill was thus excluded from artistic value. Moreover, as every intuition-expression was a unique product, there were no degrees of beauty or positive artistic value. Artistic value, for Croce, means adequate intuition-expression of its content, and if a work adequately intuits-expresses this, then nothing could be more expressive or beautiful. “The beautiful does not possess degrees, for there is no conceiving a more beautiful, that is, an expressive that is more expressive,

an adequate that is more than adequate. Ugliness [as unsuccessful expression], on the other hand, does possess degrees” [Croce (1901/1970, p. 79)]. Croce’s insistence on the uniqueness of each artwork as intuition-expression entailed rejecting all general principles of art evaluation such as those based on definitions of genres; he regarded genres as arbitrary conventions or illusions.

Most of continental, particularly German philosophy of art in the twentieth century reflected the enduring influence of Hegel by being very critical of identifying art’s value in the intrinsic pleasures of immediate experience of beauty, emphasizing instead the more than aesthetic ideals of truth and understanding. Heidegger (1889–1976), for instance, affirms art as “a distinctive way in which truth comes into being” (1975, p. 78). Hans-Georg Gadamer (1900–2002) similarly stresses the cognitive dimension in claiming that the pleasure of art’s play is “the joy of knowledge” [Gadamer (1982)]. Theodor Adorno (1903–1969) shares this philosophical bias for truth over beauty as art’s most essential value, arguing that art’s production of beauty, in our modern world that has witnessed such horrors as Auschwitz, seems too deceitfully affirmative of the world. Adorno thus can explain the way modernist art has eschewed the simple goal of beauty and instead pursued other expressive ends: “Great works of art are unable to lie” [Adorno (1973/1984, p. 188)]. In the contest of artistic values, Adorno clearly affirms that pleasures of beauty must be sacrificed to truth. “In a false world, all hedone is false. This goes for artistic pleasure too . . . In short, the very idea that enjoyment is of the essence of art needs to be thrown overboard . . . What works of art really demand from us is knowledge or, better, a cognitive faculty of judging justly” (pp. 18–21). Moreover, though Adorno recognizes that art has always had social and practical uses, he rejects the idea of understanding art’s value in terms of functionality. Instead he paradoxically maintains that “if any social function can be ascribed to art at all, it is the function to have no function” and thus offer an alternative to the “ungodly reality” of ordinary practical existence and utilitarian thinking (p. 322). Hannah Arendt (1906–1975) similarly argues that, in contrast to ordinary commodities, the value of artworks “is the very opposite of functionality” or use in the consumptive process of life. Artworks are pure ends, things of “intrinsic, independent worth”, “things which exist independently of all utilitarian and functional references, and whose quality remains always the same” and thus displays the value of “imperishability” [Arendt (1961, pp. 208–218)].

Walter Benjamin (1892–1940), who was closely associated with Adorno and much admired by Arendt, offered a more nuanced, balanced view of art’s value, by recognizing its functionality and distinguishing between varieties of its use value (1969). Benjamin’s key distinction here is the opposition between the artwork’s cult value and its exhibition value. The former is connected with art’s auratic quality, its prehistory in magic, its use in ritual, its sense of authentic uniqueness and its special connection with genius and the distant past. Part of the value here is the esoteric value connected with the artwork being not easily accessible or readily and widely seen. In contrast, exhibition value concerns the value obtained from the perceptual experience or enjoyable consumption of art. Though loss of cult aura has in some way diminished art’s power,

Benjamin recognized the valuable democratic potential of art's move toward exhibition value as some compensation for such loss.

By the turn of the century, Cambridge had become a center not only for general analytical philosophy, but for aesthetic philosophy as well. A dominating figure was George Edward Moore (1873–1958). He claimed that beautiful artworks had objective, indeed intrinsic value, but that such value could never be captured by a definition or criteria based on natural properties [Moore (1959)]. The ethically good and the aesthetically beautiful, he argued, were non-natural values that could not be analyzed in terms of criteria or standards. The aesthetic value of each artwork must be judged, as Kant and Croce had earlier argued, through a particular judgment or intuition, rather than being derivable from a general definition or principle. And to intuit the artwork properly, one had to consider it in terms of its organic unities while isolating it from its external uses. Moore's theory was not purely formalist since he thought that the truth of an artwork added to its value. But his emphasis on organic unities helped inspire the more distinctively formalist theories that Roger Fry (1866–1934) and Clive Bell (1881–1964) applied to the plastic arts, since these art theorists, as well as the J.M. Keynes and other Bloomsbury intellectuals, were avowedly influenced by Moore's philosophy.

Ludwig Wittgenstein (1889–1951), a Viennese philosopher at Cambridge who at times worked closely with Moore but soon succeeded him in philosophical influence, introduced even more particularity and variability to the evaluation of artworks. Wittgenstein argued that the concepts of aesthetics, such as art and beauty, were especially vague and ambiguous. There was no single essence of art or beauty on which we could ground our value judgments, but these concepts did not require essences for us to use them validly in evaluations. Opposing Moore's concentration on beauty and his assumption that it had a common essence (albeit one that could not be reduced to natural properties or captured by definitional criteria), Wittgenstein (1970) argued that our aesthetic evaluations were of significantly different kinds that could not be reduced to a single form. In his terminology, there are a number of different language games with respect to art: evaluating a sonnet as properly formed or a performance as technically flawless virtuosity is different from judging a portrait as luminously subtle or a novel as deep or great. Outside the plastic arts and music, he pointed out, the predicate "beauty" is not frequently used in our aesthetically evaluative language games, and these language games are practiced with somewhat flexible rules. Moreover, Wittgenstein noted that some of aesthetic evaluations are expressed as much in our behavior as in our linguistic statements. Despite such fuzziness and openness, Wittgenstein recognized that there must be significant convergence in our aesthetic evaluations since they are embedded in shared ways of life that give our evaluative terms their meaning. Though we require no essence of the beautiful for the term "beautiful" to have an understandable and shared meaning, we do need some convergence on its use and applications.

Moore and Wittgenstein were founders of analytic philosophy that largely dominated Anglo-American philosophy in the twentieth century. Pragmatism, however, has also been an influential philosophy in the United States and increasingly elsewhere. John Dewey (1859–1952), its major exponent, offered a comprehensive aesthetic theory in

*Art as Experience* [Dewey (1934/1987)]. Rejecting the Kantian ideas of disinterestedness and purposelessness, Dewey argued for art's wide-ranging functionality. Not only do artworks serve a variety of instrumental functions (entertainment, edification, religious inspiration, decoration, personal and social expression, etc.), but art is also enjoyed intrinsically for the sake of the "consummatory" experience that it provides, an experienced fulfillment that is valued for its own sake but that also, through its vividness and vitality, functions to enhance life in a general way by making it more satisfying and by stimulating the energy and intelligence of the individuals and groups who participate in art's aesthetic experience. Art's value, Dewey argued, is not in artworks as mere physical objects but in the lived experience that those objects serve, whether this is the creative experience of the artist or the appreciative experience of the audience. Because art provides enhanced experience that can be powerfully and widely shared, it has an important function of forming and expressing community, constituting "a remaking of the experience of the community in the direction of greater order and unity" (1987, p. 87). Art's imaginative power, moreover, can improve our ethical sensibility. Dewey makes the important point that instrumentality is not inconsistent with intrinsic value when the latter is construed as valuing something for its own sake rather than 'only' for its instrumental uses. We can enjoy art's aesthetic experience for its own sake while also appreciating its non-aesthetic functions.

While analytic philosophers often shied away from both the issue of aesthetic evaluation and the concept of aesthetic experience, fearing that both were too problematically subjective, Monroe C. Beardsley (1915–1985) developed Dewey's idea of aesthetic experience into a distinctive theory of evaluation [Beardsley (1958)]. The value of an artwork is defined in terms of its ability to create an aesthetic experience of a certain magnitude, such experience being presumed to have value; the larger the experiential magnitude, the higher the value. Beardsley combined this approach with three largely formalist criteria of value: the artwork should display unity, complexity, and intensity – these attributes being conducive to the production of strong aesthetic experiences. Suspicious of the notion of intrinsic value and insufficiently attentive to Dewey's harmonizing of intrinsic and instrumental value, Beardsley held that aesthetic experience was valuable, but not intrinsically; its value rather derived from the valuable role or consequences that aesthetic experiences had in the life of individuals and of society as a whole. Nonetheless, Beardsley conceived aesthetic experience as essentially compartmentalized from ordinary life, so that artworks should be interpreted in terms of their immediate perception and not in terms of their wide-ranging referential relations to real world contexts. Such views made Beardsley an important theorist for the New Criticism.

In sharp contrast to Beardsley, Nelson Goodman (1906–1998) insisted on art's referential or symbolic functioning. Goodman (1969) criticizes traditional aesthetic theory for having devoted too much attention to questions of artistic value, which he argued has obscured our understanding of art's modes of meaning and thus diminished the value of aesthetic theory itself. However, based on his analysis of art as symbol, Goodman boldly proposed "the subsumption of aesthetic under cognitive excellence" (1969,

p. 259) a remote echo of the familiar strategy of defining art's value in terms of truth or knowledge.

George Dickie, who skeptically criticized Beardsley's notion of aesthetic experience as a metaphysical phantom (1965), later affirmed experience as essential to explaining art's value: art is valuable as instrumental to the production of "the experience of aesthetic qualities" (for example, unity, complexity, intensity) such experience having intrinsic value (1997, p. 158). Recognizing that not all artworks display the same sort of valuable properties, Dickie argued that we cannot provide a single universal matrix for ranking the values of all different artworks, though we can compare artworks in terms of how many valuable aesthetic properties they display and the degree to which they display them.

Pierre Bourdieu (1930–2002) has contributed works as a philosopher and as an economic sociologist. He argues (1996) that art's value is socially and historically constructed and is thus ultimately grounded in the social structures of cultural hierarchy and hegemony. Artworks that have been admired for centuries become established as icons of culture and genius whose worth cannot be easily contested because they are so deeply socio-culturally entrenched. Not merely items possessing symbolic capital, artworks are also markers of cultured taste that bestow symbolic capital on those classes and individuals who know the right artworks to appreciate and the right ways to appreciate them. Though taste may seem spontaneous and individual, Bourdieu maintains it is the product of pervasive, extended (though typically informal and implicit) social training. Art's value is objective but only as a social fact not as an independent ontological given.

Niklas Luhmann (1927–1998) placed the social construction of artistic value in the context of an encompassing theory of society. In his view [Luhmann (1995/2000)], art theory's connection with philosophy has encumbered it with constraints in theory design which do not have their origin in the Arts themselves. A theory adequate to the unique qualities of Art must account for the simultaneous presence of surprise and recognition which characterizes the experience of works of Art. The special reality communicated by such events is an intentionally fabricated duplication of common reality, a performance of "a world within a world" (p. 241). The theorist and other observers are at liberty to apply the communicative power of Art in an idealizing, critical, affirmative or exploratory manner. But all liberties, as well as all constraints, are the result (Eigenprodukt) of decisions which have been taken within the works themselves.

## 6. Varieties of artistic external effects in contemporary economics

Since the 1970s, economic theory has expanded to explanations of institutional change. Rights to natural and intellectual property are interpreted as part of a social contract that underlies economic action. Implicit constraints on individual behavior can be explained as the stable solutions of non-cooperative strategic games: it pays off to respect the rights

accorded to others. Apart from the rights, rules are reached that reduce the damage done by negative external effects and capture a share of the possible positive external effects.

The concept of “external economies”, originally suggested by Alfred Marshall, has experienced a remarkable rise to prominence in the past thirty years. Marshall had employed it to identify deviations from constant returns to scale in production within an industry. Examples for positive external economies are education and skills, examples for external diseconomies are environmental pollution and political instability. The price system does not reflect the positive or negative “value” of such activities. The lessons of this discovery have been applied to the economics of developing countries, to the economics of education and research, and to environmental economics. They have also been applied to the role of the Arts.

The following section presents three examples of “external effects” which came into view when markets for cultural goods were investigated. The three cases are not exhaustive. They are selected because they bring the valuation of art into play.

Tibor Scitovsky (1910–2002) had participated prominently in the discussion of technological external effects in the 1950’s. In *The Joyless Economy* (1976), his scope is larger: he investigates the “value of consumption skills”. He starts from the notion that subjective value is pure mental value. In consequence, psychological research should yield insights into the process of attaining pleasure. Reviewing the literature, he arrives at a distinction between pleasures of arousal and pleasures of stimulation. The former evoke comfort, the latter evoke a sensation of novelty and discovery. Individuals choose the two varieties in changing proportion, a fact that can be measured through sociological statistics. The data show that the “growth rate” of stimulation pleasure far exceeds that of comfort pleasure. Stimulation is often supported by purchased goods and services. In many cases, it is attained in non-market contexts. Scitovsky discusses self-stimulation, mutual stimulation, non-market goods and external economies. He concludes that the welfare of the community is significantly determined by “the economy’s ability to produce the economic product with a maximum of beneficial and minimum of harmful accompanying effects” (p. 105).

Cultural experiences are singled out as activities leading to beneficial side effects. Culture is defined as “that part of knowledge which provides the redundancy needed to render stimulation enjoyable” (p. 226). Consumption skills are the means, which turn further stimulation into enjoyment, into the source of subjective value. Consumption skills rely on personal practice, on acquired taste and on critical judgment. They are, at base, information differentiation skills: out of the constant flow of new information, a very few items are selected to become the source of common enjoyment in a community or civil society.

Amongst the cultural consumption skills, the skills necessary to exercise and understand the valuation of art objects and performances are central. Scitovsky uses the example of music composition and of painting to illustrate the skill of assembling new works with sufficient degree of redundancy and novelty (pp. 48 and ff.). They are practiced by artists, by experts and by amateurs.



Just as production skills are observed to increase yield per input, consumption skills are able to increase “consumptivity”, or consumptive yield. Scitovsky observes that in the course of the past 300 years of industrial development, production skills have progressively crowded out consumption skills. He suggests several reasons for that phenomenon. Firstly, the value of consumption skills is underestimated for reasons that echo the moral principles of Hutcheson’s age: skills for enjoying leisure are frowned upon in societies which distinguish between activities elevating the soul to bliss, and activities endangering such elevation by stimulating selfish pleasure (p. 228). Secondly, measures of increases in earnings are more precise than measures of the value of skills for the enjoyment of concerts or ballets. Such skills might open “a large reservoir of novelty and years of enjoyment” (p. 235), but the results are uncertain and therefore discounted heavily. Thirdly, the exercise of consumption skills is seen as a threat to currently produced goods: the skills provide access to pleasures of self-discovery and novelty. The pleasure can even be generated out of the process of the activity itself. In consequence, certain cultural commodities and service become superfluous. Once a person has learned to write, for instance, she does not need the services of a scribe or a reader anymore. It is overlooked that new commodities which were beyond the horizon of the old preference ordering begin to be valued: a person who knows how to read develops a taste for books and newspapers.

William Baumol (1986) presented a ground-breaking study on the rates of return from the resale of paintings since 1652. According to his results, returns follow a random pattern. The average return lies at 0.55%, well below average return rates of alternative assets. Since then, an entire subdiscipline has set out to test and expand the Baumol results.<sup>11</sup> The majority of studies confirm the result of below average returns, while the random walk result is not confirmed. The explanations for the “anomaly” point consistently to “psychic returns”, and to the “social benefits” of signals that indicate a specific status or position in society. Psychic pleasure can be triggered by the subject matter of a painting (arousal) or by knowledge of a painting’s history and its position in the world of taste (stimulation). Social benefits flow from being able to signal one’s income level, cultural erudition or attitude to novelty and risk. The value for the owners of artworks lies in the improvement of their negotiating position, on the assumption that negotiating partners are able to read the signals which are brought into the medium of a private collection or a national museum. Expensive materials are easy to read, while the subtleties of a particular style, like black-ground Greek vases or “Arte povera” objects, demand the operation of elaborate standards of aesthetic discrimination. Thus, the successful operation of standards of critical taste exerts positive external effects on the consumption value of an individual.

Furthermore, standards of taste make art objects suitable for value storage and potential value increase. The decisions of investors follow the quality judgments of professional experts. The stability of these judgments has made it possible that art values

<sup>11</sup> See Ginsburgh, Mei and Moses, Chapter 27 in this volume.



maintain their resale value for longer periods of time than bonds and obligations, and that the risk of “bets” on the success of contemporary works can be significantly reduced by following the advice of gallerists and critics who are skilled in aesthetic valuation.

One of the methods developed in the extensive discussion on the value of portions of the natural environment is the “Contingent Valuation Method” [Noonan (2003)]. Environmental benefits are sometimes enjoyed by individuals, but markets for generating such effects cannot be installed. The standard institutional response to this problem is a transfer to the political agenda: taxes or transfers are executed through public households, rules and regulations are put into effect. It is difficult for political agendas to reflect individual preferences adequately. Contingent Valuation is a survey technique that helps to solve the problem by gathering information about exchange values in imaginary markets. The respondents are asked to reveal their willingness to pay for the source of non-market beneficial effects – the survival of an animal species, the designation of a nature preserve and, by extension to the cultural field, the preservation of a historical monument.

Responses to survey questions are radically different from the actual reduction of buying power in a market transaction. Still, the results consistently indicate a positive willingness to pay for artistic works and institutions, and for monuments of cultural heritage. Particularly noteworthy is the observation of option values. Cultural option values are opportunity values attached to places, objects and events which are known to the individual, but not experienced by the individual. The individual “consumes” the ability to express the identity, the history and the ideas and aspirations of a “culture” [Throsby (2001)]. A culture may have the size of a town, a region, a nation or even a network connected by a common belief or life-style. “Cultural value” is a club good for the members of such communities. They benefit from being informed about their common heritage. Cultural symbols provide common themes for conversation and they facilitate the formation of expectations about negotiation partners. For non-members, artworks inform about fundamental issues and perceptions that characterize that other culture. Beyond this instrumental dimension, members reap immediate intrinsic benefits because they take pride and pleasure in artifacts and events that are representative of their community and its “canon” of excellence.

In consequence, the particular method of valuation of art matters for the economic outcome: only autonomous processes of aesthetic valuation can credibly select those artifacts, stories and compositions which are to be regarded as the height of taste. In that process, the set of currently canonical works is formed and constantly contested. It is this set of selected works and performances which serves as a source of value to individuals within and outside of the cultural boundaries of that canon.

## 7. Varieties of artistic value in contemporary aesthetics, and their economic valuation

In the vague concept of artistic value, different kinds of value seem to be nested. We distinguish ten kinds which have been to some extent suggested in the aesthetic theories so far sketched. They are conceived as heuristic distinctions to illuminate the concept of artistic value rather than as ontologically distinct categories of value. The forms of artistic value do not include art's economic value as understood in monetary terms. This form is discussed, as a distinct category, below. It is hard to imagine how all these types of value could be organized into one accepted calculus for ranking the value of all works of art, not least because the relative weighting of these different types would be much contested. Yet, to distinguish them could enable us to be more precise about what we are in fact valuing when, with respect to particular artworks, we speak of artistic value.

(1) Art's *moral or religious vision*, its power to edify and spiritually uplift, can still form part of a work's artistic value, while the appeal to low human drives and the toleration of morally condemned behavior diminishes the value of an artwork. Form cannot be adequately isolated from content. The moral or religious vision expressed in a work forms part of the work's content and structure, and as such its valuation can be legitimately included in our appraisal of the work's value. This is not to say that the moral vision must be true or fully acceptable to the appraiser, but it must at least be regarded as reasonable, mature, and coherent. Art has traditionally been valued for its religious uses, as in altar pieces, poetry and music of prayer. Positive moral effects have been ascribed to literature, to musical education and to visual works, while others have been condemned for irreligious and moral corruption. Art can improve our character by its harmonizing of our psyche, as Aristotle and, in more individualistic terms, Schiller have suggested through their ideas of catharsis and play. The education of moral sensibility through artworks that portray fine subtleties of ethical behavior and character expression has been considered a source of value since Shaftesbury.

(2) Art has long been valued for its deep *expressiveness*. Expression, it is argued, requires a medium through which the self can be expressed, and the various media of art, rich with perceptual and semantic potential, provide a superb matrix for such expression. Advocates of expression theories of art, such as Croce (1970) and Collingwood (1958), argue that the artist begins with an unclear feeling or sense of what she wishes to express, and it is only through art that the expression acquires clarity and distinction. Apart from this transitive sense of expression, where an artwork's expression is the expression of something anterior – a specific emotion, idea, etc. – there is an intransitive sense of artistic expressiveness that is valued. It makes sense to say of a painting or a piece of music that it is expressive without our being able to specify what exactly it expresses. Here expressiveness connotes the degree of power and impact which is suggestive of artistic value.

(3) Art's *communicative power* for the sharing of feelings and ideas between artists and their public is part of artistic value. Art's emotional quality, direct experiential appeal, and link to pleasure give it a penetrating, pervasive infectiousness that promotes

easy, rapid, powerful, and widespread communication. Kant located the grounds of aesthetic judgment in the “*sensus communis*” of human nature, Schiller argued that “only the aesthetic mode of communication unites society because it relates that which is common to all” [Schiller (1986, p. 217)]. In Dewey’s words: “Art breaks through barriers that divide human beings, which are impermeable in ordinary association” [Dewey (1987, p. 249)].

(4) Communicative power is also essential to art’s *social and political* value. Artworks typically embody the meanings and ideals of the society in which they are created; even works that have a revolutionary message must rely to some extent on shared meanings and values or else they would be unintelligible and totally rejected. Art thus provides an attractive repository of ideas and ideals that build social unity and stability, while enabling their transmission over generations. Through its imaginative dimension, works of art can also inspire new visions of social and political order. The social and political import of an artwork cannot be neatly isolated from its directly experiential artistic value. Our aesthetic experience of listening to an anti-war protest song from the sixties or a politically charged rap song from the eighties derives added enjoyment and meaning from recognizing the socio-political motives and roles such artworks have played.

(5) Plato’s condemnation of art as a deceptive purveyor of falsehood has been frequently countered by affirming art’s *cognitive* value. Even if we dismiss the notion of a special form of truth that is accessible only through artistic means, art has undeniable value in effectively communicating a wide variety of truths and in honing our symbolic skills of conveying and processing very subtle forms of information. Because emotion has a strong bodily dimension, art’s emotional power makes the truths it expresses more powerful and convincing, because as emotionally grasped truths they become more deeply embodied and impressed in our consciousness and memory. The very appreciation of form and meaning is an exercise whose practice enhances our cognitive skills and our proficiency in symbolic processing.

(6) Many theorists, as do most other people, locate art’s value largely in the special, directly satisfying or pleasurable experience it gives. We call this art’s *experiential* value. It includes art’s entertainment value – the entertaining pleasure and distraction it provides as a pastime. But art also has experiential rewards that are not primarily pleasurable. Avant-garde works, for example, may produce experiences of shock, intensity or outrage that we recognize as valuable without their being pleasant or enjoyable. The central role of aesthetic experience in art’s value has been reaffirmed in recent years in rather different ways. Budd (1995) insists, contrary to Beardsley, that such experience has intrinsic value in some meaningful sense. Therefore, Budd argues, art’s value as art should be confined to purely aesthetic dimensions of art’s immediate intrinsic experience. Shusterman (1992, 2000) allows that art’s value also should include the instrumental effects or consequences of an artwork’s aesthetic experience, including the truths, insights, and uses the work provides.

(7) Aesthetics has long emphasized certain formal or design values embodied in art: unity, harmony, complexity, balance, intensity, dramatic tension, etc. Such formal values

are sometimes distinguished by philosophers as distinctively *aesthetic* values in contrast to artistic values. This is because these formal values clearly seem applicable to objects other than artworks (a flower or sunset or ocean storm) and do not seem to require historical knowledge of art in the ways demanded by assessments of art-historical value, art-technical value, or cult value. Nor do these values demand for their appreciation the sort of external, non-aesthetic knowledge we need for assessing the cognitive, moral, religious, or communicative value of artworks. Experiential value might be considered along with formal values in this group of distinctively aesthetic values of art since the dimensions of experienced value do not make essential reference to art-historical knowledge or standards outside the immediate experience of the artwork.

Expressiveness, in the intransitive sense of evocative suggestiveness, can also be included under specifically aesthetic properties of artworks. We can appreciate an artwork as expressive without external art-historical knowledge about what its creator wanted to express and without even assuming that there was a distinct idea the work aims to express. Indeed, we can even speak of a natural scene – such as a rock formation or a gnarled tree – as aesthetically expressive without implying that it expresses a prior identifiably intention by its creator. A clear distinction between artistic and aesthetic value has been hard, however, to maintain, because the term “aesthetic” is so commonly associated with the artistic and because it can be argued that even judgments of unity, intensity, complexity, etc. in artworks implicitly rely on some basic art-historical knowledge.

(8) A specific kind of artistic value could be called *art-technical* value. Such value relates to the skill, technique, or technical innovation displayed by an artwork. We can, for example, regard the content or form of an artwork as not particularly worthy of appreciation but still value the virtuosity of technique or invention that the work or its performance displays.

(9) *Art-historical* value concerns the value an artwork has for art’s history, either by its providing evidence of historical innovation or influence, whether technical, stylistic, or in terms of new content, or by simply being a crucial historical artifact for art history. Though some viewers find Picasso’s *Desmoiselles d’Avignon* a very unattractive painting, its artistic value in terms of art-historical value (as the harbinger of cubism) cannot be denied. Physical rarity, because very few other surviving exemplars of its period or style have been found, adds to appreciation.

(10) Related to art-historical value is *artistic cult* value. Through a history of appreciation and dissemination, a particular artwork, for example, Leonardo da Vinci’s *La Gioconda*, becomes identified as a hallowed locus of artistic genius and a paradigm of self-representation. The strength of the aura, to which Benjamin refers, gives value to the reproduced versions of the image, and the volume of reproductions, in turn, increases the cult value of the original.

Some of the works of art created also have economic value. Economic value is a property which all works can attain, irrespective of the kind of artistic value attributed to a particular work. Money is paid in exchange for original works, copies of originals (books, prints and disks), and performances of musical or theatrical scores. Certain

patterns of demand and supply are directly connected to some of the artistic values sketched above.

The ability of public expression has made artworks a valuable resource for the institutions of power and their public households. When religious institutions controlled a large portion of social power, their demand for works which could transport their messages and rules was high (1). It has dwindled since, to be replaced by political institutions (4), educational institutions (3) and by a recognition that the expressive power of art makes it a legitimate candidate for public financial support (2).

Art's cognitive value (5) communicates a "wide variety of truths" and it trains "symbolic skills". An economically relevant truth is social status, transported with symbolic skills. Individuals and communities know how to represent their status through the purchase of art works, through buying the right to attendance and through the acquisition of decorating items that are associable with artworks. High market values signal high social status, and therefore the prices of certain works may be bid up without any change in the artistic value of the works. While cognition is directed outward, pleasurable experiences (6) are directed inward. The desire to make time pass pleasantly is a powerful motive for spending income. The share of entertainment goods and services in total expenditure increases over time, as more immediate pleasures of "arousal" are saturated, to use Scitovsky's term. Since stimulation produces constantly new variations, works of art become a valuable source of inspiration for authors, and they attain, due to the media attention, widespread popularity among consumers which steps up demand for originals and copies.

Artworks achieve part of their impact because of the pleasure gained in experiencing virtuosity. Artworks are judged by the degree to which they satisfy formal qualities (7). The technical challenges contained in the execution of an artwork are mastered by only a few (8). Excellence implies scarcity. Kinds of works which are in shorter supply can command higher prices. Rarity is also a key factor in art-historical value (9). The economic value of a historical item, like a book from the first edition of a successful novel, also increases with a decrease in specimens available. Finally, the paradigmatic uniqueness of works that have been attributed "cult value" (10) increases the demand for copies, for travels to the location of the cult work or cult event and for other works associated with an "icon".

Economic value increases when several varieties of artistic value are combined in a single work. The highest prices are attained for rare works or performances by most highly ranked masters that combine emotional impact with status and entertainment value. In contrast to ordinary consumption goods, the value of physical artworks is sustained or even increased over time, making such works effective stores of economic value.

## 8. The logic of aesthetic and economic evaluation

Philosophical theories concerning the logic of evaluating art are focused on three related topics: the logical status of evaluative judgments, the role of reasons in evaluation, and the general structure of evaluative argument [see Shusterman (1980, 1981)].

(1) The *logical status of evaluative judgments* concerns the question of whether they are propositions that have descriptive truth value, are prescriptive expressions of feeling or of recommended decisions to take a particular attitude to the artwork or, finally, are neither descriptions nor prescriptions but rather performative verdicts that themselves establish the evaluation they make. These different options are respectively advanced by descriptivist, prescriptivist, and performativist theories of evaluation. Descriptivist theories can be divided further into subjectivist, relativist, or absolutist theories according to whether the alleged descriptively true statement of value refers only to a particular subject, or is relative only to particular criteria of evaluation, or is meant to have absolute or universal reference. The subjectivist descriptivist construes the assertion that “artwork W is valuable” as meaning “artwork W is valuable to me”. Literary masters such as D.H. Lawrence and Walter Pater argued for this position, Lawrence (1936, p. 539) claiming that “literary criticism can be no more than a reasoned account of the feeling produced in the critic by the book he is criticizing.” In contrast, absolutist (or objectivist) descriptivists – such as the renowned critics Matthew Arnold, the early T.S. Eliot, and Yvor Winters – claim that evaluative judgment is about the artwork itself “as it really is”. Though the subjectivist position may seem overly personal and impressionistic, Pater (1912, p. 10) shrewdly defends such evaluation by arguing that “in aesthetic criticism, the first step to seeing the object as it really is, is to know one’s own impression as it really is”. The relativist occupies an intermediate position of evaluative descriptivism. He admits that some evaluations are better than others and that some are plainly wrong. But recognizing that there are often conflicting evaluations of an artwork that seem well-reasoned and adequate, he rejects the absolutist assumption that there must be only one true evaluation and that all others must be regarded as false. The relativist therefore construes evaluative judgments as true or objective relative to some standard that is deemed adequate. Among the many theorists and critics who advocate forms of relativism, E.D. Hirsch (1969, p. 33) clearly formulates the key idea that though “there is no privilege [of a particular evaluative standard] in literary evaluation, there is nevertheless objectivity and accuracy, and these reside entirely in the judged relationship between the literature and the criteria we choose to apply to it”.

Other theorists construe evaluative statements not as descriptive propositions at all. For an emotivist–prescriptivist, such as A.J. Ayer (1971), critical judgments “are pure expressions of feeling and as such do not come under the category of truth and falsehood” but are instead meant to make us share the feelings and attitude expressed in the evaluation. To lack truth status does not entail lacking efficacy in communicating a work’s felt value and convincing others of it. An alternative position, performativism builds on J.L. Austin’s theory of performative utterances. It argues that evaluative judgments do not describe the work’s value in itself or for the subject who judges, but instead

display or create the work's value. As Margaret MacDonald (1954, pp. 121–122) formulates it, “to affirm a work good is more like bestowing a medal than naming any feature of it or of the status of its creators or audience. Verdicts and awards are not true or false. They may be reversed but not disproved. But they can be justified and unjustified”. To give some common examples, when an important book reviewer writes in his review that a book is highly recommended, he institutionally renders it such, thus augmenting the value it is critically accorded. The same goes for the nominations and awarding of artistic prizes.

With respect to these different theories of evaluative judgment, one is led to conclude that none is entirely acceptable because each has more than a grain of truth. Evaluative judgments are in fact diverse in form and function; descriptivism, prescriptivism, and performativism can all find some evaluative judgments that support their different analyses.

(2) Critics and art lovers do not simply offer evaluative judgments, they give reasons for them. What role do these *evaluative reasons* play? Some theorists argue that evaluative reasons play a logical role of evidence or validating principles. Monroe Beardsley, for example, views the reason “this work is unified” as providing some evidential confirmation that the given artwork has value because it provides inductive evidence that the work is likely to produce an aesthetic experience (which has value). Beardsley argues that the logical relation of unity to aesthetic value is similar to that between “The food is dangerous” and “It is crawling with salmonella bacteria”. Reasons relating to complexity and intensity are likewise evidence, though not necessary or sufficient conditions, for affirmations of aesthetic value, since objects possessing those qualities tend to produce aesthetic experiences [Beardsley (1958, pp. 533, 535)]. Similarly, when critics appeal to the fact that a classic work has passed “the test of time” or that a work has been historically very influential, they are arguing for its value on the basis of evidence from the strength of its positive past valuation. As Samuel Johnson (1709–1784) classically formulated this empirical position: “to works not raised upon principles demonstrative and scientific [e.g., mathematical or deductive], but speaking wholly to observation and experience, no other test can be applied but length of duration and continuance of esteem” [Johnson (1957, p. 9)]. This evidential approach is still widely used in contemporary evaluations of art. Lionel Trilling, for example, in advocating the aesthetic value of *The Great Gatsby*, notes that the book has “gained in weight and relevance over time” and then reasons that “this could not have happened had the book's form and style not been right as they are” [Trilling (1950, pp. 251–252)]. Another critic, Graham Hough, argues that the value of Dante's work is not logically challengeable, since “it would be a very strange position to hold that Dante's fame and influence were no evidence of literary merit” [Hough (1966, p. 76)]. Indeed, to say that an artwork has been influential and important is already to make a claim about its art-historical value.

In contrast to these logical theories of evaluative reasons, Charles Stevenson (1855–1950) offers (1957) a causal theory, urging that such reasons refer to the causes or motives that determine the critic's attitude to the work and are expressed by the critic in order to recommend those motives and attitude to the public. This sort of reasoning

can be found in a famous essay by T.S. Eliot that savagely criticized Milton's poetry, arguing that although "great" in its way, it was a grave danger and bad influence for contemporary poets because it emphasized qualities such as grandiloquence, sonority, lack of visual concreteness that were opposed to the sort of poetry that Eliot and his cohorts were trying to establish and that was exemplified by the seventeenth-century "Metaphysical Poets". Once Eliot's poetic revolution was successfully achieved, he wrote more positively about Milton while defending his earlier negative evaluation as right because of its justifiable motives in revolutionizing English poetry: "poets engaged in such a revolution will exalt the virtues of those poets of the past who offer them example and stimulation, and cry down the merits of those poets who do not stand for qualities they are zealous to realize. This is not only inevitable, it is right" [Eliot (1975, pp. 272–273)].

A third position, perceptualism, claims that reasons in artistic evaluation function as rhetorical devices for focusing attention on the work in such a way that the affirmed value will be directly perceived. The reasons do not function as logical evidence but rather as instruments of perceptual persuasion which lead the critic's readers or audience to see or experience the value the critic perceives in the work and thus to share the critic's evaluative verdict. For example, in justifying a positive evaluation of a painting, a critic may point to a particular line or feature of the work. This particular feature can lead an observer to see the work in the way the critic does and thus be perceptually convinced of the critic's evaluative verdict. Wittgenstein (1970) and many of his followers have propounded this perceptualist view of critical reasoning, and many examples of such reasoning can be found. Thus, with respect to the logical role of reasons in aesthetic evaluation, we again find that none of the major theories is entirely right or wrong, since the arguments of critics partially fit each of these theories.

(3) Closely related to the role of evaluative reasons is the issue of the general *logical form of evaluative argument*: is it inductive, deductive, or something altogether different. Beardsley (1958, pp. 471–472) argues that evaluative argument is "elliptical induction", since he denies there are any universal, absolute criteria from which we could deductively derive a true evaluative verdict.<sup>12</sup> However, Beardsley maintains that there are nonetheless criteria such as unity, complexity, and intensity that provide inductive evidence that an artwork which possesses them will be good. We have already noted that arguments relating to the test of time or historical influence are also inductive. Such arguments have themselves stood the test of time, extending from Longinus in antiquity to contemporary advocates [Savile (1984)].

In contrast, other authors claim that evaluative arguments should be deductive, based on the critic's clearly formulated norms of judgment and her description of the work in question. When we look at the practice of evaluating critics we can sometimes find arguments that look deductive in general form. Johnson and Coleridge argue deductively for the greatness of Shakespeare's dramas by trying to show these plays display the

<sup>12</sup> An example of such a universal criterion would be "all works having unity must be valuable".



essential principles of dramatic greatness. Evaluative arguments that are based on genre rules seem to have an overarching general deductive structure, where one proves the excellence of the work by showing that it meets all the necessary rules of excellence or properties of value of the esteemed genre. Joseph Addison, for example, argues for the excellence of *Paradise Lost* by demonstrating it satisfies Aristotle's "rules" for epic poetry and has all "the beauties which are essential" to that genre [Addison (1943, p. 280)]. Harold Osborne's declared norm of artistic judgment is that a work of art should be "an organic whole of interlocking organic wholes", and he maintains that any work that truly meets that description would have artistic value. However, Osborne admits that to demonstrate that a work indeed has this organic nature requires reasons that operate perceptually rather than deductively or inductively [Osborne (1952, p. 203; 1955)].

Some evaluative arguments about art do not display a clear inductive or deductive form but instead consist of a complex arrangement of focusing remarks, analogies, contrasts, leading questions, and suggested responses that attempt to bring the reader to a particular desired conclusion. Such arguments, which rely very much on the perceptualist role of reasons, have been described as dialectical or rhetorical [Shusterman (1980); Wisdom (1957)]. This style of reasoning is very salient and self-conscious in the influential English critic Frank Raymond Leavis (1845–1978) who characterizes his evaluative argument as not "measuring with a norm" but as convincing through a collaborative-persuasive appeal to the reader "in terms of concrete judgments and particular analyses: 'This – doesn't it? – bears such a relation to that; this kind of thing – don't you find it so? – wears better than that' etc." [Leavis (1976, p. 115)].

There is thus not one general logical form of argument in evaluative reasoning about art. This should not surprise us, since we saw similar plurality with respect to the logical role of evaluative reasons and the logical status of evaluative statements. Evaluative logic, to conclude, is clearly pluralistic. This need not be seen as a weakness. The different logics of aesthetic evaluation reflect the different motives, aims, and contexts that we have in evaluating art as well as the different competing values that art embodies and promotes.

Analytical philosophy clarifies the rich variety of aesthetic argument. It locates the source of value in subjective, yet communicable perceptions of truth, emotion and honor, it reveals that motives of artistic or academic recognition lead to the development of reasons in order to persuade others, and it demonstrates how the full range of heuristic devices is employed, from persuasive induction to stringent deduction, and from simple analogies to complex rhetorical patterns. Most importantly, it demonstrates that artistic valuation is a constantly shifting, competitive process. This process is played out in a variety of arenas. It takes place before festival juries, prize committees and journal editors. It generates evaluative rankings, both of the artistic value of the works and of the judgmental skill of the juries.

Economic theory is concerned with an entirely different mode of evaluation. In this mode, evaluation does not take place through judgment but through actual exchange: the buyer gives up a sum of money and gets control over the item bought. All relevant

information has been obtained and all the arguments have been delivered when the exchange takes place. Therefore, the economic value of a work is contained in the price last paid, or in the sum of the prices paid for copies or performances of an artwork.

The simplicity and unambiguity of the market mode has been identified as one of the reasons for its success. However, economic theory is well aware that the emergence and the successful performance of actual market institutions is the exception rather than the rule under conditions which deviate as far from the standard commodity assumptions as artworks do. Art works are typical “information goods”: they are uncertain in their effects, they are public in nature and their cost of multiplication has steadily decreased since the invention of print with movable letters. Under such conditions, market valuation is improbable. There may well be heated debate and exchange of words, but why should money equivalents be offered if the effect of a good is uncertain, access to it is free, and imitations are cheap?

Following recent institutionalist theory, behavioral arrangements and rules should emerge which effectively decrease uncertainty, reduce access and discriminate against imitations. In fact, a vast volume of market valuations for art works does take place in contemporary economies. Specialized professions, like agents, dealers, fair organizers and art critics, have established themselves. Special laws protect the rights of artists and performers to the commercial use of their works. Finally, the judgments of experts for rankings of artistic value provide stability for expectations of future value. There are still cases of grave impediments to exchange, like Land Art works which lack transportability and durability, but artists have shown that even such ephemeral projects can be financed through the sophisticated sale of reproductions. For the most part, artistic expression has accepted formats which improve marketability, like transportable framed canvases, or music pieces of supportable length, or gripping narratives.

While the valuation processes observed by aesthetic philosophy and by economics remain logically distinct, their performative interdependence in real life has grown. In consequence, the academic disciplines observing them gain insights by taking each other's results into consideration.

## References

- Addison, J. (1943). “Chevy Chase”. Reprinted in: Jones, E.D. (Ed.), *English Critical Essays (Sixteenth, Seventeenth, and Eighteenth Centuries)*. Oxford University Press, Oxford.
- Adorno, T.W. (1984). *Aesthetic Theory*. Routledge, London. Translated by C. Lenhardt.
- Agnew, J.-C. (1986). *Worlds Apart the Market and Theater in Anglo-American Thought, 1550–1750*. Cambridge University Press, Cambridge.
- Arendt, H. (1961). “The crisis of culture”. In: *Between Past and Future: Six Exercises in Political Thought*. Viking, New York.
- Aristotle, X. (1947). McKeon, R. (Ed.), *The Basic Work of Aristotle*. Random House, New York.
- Ayer, A.J. (1971). *Language, Truth, and Logic*. Penguin, London.
- Baumgarten, A. (1988). *Theoretische Ästhetik: Die grundlegenden Abschnitte aus der “Aesthetica”*. Felix Meiner, Hamburg. Translated from the original Latin by H.R. Schweizer.

- Baumol, W. (1986). "Unnatural value: Or art investment as floating crap game". *American Economic Review* 76, 10–14.
- Beardsley, M.C. (1958). *Aesthetics*. Harcourt Brace, New York.
- Benjamin, W. (1969). *Illuminations*. Schocken, New York. Translated by H. Zohn.
- Berg, M. (2002). "From imitation to invention: Creating commodities in eighteenth-century Britain". *The Economic History Review* 65, 1–30.
- Bourdieu, P. (1996). *The Rules of Art*. Polity Press, Cambridge. Translated by S. Emanuel.
- Budd, M. (1995). *Values of Art*. Penguin, London.
- Burke, E. (1998). *A Philosophical Enquiry into the Origin of Our Ideas of the Sublime and the Beautiful*. Penguin, London.
- Caygill, H. (1989). *Art of Judgement*. Basil Blackwell, London.
- Collingwood, R.G. (1958). *The Principles of Art*. Oxford University Press, Oxford.
- Croce, B. (1970). *Aesthetic*. Farrar/Giroux, Strous/New York. Translated by D. Ainslie.
- Debreu, G. (1959). *Theory of Value; An Axiomatic Analysis of Economic Equilibrium*. Wiley, New York.
- De Marchi, N. (2006). "Confluences of economic and artistic value: Three historical moments". In: Hutter, M., Throsby, D. (Eds.), *Beyond Price. Value in Culture, Economics and the Arts*. Cambridge University Press, New York.
- De Marchi, N., Van Migoet, H. (1999). "Ingenuity, preference and the pricing of pictures: The Smith–Reynolds connection". In: De Marchi, N., Goodwin, C.D.W. (Eds.), *Economic Engagements with Art*. Duke University Press, Durham.
- de Piles, R. (1743). *Principles of the Art of Painting*. English translation of French edition of 1708, London, J. Osborn.
- Dewey, J. (1987). *Art as Experience*. Southern Illinois University Press, Carbondale.
- Dickie, G. (1965). "Beardsley's phantom aesthetic experience". *Journal of Philosophy* 62, 129–136.
- Dickie, G. (1997). *Introduction to Aesthetics*. Oxford University Press, Oxford.
- Eliot, T.S. (1975). Kermode, F. (Ed.), *The Selected Prose of T.S. Eliot*. Faber, London.
- Gadamer, H.G. (1982). *Truth and Method*. Crossroad, New York.
- Ginsburgh, V., Weyers, S. (2003). "On the contemporaneousness of Roger de Piles' 'Balance des Peintres' ". In: Amariglio, J., Cullenberg, S., Childers, J. (Eds.), *The Aesthetics of Value*. Routledge, London.
- Goodman, N. (1969). *Languages of Art*. Oxford University Press, Oxford.
- Guillory, J. (1993). *Cultural Capital: The Problem of Literary Canon Formation*. Chicago University Press, Chicago.
- Hegel, G.W.F. (1993). *Introductory Lectures on Aesthetics*. Penguin, London. Translated by B. Bosanquet.
- Heidegger, M. (1975). "The origin of the work of art". In: *Poetry, Language, Thought*. Harper, New York. Translated by A. Hofstadter.
- Hicks, J.R. (1939). *Value and capital; An Inquiry into Some Fundamental Principles of Economic Theory*. Clarendon, Oxford.
- Hirsch, E.D. (1969). "Privileged criteria in literary evaluation". In: Strelka, J. (Ed.), *Problems of Literary Evaluation*. Pennsylvania State University Press, University Park.
- Hough, G. (1966). *An Essay on Criticism*. Norton, New York.
- Hume, D. (1752). *Political Discourses*. Edinburgh.
- Hume, D. (1963). "Of the standard of taste". In: *Essays Moral, Political, and Literary*. Oxford University Press, Oxford.
- Hutcheson, F. (1971). *An Inquiry into the Original of Our Ideas of Beauty and Virtue*. Garland, New York.
- Hutcheson, F. (2005). *A System of Moral Philosophy*. Continuum, London and New York. In two books.
- Johnson, S. (1957). "Preface to Shakespeare". In: Raleigh, W. (Ed.), *Johnson on Shakespeare*. Oxford University Press, Oxford. Reprinted.
- Kant, I. (1986). *The Critique of Judgment*. Oxford University Press, Oxford. Translated by J.C. Meredith.
- Lawrence, D.H. (1936). *Phoenix*. Penguin, London.
- Leavis, F.R. (1976). *Common Pursuit*. Penguin, London.
- Longinus (1992). "On the sublime". In: Adams, H. (Ed.), *Critical Theory since Plato*. Harcourt Brace, New York.

- Luhmann, N. (2000). *Art as a Social System*. Stanford University Press, Stanford. Translated by E. Knodt.
- MacDonald, M. (1954). "Some distinctive features of arguments used in criticism of the arts". In: Elton, W. (Ed.), *Aesthetics and Language*. Blackwell, Oxford.
- Mandeville, B. (1988). *The Fable of the Bees: Or, Private Vices, Publick Benefits*. Liberty Classics, Indianapolis. Edited by F.B. Kaye.
- Marshall, A. (1890). *Principles of Economics*. Macmillan, London.
- Menger, C. (1968). *Grundzüge der Volkswirtschaftslehre*. Mohr/Siebeck, Tübingen. Edited by F.A. Hayek.
- Mirowski, P. (1989). *More Heat than Light: Economics as Social Physics, Physics as Nature's Economics*. Cambridge University Press, New York.
- Moore, G.E. (1959). *Principia Ethica*. Cambridge University Press, Cambridge.
- Nietzsche, F. (1956). *The Birth of Tragedy and the Genealogy of Morals*. Anchor, New York. Translated by F. Golffing.
- Nietzsche, F. (1968). *The Will to Power*. Vintage, New York. Translated by W. Kaufmann.
- Noonan, D.C. (2003). "Contingent valuation and cultural resources: A meta-analytic review of the literature". *Journal of Cultural Economics* 27 (3–4), 159–176.
- Osborne, H. (1952). *The Theory of Beauty*. Routledge, London.
- Osborne, H. (1955). *Aesthetics and Criticism*. Routledge, London.
- Pareto, V. (1909). *Manuel d'économie politique*. Traduit sur l'édition italienne par Alfred Bonnet (revue par l'auteur). Giard & Brière, Paris.
- Pater, W. (1912). *The Renaissance*. Macmillan, London.
- Plato (1997). *Complete Works*. Hackett, Indianapolis. Edited by F.B. Kaye.
- Ricardo, D. (1817). *On the Principles of Political Economy, and Taxation*. J. Murray, London.
- Savile, A. (1984). *The Test of Time*. Oxford University Press, Oxford.
- Schiller, F. (1986). *On the Aesthetic Education of Man*. Clarendon Press, Oxford. Translated by E. Wilkinson and L.A. Willoughby.
- Schopenhauer, A. (1966). *The World as Will and Representation*, vol. 2. Dover, New York. Translated by E.F.J. Payne.
- Schumpeter, J.A. (1954). *History of Economic Analysis*. Allen and Unwin, London.
- Scitovsky, T. (1976). *The Joyless Economy: An Inquiry into Human Satisfaction and Consumer Dissatisfaction*. Oxford University Press, New York.
- Shaftesbury Earl of (1964). "Characteristics of Men, Manners, Opinions, Times, vol. 2". Bobbs-Merrill, Indianapolis. Edited by J.M. Robertson.
- Shusterman, R. (1980). "The logic of evaluation". *Philosophical Quarterly* 30, 327–341.
- Shusterman, R. (1981). "Evaluative reasoning in criticism". *Ratio* 23, 141–157.
- Shusterman, R. (1992). *Pragmatist Aesthetics*. Blackwell, Oxford.
- Shusterman, R. (2000). *Performing Live*. Cornell University Press, Ithaca.
- Shusterman, R. (2002). *Surface and Depth*. Cornell University Press, Ithaca, New York.
- Skidelski, R. (1992). *John Maynard Keynes*, vol. 2. Macmillan, London.
- Smith, A. (1976). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Clarendon Press, Oxford. Edited by R.H. Cambell and A.S. Skinner.
- Smith, A. (1982). *The Theory of Moral Sentiments*. Liberty Classics, Indianapolis. Edited by D.D. Raphael.
- Solkin, P.A. (1993). *Painting for Money. The Visual Arts and the Public Sphere in Eighteenth-Century England*. Yale University Press, New Haven.
- Stevenson, C.L. (1957). "Interpretation and evaluation in aesthetics". In: Black, M. (Ed.), *Philosophical Analysis*. Cornell University Press, Ithaca.
- Stigler, G., Becker, G. (1977). "De gustibus non est disputandum". *American Economic Review* 67, 76–90.
- Taylor, C. (1989). *Sources of the Self. The Making of the Modern Identity*. Harvard University Press, Cambridge.
- Throsby, D. (2001). *Economics and Culture*. Cambridge University Press, Cambridge.
- Tolstoy, L. (1997). "What is Art?" In: Adams, H. (Ed.), *Critical Theory since Plato*. Harcourt Brace, New York.

Trilling (1950). *The Liberal Imagination*. Harcourt, New York.

Walras, L. (1874). *Éléments d'économie politique pure; ou, Théorie de la richesse sociale*. L. Corbaz & Cie, Lausanne.

Wisdom, J. (1957). *Philosophy and Psycho-Analysis*. Oxford University Press, Oxford.

Wittgenstein, L. (1970). *Lectures and Conversations on Aesthetics, Psychology, and Religious Belief*. Blackwell, Oxford.

## THE ECONOMIC ANALYSIS OF ART LAW\*

WILLIAM M. LANDES and DANIEL B. LEVINE  
*University of Chicago Law School, Chicago, IL, USA*

### Contents

Abstract	212
Keywords	212
1. Introduction	213
2. Copyright and trademark issues	213
2.1. The basic law and economics of copyright	213
2.2. Copyright protection for visual art	216
2.3. Appropriation art	217
2.4. What is an unlawful adaptation of a work of art?	221
2.5. The appropriation of style and trade dress	222
3. Moral rights	224
3.1. The statute	224
3.2. The economics of moral rights	226
3.3. The cases	228
4. Resale royalties	232
5. Resolving disputes over ownership of art	234
5.1. Introduction	234
5.2. The disputes	235
5.3. The cases	237
6. Art market: Contract and authenticity issues	239
6.1. Introduction	239
6.2. Authenticity and the law	240
7. Estate issues	244
7.1. Introduction	244
7.2. Estate valuation	244
7.3. Estate management: The doctrine of cy pres and administrative deviation	248
References	250

\* We thank Victor Ginsburgh for many helpful comments and suggestions and Meghan Maloney for especially valuable research assistance.

**Abstract**

This paper surveys from an economic standpoint a number of important legal issues that influence the market for art, which include the creation, sale, valuation, maintenance and, in some instances, the destruction of works of art. We show that the important legal doctrines that bear on the visual arts can best be understood as rough efforts to promote efficiency in the art market. The paper focuses on U.S. legal doctrines and only occasionally mentions foreign law. Among the legal topics we consider are the following: copyright and trademark issues; moral rights which cover the right of an artist to prevent the mutilation and destruction of his work; resale royalties; rules governing ownership disputes between innocent parties such as a good faith purchaser and an earlier owner of the work of art; disputes over the authenticity of a work of art which cause material changes in the market value of the work; and the valuation of art-rich estates.

**Keywords**

art law, visual arts, copyright, trademark law, moral rights, resale rights, ownership disputes, authenticity issues, estate valuation issues

*JEL classification:* K00, K10, Z11

## 1. Introduction

This chapter surveys from an economic standpoint a number of important issues in the field of art law. We focus exclusively on the visual arts – by which we mean painting, photography, sculpture and prints – and do not discuss issues specific to other art forms such as music, dance, movies and theater. We approach the subject from a positive rather than normative standpoint. We analyze the economic implications of legal rules that influence the market for art, which include the creation, sale, valuation, maintenance and, in some instances, the destruction of works of art. We also show that a number of important legal doctrines that bear on the visual arts can best be understood as rough efforts to promote efficiency in the art market. The chapter focuses on U.S. legal doctrines and only occasionally mentions foreign law. Perhaps this limitation can be partially justified from data showing the U.S. domination of the art market.<sup>1</sup> Three broad areas we do not discuss in the chapter are illicit international trade of art, public funding for the arts and government censorship of art.<sup>2</sup>

The chapter is organized as follows. Section 2 discusses copyright and trademark issues. Section 3 examines moral rights and Section 4 briefly examines resale royalties. Section 5 looks at legal rules governing disputes over ownership of art. Section 6 examines legal questions that arise in connection with disputes over the authenticity of a work of art. Finally, Section 7 examines legal issues that arise in connection with the valuation and disposition of art-rich estates.

## 2. Copyright and trademark issues<sup>3</sup>

### 2.1. *The basic law and economics of copyright*

Copyright protects original works of authorship that are fixed in a tangible form. A copyright covers not just unauthorized copying but also includes rights over the distribution of copies, derivative works (called adaptation rights), and public performance and display. “Originality” and “fixation” are threshold questions that help economize on administrative and enforcement costs. “Originality” does *not* mean novel or creative but simply that the work originates with the author – i.e., the author did not copy it from another person – and has a minimal level of creativity. Its purpose is to save administrative and enforcement costs by screening out works that probably would be created even without copyright protection. “Fixation” also saves enforcement costs because it would be more burdensome for a court to decide, for example, if an alleged infringer had copied a verbal description of a not yet executed painting than the painting itself [see Lichtman (2003)].

<sup>1</sup> See Table 2 in Ginsburgh (2005).

<sup>2</sup> See Chiang and Posner (2003).

<sup>3</sup> This section draws heavily on Landes (2000, 2003) and Landes and Posner (2003, Chapters 2–6).



A basic principle of copyright is that ownership of the physical object is distinct from ownership of the copyright, and either can be transferred without the other. Thus, the purchaser of a painting acquires the physical object but not the copyright, which is typically held by the artist or his estate.<sup>4</sup> To be sure, the purchaser could also acquire the copyright but that would require a written transaction.

Copyrightable works have in common what economists call a “public goods aspect” to them. Creating these works involves significant expenditures of money, time and effort (sometimes called the “cost of expression”) while the cost of reproducing a work is typically very low or negligible. Since each copy embodies the “cost of expression”, additional users can be added at the small cost of making a copy. In the absence of copyright protection, therefore, unauthorized copying of the work would drive down the price of copies to marginal cost. This would reduce the incentives to create the work in the first place because the creator would be unable to recover his initial cost (i.e., the cost of expression). To be sure, some original works will still be created even in the absence of copyright protection. There may be substantial benefits from being recognized as the creator or from being first in the market or the copies may be of “inferior” quality. Creators may also use contract law or other private enforcement means to discourage unauthorized copying. Finally, the creator may be able to capture some of the value of copies made by others by charging a higher price for the copies he makes.

Unlike most ordinary goods, copyright protection generates access costs because the price charged for copies will be greater than the marginal costs of making and distributing copies. Access costs fall on both consumers and creators of subsequent works who substitute other inputs that cost society more to produce or are of lower quality, assuming (realistically) that copyright holders cannot perfectly price discriminate. As a result, some creators may be deterred from building upon prior works because they are unwilling to pay the price the copyright holder demands. Paradoxically, too much copyright protection can reduce the number of new works created by making it more costly for creators to build on prior works.

Another major cost of a copyright system is administrative and enforcement costs. These include the cost of setting up boundaries or erecting imaginary fences that separate protected and unprotected elements of a work; the cost of excluding trespassers; and the costs of proving infringement and sanctioning copyright violators. Unlike real property, these costs tend to be more costly for intellectual property because of the greater difficulty in defining and enforcing the boundaries of a copyrighted work.

Because the costs of establishing and administering property rights are greater for intellectual than real property, we expect and find that property rights are more limited for intellectual than real property. In this regard, several important limitations on copyright protection are worth noting.

(i) Copyright protects expression but not ideas, concepts, principles, techniques or processes (“ideas” for short). While distinguishing the two may occasionally prove

<sup>4</sup> See Landes and Posner (2003, pp. 126–128), for a general discussion of why separation of ownership of the thing and the copyright lowers transaction costs and promotes economic efficiency.

challenging, the economic justification for doing so is clear. The evidentiary costs of proving appropriation of ideas could be excessive, and the potential benefits of doing so – enjoining the production of works that wouldn't already be captured by protecting expression – could be relatively minor.

(ii) Copyright protects against unauthorized copying but not independent duplication. The likelihood that a second author might independently create a similar work to another is relatively low unless the work contains a negligible amount of original expression (in which case it will fail the originality requirement). Nonetheless, if independent duplication were actionable, authors would spend less time creating new works and more time checking prior works in order to avoid copyright liability.

(iii) The “first sale doctrine” and “right of public display” limit the copyright owner's distribution and display rights. Under the first sale doctrine, the owner of a work can resell or lease or otherwise dispose of the work without violating the distribution right (17 U.S.C. §109(b)(1)(A)).<sup>5</sup> Similarly, the owner of the work can display the work publicly or loan it to a museum for public display without infringing the copyright owner's display right.

(iv) The most important limitation on copyright protection is the doctrine of fair use, which allows unauthorized copying or adaptation of a work in circumstances that are roughly consistent with economic efficiency. These circumstances include:

- (1) high transaction cost cases in which the benefits to the copier are greater than the costs involved in locating and negotiating with the copyright holder;
- (2) implied consent, in which unauthorized copying tends to benefit the copyright holder; and
- (3) transformative uses, where the benefits from unauthorized copying are greater than the harm done.

In case (1), unauthorized copying yields net benefits but zero (or negligible) harm to the copyright holder. Case (2) involves unauthorized quotes from the copyright work in a book review or the reproduction of an artist's work in a catalog or magazine article about the artist. While it is possible that negative reviews will reduce the artist's sales, we expect that overall artists will benefit from the information art critics provide to consumers. A more subtle point is that although transaction costs between the artist and the reviewer would not prevent the latter from negotiating to copy the former's work, reviews in general are more informative and valuable if the reviewer does not have to pay or get permission from the artist in order reproduce a small amount of copyrighted material. Category (3) is limited to productive as opposed to reproductive uses of an artist's work. A parody or appropriation art are examples of productive uses while photocopying an artist's work is a reproductive use. This category invites a cost/benefit analysis in which the benefits from unauthorized copying are weighed against the harm to the copyright holder. Fair use will be found when the benefits are greater than the harm from unauthorized copying.

<sup>5</sup> There are exceptions for renting and leasing of sound recordings (CDs, tapes) and computer programs without the copyright owner's authorization.

## 2.2. Copyright protection for visual art

In the case of a unique work, such as a painting, the case for copyright protection is weakened because the main source of the artist's income and that of intermediaries, such as dealers, typically comes from the sale of the work itself rather than from the sale of copies. The opposite is true of most copyrightable works, such as books, movies, software, musical works, and, in the visual-arts domain, works of graphic art. That said, unauthorized copying of unique works will reduce an increasingly important source of income an artist receives from posters, note cards, puzzles, coffee mugs, mouse pads, t-shirts and other derivative works or adaptations that incorporate images from the original work. Without such income, some of which flows directly to artists and other indirectly by increasing the resources that museums and galleries have for acquiring art, there will be less incentive to create unique works. How much less? Perhaps not much since there also are substantial pecuniary and non-pecuniary benefits from recognition as an original artist and even unauthorized reproductions that appear on merchandise call attention to the original work, enhancing the artist's reputation and increasing the value of his works. Still, the prospect of future ancillary income will have a positive though possibly small influence on the incentive to create new works.

It is worth mentioning that copies of works of art often are viewed as greatly inferior in quality to the original, which in turn tends to reduce the economic impact of copying. When a painting attributed to Vermeer or Van Gogh is discovered to be a forgery, however skillful, its price nosedives. To take another example, vintage photographs (prints made at the time the photograph was taken) command a substantially higher price than the identical photograph printed later from the same negative.<sup>6</sup> Another possible reason for the great disparity in price between originals and copies is that the latter, because producible in essentially unlimited quantities, sell at a price equal to their (low) cost of production, whereas the supply of originals is fixed at a low level, so if originals are valued for their scarcity, copies are poor substitutes even at very low prices [Becker, Landes and Murphy (2000)]. But whatever the reasons, the important point for our purposes is simply that copying is a much smaller threat to the ability of artists to recover their fixed costs of expression than it is to the ability of writers and composers to do so.

There is also a transaction cost argument for vesting control over derivative works in the creator of the original. For example, there are several hundred ancillary products ranging from umbrellas to condoms that incorporate images from works of art created by Andy Warhol.<sup>7</sup> By concentrating the copyrights in the Warhol Foundation rather

<sup>6</sup> Consider two examples. Dorothea Lange's widely reproduced 1930s vintage photograph known as "Migrant Mother" sold at a Sotheby's photography auction October 7, 1998, for \$244,500, while an exhibition-quality print of "Migrant Mother" can be obtained for under \$50 from the Library of Congress Photoduplication Service. Edward Weston's vintage photograph from the 1929 entitled "Pepper" was sold at a Christie's photography auction in 1997 for \$74,000, while a print from the same negative, printed later by the photographer's son, had been sold at an auction eighteen months earlier for only \$1840.

<sup>7</sup> See "The Warhol Store" on the website of the Andy Warhol Museum, at <http://www.clpgh.org/warhol>.

than letting each creator of a derivative work own a separate copyright, the law avoids infringement suits involving multiple plaintiffs in which a court might have to decide which of many similar and widely accessible works the defendant had copied. Licensing costs would also rise because a potential licensee would be well advised to seek licenses from all owners of copyrights on derivative works, as well as from the owner of the copyright on the original, in order to avert the risk of being sued by one of them. The copyright on the original Warhol image is sufficient to prevent unauthorized copying of the various derivative works, since a derivative work will infringe the copyright on the original work.

Of course an alternate solution would be to deny copyright protection to works of art and their derivative works, given that the overall case for copyright protection of works of art is weaker than that for copyright protection of most other expressive works. This point should be kept in mind in deciding how the law ought to resolve close questions relating to art copyrights.

### 2.3. *Appropriation art*

Appropriation Art borrows images from popular culture, advertising, the mass media, and other artists and incorporates them into new works of art. Often the artist's technical skills are less important than his conceptual ability to place images in different settings and thereby alter their meaning. Appropriation Art has been described "as getting the hand out of art and putting the brain in". It is legally unproblematic when it copies works that are in the public domain. For example, Marcel Duchamp, an important precursor of Appropriation Art as well as of Pop Art, exhibited ready-made objects such as a urinal, bicycle wheel, and snow shovel as works of art. But when the copied image is copyrighted, the risk of a suit for copyright infringement looms.

Artists and judges tend to have different views about how the law should treat Appropriation Art. The artist perceives legal restraints on borrowing as a threat to artistic freedom:

Whenever people's response is "how dare you!" I consider that a high compliment. First of all, taking from other artists is not illegal in the art world, as it is in the music industry, and second, it is a direct acknowledgment of how we work in painting. Everything you do is based on what came before and what is happening concurrently. I don't see history as monolithic. I feel very free to take and change whatever I want, and that includes borrowing from my contemporaries. If some people are upset because my work has similarities to what they're doing, that's their problem. And if they take from me, that's great! I don't respect these artificial boundaries that artists and people around artists erect to keep you in a certain category [Rubinstein (1994, p. 103), quoting the artist Richmond Burton].<sup>8</sup>

<sup>8</sup> Another prominent member of the art community has been quoted as saying that "if these copyright laws had been applied from 1905 to 1975, we would not have modern art as we know it" [Norman (1996, p. 125), quoting art dealer Jeffrey Deitch].

Yet from the perspective of copyright law, “appropriation” of protected work connotes stealing. And so, rejecting the defense of fair use, *Rogers v. Koons* (960 F.2d 301 (2nd Cir. 1992)) held that Jeff Koons’s well-known sculpture of puppies infringed the plaintiff’s copyrighted black and white photograph, which Koons had transformed into a large, colored sculpture that arguably had little or no new expression since a black and white photograph of the sculpture looked nearly identical to the plaintiff’s photograph. The court said that “the essence of Rogers’ photograph was copied nearly in toto, much more than would have been necessary even if the sculpture had been a parody of plaintiff’s work. In short, it is not really the parody flag that [Koons is] sailing under, but rather the flag of piracy” (Id. at 310).<sup>9</sup>

We approach the copyright issues presented by Appropriation Art by way of six examples. The examples are:

- (1) *A* creates a unique collage that includes *B*’s copyrighted photograph;
- (2) *A* creates a limited edition series of prints that incorporates *B*’s copyrighted photograph;
- (3) the same, plus reproductions of *A*’s prints appear on posters, calendars, and other mass produced merchandise;
- (4) *A* creates a work that copies the outline of a nude from *B*’s photograph, the distinctive color from *C*’s monochromatic painting, and a miniature yellow square from *D*’s painting;
- (5) *A* constructs several identical sculptural works based on *B*’s copyrighted photograph or comic book character;
- (6) *A* creates a work that contains elements substantially similar to one of his earlier works owned by *B*, who also happens to own the copyright in that work.<sup>10</sup>

We begin with the examples in the first group.

1. *Creating a unique work.* Suppose an artist incorporates a copyrighted photograph from a popular magazine into a collage by cutting the photograph out of the magazine, affixing it to a board, and adding other objects, colors, and original images. No copy

<sup>9</sup> Koons’s sculpture had been prepared for a 1988 exhibition entitled “The Banality Show”. Copyright infringement suits were also brought successfully against two other Koons sculptures from the show. See *Campbell v. Koons*, No. 91 CIV.6055, 1993 WL 97381, at \*1 (S.D.N.Y. Apr. 1, 1993) (involving a copyrighted photograph of two boys and a pig); *United Feature Syndicate, Inc. v. Koons*, 817 F.Supp. 370 (S.D.N.Y. 1993) (involving the character “Odie” from the Garfield comic strip).

<sup>10</sup> The first three examples are based upon lawsuits brought by photographers against, among others, Robert Rauschenberg and Andy Warhol for using copyrighted photographs in their works. Both Warhol and Rauschenberg settled out of court. Warhol paid \$6000 in cash and royalties on the print edition of *Flowers* to the photographer Patricia Caulfield, who had threatened to sue Warhol over his flower paintings. Rauschenberg gave the photographer Richard Beebe \$3000 and a copy of the allegedly infringing work worth about \$10,000. These cases and others involving the artists Sherrie Levine, David Salle, and Susan Pitt are discussed in Ames (1993, pp. 1484–1485). Example 4 is based on a lawsuit in Germany brought by the well-known photographer Helmut Newton against the artist George Pusenkov, who claimed that his paintings “quote” rather than borrow from other artists. See Norman (1996, p. 123). Example 5 is based on the Koons litigation. Example 6, the case of an artist borrowing from his own works, is based on *Franklin Mint Corp. v. National Wildlife Art Exchange, Inc.*, 575 F.2d 62 (3d Cir. 1978).

of the photograph is made, and the photograph itself may constitute only a small part of the collage. This should be an easy case against a finding of copyright infringement. Since the publisher has paid the photographer for his work and charged consumers for copies of its magazine, allowing the photograph to be used in the collage will have no significant impact on incentives to create new commercial photographs or to publish magazines, but it will reduce access and transaction costs.

2. *Creating multiple copies.* Henri Dauman, a French photographer, sued Andy Warhol's estate over Warhol's "Jackie" series of silkscreen prints that incorporated a copyrighted photograph by Dauman of Jacqueline Kennedy that had appeared in *Life Magazine* in 1963 [see King (1997)].<sup>11</sup> Dauman also sued the estate for reproducing the silkscreen images on calendars, posters, and other widely distributed merchandise. We are more sympathetic to the copyright claim here. It might seem arbitrary to draw a bright line between a one-time use of an image lawfully acquired and reproducing that same lawfully-acquired image in multiple copies, but the distinction goes to the heart of the economic rationale for copyright. Commercial photographers are in the business of licensing reproduction rights for a variety of unanticipated uses. Without copyright protection the price of copies would be driven down to the cost of copying, leaving nothing to cover the cost of creating the work. Allowing an artist to make multiple copies without authorization poses a more substantial threat to the incentive to create new works than the one-time unauthorized copy, as in our first example. To be sure, Warhol had added substantial original expression to the original image, and his silk-screens, one of which sold in 1992 at Sotheby's for more than \$400,000, were not likely to cut into the market for the photograph. But remember that Warhol reproduced the silk-screens on posters, calendars, and other merchandise, and these *were* likely to cut into Dauman's market.

3. *Appropriating from multiple sources.* The Russian painter George Pusenkoff included in one of his paintings the outline of a nude from a Helmut Newton photograph, a distinctive bright blue background from an Yves Klein monochromatic painting, and a small yellow square from a painting by the late Russian artist Casimir Malevich [see Norman (1996, p. 123)]. Neither Klein nor Malevich's estate objected to Pusenkoff's borrowing, but Newton did and sought to have the painting destroyed. Pusenkoff's defense was that he had created a unique work rather than made multiple copies, that he had borrowed only the outline of a photograph and not the entire photograph, and that he had transformed the photograph by adding public domain material and altering the medium. But he clearly copied Newton's well-known image without paying for it and indeed his stated purpose was to copy recognizable elements from other artists – "to make canvases buzz with cultural associations by 'quoting' from other artists – a perfectly respectable post-modernist approach to picture-making" (Ibid., p. 123).

The German court in which Newton's case was brought held that Pusenkoff's painting was a "free adaptation" rather than a reworking and therefore did not infringe New-

<sup>11</sup> The case was settled.

ton's copyright (Ibid., p. 125).<sup>12</sup> Ordinarily an adaptation would be a derivative work and thus infringing if made without the authorization of the owner of the copyright on the original work, and the fact that Pusenkoff's adaptation did not substitute for the original photograph would be irrelevant. The impact of Pusenkoff's appropriation on Newton's income was surely minute and Pusenkoff unlike Warhol did not create posters and other merchandise but a unique work; but again that is not the sort of consideration that entitles the making of a derivative work without the authorization of the owner of the copyright on the original. However, Pusenkoff's adaptation was sufficiently transformative that it should be considered a fair use. This suggestion is supported by the fact that Pusenkoff's work borrowed (or "quoted") from more than one previous work. Transaction costs are likely to be high if the law requires artists to obtain permission to appropriate from multiple sources. And a work that copies from several sources is much less likely to be a substitute for any one of them. Other things being equal, therefore, the law should be more sympathetic to the artist whose work borrows from multiple copyrighted sources rather than from a single copyrighted source.

4. *Creating sculptures from a single copyrighted source.* This is *Rogers v. Koons* (1992). Koons bought a note card displaying a photograph of a group of puppies with their owners, tore off the copyright notice from the card, and hired an Italian firm to make four large sculptures called *A String of Puppies* based on the photograph. Koons's role was strictly conceptual. He did not make the sculptures himself, although he chose the subject matter, medium, size, materials, and colors. And he did not design the sculptures, at least in the usual sense, for he instructed the studio that they "must be just like photo – features of photo must be captured" (Id. at 305, quoting Koons). Altering the image to avoid a copyright lawsuit would have defeated his purpose of showing that meaning depends on context.

In rejecting the fair use defense, the court emphasized the commercial nature of the copying; the fact (the same point, really) that Koons had earned a substantial sum from the sculptures (three of the four sculptures sold for a total of almost \$400,000); that he had faithfully copied the original image; and that the sculptures were likely to impair the market for the copyrighted photograph. Although the copies and the original were sold in different markets, the court believed that Koons's type of appropriation could potentially eliminate an important source of licensing revenues for photographers.

Koons' main argument was that his work should be privileged as a satirical comment or parody. He claimed that by placing the image of the puppies and their owners in a different context from that of the original photograph, he was commenting critically on a political and economic system that overvalues mass-produced commodities and media images. Copyright law, however, requires that a privileged parody must "target" the original work rather than use the original as a "weapon" to comment on society in

<sup>12</sup> Newton was not happy with the German court's decision, and remarked: "Poor fellow, he hasn't got an idea of his own, so he has to use other people's" [Norman (1996, p. 125)]. But copyright law does not protect ideas.



general. The economic reason for this is clear. When the parody targets the plaintiff's work, the parties are unlikely to be able to agree on a price that allows the defendant to make fun of, embarrass, or even humiliate the plaintiff. (Of course, the photographer might not have realized that Koons's sculptural version of the photograph *was* parodic.) If the parodist wants to use the parodied work as a weapon to batter society rather than the work, he should have less trouble obtaining a license. But Koons wanted both to comment on the vacuity of modern American culture and – as the court failed to note – to do so by offering the copied work as an *example* of that fatuity. This makes us doubt that the case was correctly decided. There was no chance that Koons's costly sculptures would be substitutes in the market for the plaintiff's note cards. Nor was Koons planning to do Warhol-like reproductions – that would have been inconsistent with the critical message of his parodic copies.

5. *Borrowing from one's own earlier work.* Artists often return to themes they had used earlier in their careers and even copy from their earlier works. Gilbert Stuart is reported to have painted some 75 substantially similar portraits of George Washington [see Grampp (1989, p. 6)]; and Giorgio de Chirico made numerous copies of many of his best-known early Surrealist works [see Levin (1988, pp. 251–253)]. An issue of unlawful appropriation arises only if the artist no longer owns the copyright on the earlier work, which likely will occur less often today than prior to 1976.<sup>13</sup> Yet when an artist does part with his copyright, should the law allow him a fair use defense to produce derivative works?

Such a rule would spare the courts from having to determine whether a new work by the artist was a copy of the earlier work by him (whose copyright he had transferred) or an independently-created work substantially similar to the earlier one only because it was created by the same person. “If Cézanne painted two pictures of Mont St. Victoire, we should expect them to look more alike than if Matisse had painted the second, even if Cézanne painted the second painting from life rather than from the first painting” (*Schiller & Schmidt, Inc. v. Nordisco Corp.*, 969 F.2d 410, 414 (7th Cir. 1992)).

Against this evidentiary point in favor of an expanded fair use privilege is the harm likely to accrue to the artists themselves from retention of a right to make derivative works. The transfer of a copyright is of little value if the transferor retains the right to make a derivative work that may destroy the market for the transferred work.

#### 2.4. What is an unlawful adaptation of a work of art?

The contrast of two similar cases illustrates the difficulty of defining what it means to create an unauthorized derivative work of a work of art. In *Lee v. A.R.T. Co.* (125 F.3d 580 (7th Cir. 1997)), the defendant purchased note cards from the plaintiff, affixed them

<sup>13</sup> Section 202 of the 1976 Act reverses the common law “*Pushman* presumption”, after *Pushman v. New York Graphic Society*, 39 N.E.2d 249 (N.Y. 1942), that copyright automatically transfers to the purchaser of a physical work. Under the 1976 Act, transfers are only valid if made in writing and signed by the owner or his agent. See 17 U.S.C. §§202, 204(a).



to tiles, and sold the tiles at retail. Since copying was not involved, the plaintiff only claimed that A.R.T. had infringed its right to prepare a derivative work. The copyright statute defines a derivative work broadly to include “any other form in which a work may be recast, transformed or adapted”. But we noted that the “first sale” provision of the statute entitles the owner of a lawfully-acquired copy to sell or otherwise dispose of the copy without the copyright owner’s consent (17 U.S.C. §109(b)(1)(A)). Applying this “first sale” rule, the court held that A.R.T. had merely placed the equivalent of a mat or frame on the work it had purchased, and then resold the composite piece. The decision makes economic sense, as the defendant’s activity most likely benefited the plaintiff. The more tiled cards the defendant sold, the more cards he would have to buy from the plaintiff. The plaintiff’s position if accepted would give an artist the right to block any minor alteration in a work of his of which he disapproved. This would harm artists in the long run. Costs of contracting over art would rise as galleries, museums, and collectors, in order to avoid copyright liability, sought permission from copyright owners to mat and frame works of art they acquired.

So why would a plaintiff sue to stop an activity that benefits him? Maybe it doesn’t; maybe his reputation will be damaged by the alteration in his work. That is the basis of the moral rights doctrine, discussed shortly. But it is unlikely that mounting note cards on tiles would have tarnished the artist’s reputation. A more plausible explanation is price discrimination. Lee might want to charge higher prices for note cards to firms that affix them to tiles for resale to consumers. Arbitrage would make this discrimination infeasible unless the law forbade affixing cards to tiles, or selling the tiles, without the plaintiff’s consent. But it is doubtful that enabling artists to engage in this form of price discrimination is needed to give artists adequate incentives. There is no indication that Lee ever contemplated producing tiled note cards or licensing others to do so.

In the second case, *Peker v. Masters Collection* (96 F.Supp.2d 216 (E.D.N.Y. 2000)), the defendant lawfully purchased inexpensive poster-reproductions of original paintings, transferred the images from the posters to canvas, added paint to the images to replicate the original paintings, and sold the replicas for up to several hundred dollars. Here the court held that the replicas were unauthorized copies or derivative works. In both *Peker* and *A.R.T.*, each additional reproduction required the purchase of another copy of the original (either a note card or a poster) but in *Peker* the “derivative work” would compete with the original and thus undermine the incentive to create the work in the first place. Put differently, if the copier’s gains begin to eat into the market for the original work – and those gains cannot be captured through the initial sale of the poster – the first-sale doctrine becomes a drag on copyright protection.

### 2.5. *The appropriation of style and trade dress*

So far our examples have involved the appropriation of easily describable features of an original work. But what if all that can be said of the alleged copy is that it appropriates the *style* of the original work? What if the alleged copy does not appropriate the style of a single work, but rather an artist’s style in general [see Ginsburg (1995)]?

Two cases illustrate the challenges presented by the alleged appropriation of style. In *Steinberg v. Columbia Pictures Industries, Inc.* (663 F.Supp. 706 (S.D.N.Y. 1987)), the artist Saul Steinberg created a well-known *New Yorker* magazine cover depicting in the foreground several blocks of Manhattan's West Side and in the background the rest of the world beyond the Hudson River in formless obscurity. Steinberg accused Columbia of appropriating this image on a poster promoting its film, *Moscow on the Hudson*, starring Robin Williams. The poster, facing east rather than west, depicted Manhattan in the foreground and Moscow distantly in the back with only insignificant references to the Atlantic Ocean, London, Paris, and Rome in between. Superimposed on the bottom of the poster was a drawing of Williams and two of his co-stars.

The court held that the poster had impermissibly appropriated Steinberg's original style. Both works were "executed in the sketchy, whimsical style" for which Steinberg was well known. Both featured "a bird's eye view across the edge of Manhattan and a river bordering New York City to the world beyond". And "[b]oth depict[ed] approximately four city blocks in detail and [became] increasingly minimalist as the design recede[d] into the background" (Id. at 712). Although the case was probably decided correctly because the defendant had copied a number of Steinberg's expressive features that were not essential to working in his style, copyright protection of an artist's style (like protection of ideas) would impose significant access costs on other artists that probably would offset any incremental incentive benefits for artists to develop new styles. These benefits are likely to be negligible because an artist whose style attracts followers is very likely to be successful and highly compensated. Put differently, the value of a well-known artist's work inheres not in the physical object but in the artist's identity, and so is not impaired by copying.

It is worth noting that an artist's style may be protected as "trade dress" under trademark law, but protection is subject to several important limitations. First, the style in question must be so closely associated with the particular artist that it serves to identify and distinguish his work from other artists – i.e., the style functions as a brand name identifying the source of the work or in trademark lingo has acquired "secondary meaning". Second, an artist copying that style must cause a likelihood of confusion between his and the originator's work (i.e., the similarity of styles leads consumers to believe that the originator is the source or connected in some way to the copier's work or the reverse). Finally, the trade dress must be non-functional, in the sense that artists do not need to copy the originator's style in order to compete effectively in the marketplace. Since nothing in a painting could properly be categorized as functional (except, perhaps, the paint and the canvas), this constraint appears to be non-binding. In contrast, the other constraints will result in very weak trademark protection for style. Consider the Abstract Expressionist paintings of Jackson Pollock and Ad Reinhardt. Any painter who adopts Pollock's stylistic technique of splashing paint on large canvases spread horizontally on the floor, or who copies Reinhardt's style (or idea) of painting the surfaces entirely black, will produce paintings that look quite similar to Pollock's or Reinhardt's, respectively. Even if these styles had acquired secondary meaning, the copier can avoid trademark infringement by adding his signature or other identifying material that, in

effect, acts as a disclaimer of any connection with Pollock or Reinhardt. If the rule were otherwise, what would have stopped Braque or Picasso from acquiring trademark rights to cubism?

The case of *Romm Art Creations Ltd. v. Simcha Intern., Inc.* (786 F.Supp. 1126 (E.D.N.Y. 1992)) applied the concept of trade dress to the copying of an artist's style. In *Romm*, the plaintiff published posters of works by the Israeli artist Tarkay, including a "Women and Cafes" series of paintings for which the artist was well known. Simcha published posters of works by the artist Patricia that, in a fashion similar to Tarkay's, depicted women seated in various cafes. Romm sued Simcha for trade dress infringement, alleging that the similarities of the works would cause consumer confusion and enable Simcha to capitalize on the public's positive association with Tarkay's works. The court sided with the plaintiff because his style had acquired trademark significance, and there was the distinct possibility of consumer confusion, given the similarity of the works and the "lack of deliberate and measured product selection by consumers in a gallery or poster shop" (Id. at 1140, internal quotation removed). A key factor in the court's decision was that the works were inexpensive posters, for which signatures or other disclaimers were not likely to eliminate consumer confusion.

### 3. Moral rights<sup>14</sup>

In 1990 Congress enacted the Visual Artists Rights Act (VARA), which amended the Copyright Act to confer attribution and integrity rights, commonly called moral rights, on authors of works of visual art (see 17 U.S.C. §106A). Attribution rights entitle the artist to claim authorship of a work he created and to disclaim authorship if his work is altered in a manner "prejudicial to his honor or reputation" or incorrectly attributed to him. Integrity rights prohibit the intentional distortion, mutilation, or other alteration of the artist's work that injures his honor or reputation, and makes the intentional or grossly negligent destruction of a work of recognized stature actionable.

In contrast to the United States, most countries in Western Europe have a long tradition of recognizing moral rights. France recognized moral rights in the nineteenth century, and since 1928 they have been codified in the Berne Convention, to which the United States became a party in 1989. Before then several states, beginning with California in 1979, had enacted moral rights laws and the Copyright Act itself had occasionally been interpreted to confer analogous protections.

#### 3.1. *The statute*

VARA protects works of "visual art" narrowly defined as a unique work or a print, sculpture, or photograph produced in an edition of no more than 200 copies that are signed

<sup>14</sup> This section draws on Landes (2001) and Landes and Posner (2003, Chapter 10).

and consecutively numbered. Unlike the other rights conferred by the Copyright Act, the rights conferred by VARA are enforceable only during the artist's lifetime. There are other limitations as well. The artist may not transfer or assign his rights, though he may waive them in a signed document. The alteration, mutilation, or destruction of a work as a result of negligence, the passage of time, the nature of the materials used in the work, or failed conservation efforts does not violate VARA. VARA also provides no remedy for injuries to an artist's reputation caused by the presentation, display or reproduction of his work; he cannot complain that a dimly lit exhibition or a poor-quality reproduction in a pamphlet or website violates his integrity or attribution right.

VARA does not create any rights in works for hire – i.e., works prepared by an employee within the scope of his employment.<sup>15</sup> The works for hire exemption is a way of economizing on transaction costs: if the individual employee's creative contribution were protected by VARA, then since the employer's use of his contribution might alter, mutilate, or even destroy the work, the employer would insist on a waiver – the execution of which is costly. Or suppose a developer commissioned a large sculptural work as part of a building project and, after the work was completed and installed, decorated it with Christmas or Easter decorations that the artist regarded as degrading. Anticipating such possibilities the developer would require a waiver of VARA rights when he commissioned the sculpture. Excluding this type of commissioned work from VARA coverage saves the transaction costs that waivers impose.

Another important limitation of VARA coverage involves installed works that are likely to be mutilated or destroyed if they are ever removed. VARA provides that there is no integrity right for a work installed after the statute's effective date of July 1, 1991, provided the artist consented in writing to both its installation and the possibility that removal might mutilate or destroy the work, or, in the case of a work installed prior to the effective date, provided the artist consented to its installation. If an installed work can be removed without being mutilated or destroyed, the artist retains his integrity right unless the building owner notifies him that he intends to remove the work and gives the artist a reasonable opportunity to remove it at his own expense.

Consider a building owner who hires an artist to create a site-specific sculptural work for the building's entrance plaza. Their contract is silent on the artist's rights if the sculpture is ever removed, though removal would destroy the sculpture. A new owner acquires the building and wants to tear it down and build a modern office building in its place. Never having consented to the possible destruction of his work, the sculptor would be in a position, given VARA, to extract a substantial payment from the new owner for allowing the project to go forward. The owner might argue that the sculpture

<sup>15</sup> A work created pursuant to a formal employment relationship (as when Disney hires an animation artist who is paid a regular wage, receives fringe benefits, and can be assigned to work on different projects) is an unambiguous example of a work for hire. But a commissioned work executed by an independent artist may also be a work for hire if, for example, the commissioning party pays a monthly stipend, pays health and other fringe benefits during the time the artist works on the project, defrays the cost of materials, and exercises overall though not necessarily daily supervision.

was a work for hire, but commissioning a work does not necessarily turn it into one. So building owners would demand written waivers of VARA rights at the time works such as we are discussing were commissioned. If they proved difficult to obtain, building owners might forgo installing artwork in buildings for fear of future legal problems.

VARA's integrity right protects an artist against alterations that injure his "honor or reputation" and against the destruction of his work if it is of "recognized stature". The terms "honor" and "reputation" were borrowed without attempt at definition from moral rights laws in European countries, but the intended meanings are reasonably clear. Reputation is a matter of what other people think of one, and an artist's reputation is primarily a matter of what art lovers think of the artist's work. An injury to an artist's reputation is likely to affect the prices of his artworks and thus, if he is still active or has retained copyright or ownership of some of his works, his income. Honor is a related concept but includes self-esteem and need have no pecuniary implications. Someone might intentionally mutilate the work of an unknown artist, injuring the artist's self-esteem yet inflicting no financial injury on him because the work had no market value. Finally, the statutory term "work of recognized stature" has been interpreted to require only minimum public acknowledgment of a work's quality or significance (see *Martin v. City of Indianapolis*, 192 F.3d 608 (7th Cir. 1999); *Carter v. Helmsley-Spear, Inc.*, 861 F.Supp. 303 (S.D.N.Y. 1994), reversed on other grounds, 71 F.3d 77 (2nd Cir. 1995)).

### 3.2. *The economics of moral rights*

Attribution rights are closely related to rights against fraud and trademark infringement (and also to the norm against plagiarism), so that much of what they seek to prevent is already unlawful. For example, the laws against fraud would forbid someone who was not Jasper Johns to paint a picture in Johns' style, sign it "Jasper Johns", and attempt to pass it off in the market as Johns' work. And removing Johns' signature from an original Jasper Johns painting and selling it under one's own name would violate trademark and unfair competition law, although such a forgery is unlikely because it would reduce the painting's market value. Not only does VARA add little if anything to the rights that an artist already has in such cases; but those rights, unlike the rights created by VARA, do not expire with the artist's death. For these reasons we are not surprised to have found no cases in which a plaintiff sought to enforce an attribution right.

The picture is different when we turn to integrity rights. It is true that, just as with attribution rights, even if there were no moral rights law there would be alternative methods for securing protection of the artist's integrity rights – most obviously contract law (though also copyright law, as we shall note shortly). An artist concerned with the possible future alteration of his work could add a term to the original sales contract giving him the right to approve or veto future modifications of the work. But there would be two drawbacks. First, most sales contracts in the art world are oral rather than written,<sup>16</sup> so to protect integrity rights the parties would have to incur the added costs of

<sup>16</sup> Sixty-one percent of the respondents to a survey conducted by the Copyright Office stated that oral contracts are more common than written ones in the art world [U.S. Copyright Office (1996, Table 3-2)].

a written contract. Second, a moral rights provision would not be enforceable against subsequent purchasers of the work unless the original purchaser included it in his contract with the subsequent purchaser and the latter, if he resold it, did likewise. It should also be noted an artist can protect the integrity-rights component of moral rights law simply by enforcing his copyright. Because copyright embraces the exclusive right to make and authorize others to make derivative works, any significant distortion, mutilation, or modification of an expressive work without the artist's consent would infringe his copyright as an unauthorized creation of a derivative work [see, for example, *Ty, Inc. v. GMA Accessories, Inc.*, 132 F.3d 1167, 1173 (7th Cir. 1997); *WGN Continental Broadcasting Co. v. United Video, Inc.*, 693 F.2d 622, 626 (7th Cir. 1982); cf. *Gilliam v. American Broadcasting Cos.*, 538 F.2d 14 (2nd Cir. 1976); but see *McCartney* (1998, pp. 43–52), expressing skepticism]. And unlike VARA rights, copyright does not expire with the death of the copyright holder; indeed, under current law, the death of the author of the copyrighted work triggers a 70-year extension of copyright protection.

VARA retains significance in cases in which the copyright to the artwork is transferred along with the work itself. And the derivative-works provision of the Copyright Act would not prevent the intentional destruction of a work of visual art, as VARA does for works of recognized stature – though it is a little hard to see how this can matter very much. The owner of a work of “recognized stature” would rarely want to destroy it, though this depends on how low the “recognized stature” threshold is set, and we’ll see shortly that the tendency of the courts has been to set it very low.

*Hansmann and Santilli* (1997) have offered the following economic argument in favor of moral rights that cannot be subsumed, as we have just been suggesting, under conventional copyright law. The value of a work of art depends in part on the artist's reputation, which is embodied in the entire stock of his works, with each piece acting in effect as an advertisement for the others. Mutilating (or destroying) any one of them thus imposes a cost on the artist that is external to the mutilator (assuming he has acquired the work lawfully and so could not be punished for theft or malicious mischief). A moral rights law causes this cost to be internalized. Yet, cutting the other way, the destruction or mutilation of a single work will reduce the effective supply of the artist's works and by doing so increase rather than decrease the value of the remaining works plus any works that he creates in the future. And as long as it was known that the artist had not committed or condoned the mutilation, all the mutilation would demonstrate was that one person disliked the artist's work intensely or wanted to subject it to ridicule – and to prevent such a mutilation would be like forbidding a parody.<sup>17</sup> Given the point noted earlier that copyright law's derivative work provision itself provides some protection against mutilation (but not destruction); it is doubtful that the incremental protection

<sup>17</sup> It is true that the parody does not alter the parodied work and the mutilation does. But why should artists be immunized from this form of criticism, when persons who want to criticize the United States by burning the American flag have a constitutional right to do so? It seems, if anything, particularly fitting that criticism of a visual work should take the form of altering its appearance. A verbal work invites a verbal parody, a visual work a visual one.

added by VARA can be cost justified. We are not surprised that in the end Hansmann and Santilli draw back from claiming that moral rights are, on balance, efficient.

### 3.3. *The cases*

The paucity of litigation under VARA suggests that moral rights have little value for most artists. The evidence is not conclusive, if only because most cases are settled. But the settlement rate is unlikely to be very high in a new area of law, where there are no precedents to guide the parties in predicting the outcome of litigation if it is not settled. Nor have we been able to find more than a handful of newsworthy disputes involving moral rights (other than those in the decided cases), and most of these antedate the enactment of moral rights laws in the United States.<sup>18</sup> The infrequency of such disputes is not surprising since self-interest provides a powerful incentive for owners of art not to mutilate or destroy it.

All but one of the decided cases involved disputes between property owners and sculptors. And in only one did the artist prevail. All involve relatively unknown artists who had created (with two exceptions) large-scale sculptural or site-specific works that would have been or were substantially damaged or even destroyed as a result of new construction or renovation. The better known an artist is, the more valuable his work is likely to be and therefore the less likely it is that destroying or mutilating it would be an attractive option for its owner. And paintings and other smaller works are likely to be more valuable than the cost of moving them out of the way of whatever activity endangers them.

*Carter v. Helmsley-Spear, Inc.* (71 F.3d 77 (2d Cir. 1995)) is the best known of the cases. Artists known as the “Three-Js” created a vast lobby sculpture, using more than 50 tons of recycled materials including a school bus, in a commercial building

<sup>18</sup> Prior to the enactment of VARA, the principal disputes were the following:

- (1) A massive black and white Calder mobile installed in the rotunda of the Pittsburgh International Airport from 1958 to 1978 was repainted green and gold, the colors of Allegheny County, and motorized to turn at regular intervals.
- (2) Clement Greenberg, the distinguished art critic and trustee of the David Smith estate, stripped the paint from six of Smith’s sculptures after Smith’s death because he believed it would improve their aesthetic and market values.
- (3) A sculpture by Isamo Noguchi that had been displayed in the lobby of the Bank of Tokyo Trust Company in New York was removed, cut into pieces, and destroyed in 1980.
- (4) Diego Rivera painted a large wall mural in Rockefeller Center in 1933 that included a portrait of Lenin near the center and people marching with red flags past Lenin’s tomb – elements that were not part of Rivera’s original proposal. Rivera refused a request to replace Lenin’s head with Abraham Lincoln’s (!). The owners temporarily covered the mural and then destroyed it.
- (5) Richard Serra’s site-specific sculpture “Tilted Arc” was removed from the Federal Plaza in lower Manhattan after complaints that the sculpture was a safety hazard and prevented the public from using the space for recreation.

Examples 1 through 3 are taken from U.S. Copyright Office (1996, Chapter 2). Example 4 comes from Robinson (1983, p. 9), and example 5 from *Serra v. U.S. General Services Admin.*, 847 F.2d. 1045 (2nd Cir. 1988).

in Queens. The work was never completed, although the artists had worked at it for more than three years. A default by the original owner of the building led to a change in management, and the new management evicted the artists. Fearing that the management would destroy the sculpture, they sought an injunction under VARA. The district court ruled in their favor but the court of appeals reversed, holding that the sculpture was a work for hire. Although the artists had full authority in matters of design, color, and style, while the building management retained authority over the location and installation of the work, the artists had received a weekly salary, based on a 40-hour work week, for three years; they had received employee benefits including unemployment and health benefits (two of the artists filed for unemployment benefits after the new managers fired them); and payroll and social security taxes had been deducted from their weekly salary checks. Nevertheless the result is questionable. The contract between the Three-Js and the building's original owner stipulated that the artists retained the copyright to the work, implying that the parties did not envisage the sculpture as a typical work for hire (on which the employer owns the copyright and can therefore do with the copyrighted work what he wants). The plaintiffs had no copyright remedy because copyright does not confer a right to prevent the physical destruction of the copyrighted work by its owner.

*English v. BFC&R East 11th Street LLC* (1997 U.S. Dist. LEXIS 19137 (S.D.N.Y. 1997)) involved a group of related artworks, including both sculpture and murals, installed in a community garden on East 11th Street in New York City. A development was planned that required moving the sculptures but leaving the murals intact, though it would obstruct the view of the murals. The artists claimed that their work had been conceived as a unity, which the planned development would mutilate. Without deciding whether it was either a single work or one of recognized stature, the court held VARA inapplicable because the artwork had been placed on the property illegally. The previous owner (New York City) had never authorized the artists to put their work on the site, although it had remained there for many years without the City's trying to remove it. The court worried that a ruling for the plaintiffs would entitle artists to freeze real estate development by affixing graffiti<sup>19</sup> to construction sites. The court also noted that to accept the plaintiff's argument would force the City either to incur prohibitive costs of continually patrolling its many vacant lots or to acknowledge a "squatters rights" limitation on its property rights. It was more efficient to place the burden on the artists of obtaining the City's explicit consent to their use of the property.

In *Pavia v. 1120 Ave. of the Americas Associates* (901 F.Supp. 620 (S.D.N.Y. 1995)), the plaintiff's large bronze sculpture, comprising four standing forms, had been on display in the lobby of the Hilton Hotel in New York City from 1963 to 1988. The plaintiff had retained the title to and copyright in the work. In 1988 the owner of the hotel removed the sculpture, placing two of the four pieces in storage and displaying

<sup>19</sup> At least graffiti that are not purely verbal, in which event, as we shall see shortly, they may not be visual works and therefore may be outside the scope of VARA.



the remaining two in a parking garage. Since the artist had retained title to the work, the court held that he had rights under VARA even though the work had been created before the Act's effective date. Nonetheless, the court rejected the plaintiff's claim because the mutilation had also occurred prior to that date. This left only the plaintiff's claim of injury due to the unfavorable display of his mutilated work. VARA does not cover display rights, but New York's moral rights law does, and the court held that the plaintiff had a valid claim under that law that was not preempted by the federal statute.<sup>20</sup>

In *Martin v. City of Indianapolis* (1999), the defendant city had destroyed the plaintiff's 40-foot outdoor sculpture as part of an urban renewal project. When installed in 1986 the sculpture had been engineered in such a way that it could be disassembled and removed. Remember that VARA provides that a work created before the Act's effective date may not be destroyed unless the artist is given notice of the impending destruction and an opportunity to remove the work at his own expense. Through a bureaucratic foul-up the artist was notified but not given sufficient time to remove the work. Liability also depended on the work's being of recognized stature, but as to this the court held that the plaintiff satisfied his burden of proof by submitting local newspaper and magazine articles describing the work. No expert witness testified and there were no critical writings on the work or the sculptor. The dissent argued that more evidence of "recognized stature" should be required since otherwise buyers of works of art would be required in virtually all cases to obtain VARA waivers at the outset or face a risk of violating the statute in the future because the work, obscure when commissioned, later attained the requisite recognition.<sup>21</sup>

In *Flack v. Friends of Queen Catherine Inc.* (139 F.Supp.2d 526 (S.D.N.Y. 2001)), still another New York City sculpture case, the sculptor Flack was commissioned by a group of local boosters, the Friends of Queen Catherine (referred to in the opinion as "FQC"), to design a monumental statue of Catherine of Breganza (a seventeenth-century princess of Portugal and queen of England) for installation in the New York borough of Queens, for which Queen Catherine has some unexplained significance. The project was abandoned when it was discovered that Catherine and her family had profited from the slave trade. Flack had created a 35-inch clay model of the statue, and in the commotion attending the abandonment of the project the head of the model had been placed outdoors and suffered damage. FQC, which owned the statue and still wanted it cast in bronze, hired another sculptor to resculpt the face. Flack charged that this sculptor was grossly negligent and had produced "'a distorted, mutilated model' in which . . . the nose, nostrils, eyes and lips [were] uneven and the wrong size" (Id. at 530).

<sup>20</sup> There is an interesting addendum to the case. Three of four 3000-pound Pavia bronzes disappeared from Manhattan storage facilities in early 2005. Shortly thereafter, a scrap dealer returned the bronzes. The four pieces are being sent to the museum at Hofstra University in New York.

<sup>21</sup> Compare *Carter*, where Hilton Kramer testified for the defense and claimed that the work had no merit and no recognized stature. He based his argument of the fact that there was no literature on either the artists or the sculpture. The judge rejected Kramer's testimony on the ground that Kramer is hostile to all modern art!

The court, denying summary judgment for FQC, ruled that if Flack could prove that the substitute sculptor had been grossly negligent he was entitled to prevent FQC from casting the altered head in bronze. As a preliminary to this ruling the court held that the clay model was a work of art in its own right even though it had been intended to be only the mold for casting the bronze statue. Clay models made by recognized sculptors are accepted in the art world as works of art, as the court pointed out. But a more straightforward point is that the bronze casting, though it would not be done by Flack personally, would nevertheless be “her” creation within the most sensible reading of VARA. She would own the copyright in it and the mutilation would violate her integrity right.

Finally, *Pollara v. Seymour* (206 F.Supp.2d 333 (N.D.N.Y. 2002)) involved a large “protest mural” (actually a painting – it was not painted on a wall) that in the words of the court “depicted stylized figures of various races and socio-economic situations standing on line outside closed doors to legal offices. The mural also contained the phrases “Executive Budget Threatens the Right To Counsel” and “Preserve the Right To Counsel” (Id. at 335). The mural was installed in a state government building in Albany without authorization and the same evening was removed by government employees; in the course of removal, the mural was badly damaged. The court held that because the work “was intended solely as a display piece for a one-time event” and “there was never any intent to preserve [the] work for future display”, the work lacked the statutorily-required “stature” (Id. at 336). Stated more simply, since the artist herself had not intended the work to endure, the damaging of it inflicted no harm on her. It was not as if the defendant intended to exhibit the mural in its damaged form. It gave it back to her after removing it.

One of the cases in which a violation of VARA was alleged but not ruled on involved an assault claim by the artist Moncada against gallery owner Lynn Rubin. Moncada had painted a mural containing only the words, “I am the best artist, Rene”, on a building opposite Rubin’s Soho gallery; Rubin allegedly assaulted Moncada when she discovered the artist videotaping her attempt to remove the mural (*Moncada v. Rubin-Spangle Gallery, Inc.*, 835 F. Supp. 747 (S.D.N.Y. 1993)). The only question before the court was whether the defendant’s liability insurance policy covered an intentional tort, but the facts of the case bring to light interesting questions that may arise in future VARA disputes. First is whether the mural was a work of visual art; unlike Pollara’s mural it consisted entirely of the plaintiff’s signature and a single sentence proclaiming his artistic skill. If this is a work of visual art, how are other writings, such as a student’s homework or poem, that obviously are not protected under VARA, to be distinguished from it? This is not to deny that in some cases words can be an integral and therefore protected part of a visual work; moreover, we regard the only test of what is art to be what is accepted as art and maybe Moncada’s mural passed that test.

Second, although a tenant may have authorized Moncada to paint the mural, there is no indication that the building owner had authorized it. If a tenant’s authorization were sufficient to establish VARA rights (which it surely should not be), VARA might protect an unlimited number of graffiti artists and doodlers who decorate the outside walls of

their apartments; so cleaning and repainting walls throughout New York City would risk violating VARA.

Third, even if the plaintiff could show that the mural was a legally-authorized work of visual art, he would have to show that it was of recognized stature, which, given its purely verbal character, might have been difficult. (Of course the *Martin* case above suggested this test to be quite loose.) Note that this was not an issue in *Pollara*, because, as the court noted in an earlier opinion in the case, the complaint was not the removal of the mural from the government building but the damaging of it in the course of removal, and it could have been removed without being damaged, for it was not, despite being called a “mural”, built in (*Pollara v. Seymour*, 150 F.Supp.2d 393 (N.D.N.Y. 2001), at 396, n.4).

Our sample of VARA cases is small, but informative. They suggest that VARA disputes are likely to be limited to works of visual art that cannot be moved without damaging or destroying them; that the works are unlikely to be valuable works by well-known artists; and that judges are reluctant to preserve art at the cost of hampering development. Conceivably, too, the cost of obtaining waivers from artists may deter museums and galleries from exhibiting installation art that cannot be removed without destroying it and may deter property owners from commissioning works for installation in open spaces, lobbies, and buildings. The sheer paucity of cases, moreover, suggests that VARA did not fill some yawning gap in liability space. It is not as if the statute were so clear, or the penalties for its violation so draconian, that full compliance could be expected to be achieved immediately, obviating litigation.<sup>22</sup>

#### 4. Resale royalties

Resale rights on art, commonly known as *droit de suite*, provide an artist or his heirs a small royalty (usually less than 5 percent) each time the artist’s work is sold at auction and sometimes by a dealer. Eleven EU countries currently levy resale rights of various magnitudes and all must provide resale rights by 2006 according to a recent Directive of the European Parliament [[Ginsburgh \(1996, 2005\)](#)]. There is no federal resale royalty law in the United States although California has one for works resold in that state.

The basic argument in favor of resale rights is that they enable artists or their heirs to capture some of the increase in the value of their work after the initial sale. Proponents of resale rights often invoke the example of Van Gogh who, despite his paintings selling for millions today, sold only a couple works during his lifetime and died in poverty.<sup>23</sup> Resale rights, it is argued, would have prevented that and similar injustices.

<sup>22</sup> The inference from the paucity of VARA cases is reinforced by the extraordinary paucity of cases under state moral rights laws.

<sup>23</sup> It is far from clear whether the Van Gogh example supports the claim made by proponents of resale royalties. Van Gogh committed suicide at age 37 after an intense period of painting in the South of France. Since many of Van Gogh works were widely exhibited shortly after his death, it seems likely that he would have achieved critical and financial success at a relatively young age had he not committed suicide.

The economic case for resale rights is very weak. Resale rights are merely an interest in a future stream of cash flows. In a world with no transactions costs, the value to the artist of retaining that interest (as in a resale rights regime) should equal the value realizable from selling it (as in a regime without resale rights). In other words, first purchasers will discount the price that they pay for an artwork by the entire value of the resale right. Once you add in the costs necessary to administer a resale rights system, the simpler alternative of paying an artist upfront for the entire value of his work wins hands down.<sup>24</sup> Another disadvantage of resale royalties is it prevents artists from fully shifting the risk of future price declines to the original purchaser. And if an artist is more optimistic than the market is about future price increases of his work, he can always retain in his inventory a few paintings for future sale that will provide him an amount equivalent to what he might receive in resale royalties.

Worse yet are the distributional effects of resale rights. Art purchasers will lower the prices they pay for art that is subject to resale rights. But artists (or their heirs) only receive payments if their works sell at a price above the initial sale price. Thus artists whose works do not sell or do not increase in value – “the vast majority” (Ibid., p. 4)<sup>25</sup> – will only experience the downside of resale rights legislation. They will in effect subsidize the more successful artists and the buyers.

Not surprisingly, many artists oppose resale rights legislation “on the grounds that it damages the market for their work and the art trade in general” [*The Art Newspaper* (January 2000)]. Resale rights, they assert, are a “[restraint] on testamentary freedom” (Ibid.) because they force living artists to bear a burden today for benefits which most likely will be reaped if at all by their heirs.<sup>26</sup> Another strike against resale rights regimes is the prospect that they will drive auction business to countries without resale rights [*Wall Street Journal* (April 6, 2000)]. Harder times for auction houses in resale rights countries could further depress art prices and undermine the incentive for artistic production. Resale rights may result in less art being produced, sold, and enjoyed.

<sup>24</sup> The situation would be slightly different if artists systematically valued their expected resale rights payments more highly than did the buyers of their art – perhaps because, being confident in their abilities, they discounted them at a lower rate (less of a discount). Putting aside administrative costs, a resale rights system would enable artists to unbundle the portion of their product that was profitable to sell from that which was not. The opposite would be true if, being risk-averse, they applied a higher discount rate to the stream of payments than do buyers.

<sup>25</sup> In France, between 1993 and 1995, 2–3% of the 2000 artists who benefited from resale rights received 43% of the payments collected [*The Art Newspaper* (January 2000)].

<sup>26</sup> In Germany, for example, only 274 of 7454 artists registered with the resale rights agency were eligible for payment (Ibid.).

## 5. Resolving disputes over ownership of art

### 5.1. Introduction

Ownership disputes are more likely to occur if an object is valuable, mobile, relatively easy to hide, but still traceable – conditions frequently satisfied in the case of works of art. First, because a valuable work of art is likely to appreciate rather than depreciate over time, a search even for a work that has been missing for many years may be profitable. Second, the same characteristics, as well as durability and ease of concealment increase the possibility that works of art will resurface, often by chance, after many years of being thought destroyed or otherwise gone for good.<sup>27</sup>

Uniqueness facilitates recollection and thereby increases the possibility of tracing a lost work: the few people who have seen the work will remember having seen it and can provide valuable information to the searcher. Tracing is also facilitated by the fact that the value of a work depends on its provenance (ownership and exhibition history) and the work being preserved in its original form. The more that is known about a work, and the less the work has been altered, the easier it will be to find.

Although a registry for art – similar to a land registry – would eliminate many ownership disputes, it is probably not feasible. Most art is mobile (unlike land), so a comprehensive registry would have to be global in scope and encompass millions of works. The feasible alternative to a comprehensive registry is a registry limited to missing works of art, such as the Art Loss Register (ALR) (<http://www.artloss.com>), since works that are both missing and sufficiently valuable for owners to search for them are a tiny subset of all works of art. The ALR lists over 100,000 lost or stolen works, to which it adds approximately 10,000 new listings annually [*The Art Newspaper* (March 1998)] and it has become the standard for auction house due diligence. Each year, ALR checks 400,000 lots slated for auction against the database (*Ibid.*).<sup>28</sup> A registry limited to missing art will be valuable, as there will usually be sufficient time between the theft or loss of a work (assuming the owner discovers and reports it missing) and its resale for auction houses, galleries, or buyers to check the registry to determine its provenance.

A final explanation for why the chain of ownership of a long-lost work of art can often be traced has to do with why people own art. Economists since Veblen have described art as a prestige good that enables the collector to signal to others that he is a person of both wealth and good taste. The collector gets utility not only from admiring the work hanging in his living room but also from believing that other people envy or admire him because he owns it. To obtain this additional utility, people who buy art don't want to keep it hidden away. They brag about owning it, show it to friends, and lend it to

<sup>27</sup> This section is based on Landes and Posner (1996).

<sup>28</sup> Object ID, a uniform registration protocol for detailing lost or stolen works developed by the Getty Information Institute has further augmented effectiveness of the ALR. See <http://www.artloss.com/>.

museums and galleries for exhibitions. Art's signaling value makes it more likely that the rightful owner will eventually find his long-lost work.<sup>29</sup>

## 5.2. *The disputes*

We are interested here in ownership disputes involving “innocent” persons. But we use the term in a very broad sense, excluding – besides thieves – only persons who know they are buying a stolen work of art as opposed to buyers who fail to take proper precautions against the possibility that the work is stolen. The “good faith” purchaser, roughly one who both believes that the work he is buying is not stolen and takes optimal precautions against a mistake about title (which nonetheless fail to reveal the work is stolen), is thus a type of an innocent purchaser.

The choice of what legal rule to apply to ownership disputes over art requires taking into account the actions of three parties: rightful or original owners (owners, for short), innocent purchasers, and thieves.<sup>30</sup> Our aim here is to identify the issues that should be considered in an economic approach to this problem. Dispute resolution rules directly affect owners and buyers, as each is aware of the possibility that any particular work either will be or has been lost or stolen and thus become the subject of an ownership dispute. They indirectly affect thieves by having an impact on the market for stolen works.<sup>31</sup>

Dispute resolution rules influence the level of precautions that a buyer takes to prevent loss or theft and also what resources he expends to recover a work that has been lost or stolen. In equilibrium he will take just enough precautions such that the marginal cost of additional precautions equals the marginal value of such precautions (the decrease in the probability of loss or theft multiplied by the value of retaining the work). If a work is lost or stolen, he will expend just enough resources searching for it such that an additional dollar of searching will yield an additional dollar in expected return (equal to the increase in the probability of finding the work multiplied by the time-adjusted value of regaining the work *and* discounted by the possibility that the legal rule will not award the found work to him). The more the legal rule favors previous owners over innocent purchasers (e.g., by extending the time period an owner has to find the lost work and demand its recovery or, equivalently, by tolling the statute of limitations until the owner locates the lost work), the less an owner will spend to prevent a work's loss or theft, and the more he will spend to recover a work that is lost or stolen.

<sup>29</sup> This suggests that the high bidders for known stolen art will be art lovers rather than status seekers because the former group are more likely to be willing to conceal their ownership.

<sup>30</sup> Since each owner (other than a work's creator) is a buyer himself, it may seem erroneous to consider the two as separate parties. Here we consider separately persons *in their capacities* as buyers or owners. The costs and benefits of their actions are thus additive.

<sup>31</sup> We ignore the interests of other (fourth) parties in how ownership disputes are resolved. These interests include loss of secondary enjoyment of a work (if the legal rule forces the work underground) and pity at either a previous owner or an innocent purchaser losing out in an ownership dispute. (Such pity could be a significant consideration in disputes over Holocaust Art.)

We suspect that the effect of making a legal rule more owner-friendly will be smaller on the level of an owner's precautions than on his search expenditures. Owners already have a significant incentive to prevent a work from being lost or stolen; given that the prospect of recovering a work is small to begin with, a slightly more or less owner-friendly rule for dispute resolutions should only have a small impact on the owner's precautions in preventing theft. On the other hand, once a work disappears, the calculus changes significantly. The only reason to spend money looking for a work is if finding it will result in its return.

Innocent purchasers believe there is some possibility that works they own have defective titles. This risk will justify some amount of precautions to avoid ownership disputes with rightful owners; such precautions may include not displaying a work publicly or not publicizing one's ownership of it – a cost in the form of lost prestige value. An owner will conceal his ownership of a work just enough such that the marginal loss in the owner's prestige is equal to the risk mitigation of keeping a work out of the public's eye (the decrease in the likelihood the work is discovered times the value of the work and discounted by the likelihood that the legal rule will not return the work to the original owner). The equilibrium investment in such precautions increases as dispute resolution rules become more owner-friendly.

What sort of dispute resolution rule (owner- or innocent purchaser-friendly) will minimize theft (and thus the costs of theft) is unclear. Owner-friendly rules will decrease both the level of precautions that owners take (good for thieves) and also the prices that innocent purchasers are willing to pay for potentially stolen art (bad for thieves). Buyer-friendly rules will do the opposite. We suspect however that the net effect of an owner-friendly rule is a decrease in theft. Owners take fewer precautions in response to more owner-friendly rules in part because the decline in the price of stolen works will directly decrease theft rates. The theft-increasing effect of lower precautions is thus moderated by the theft-decreasing effect of lower prices for stolen works. Moreover, as noted earlier, the change in precautions due to changes in legal rules is likely to be small to begin with; thus the increase in the amount of theft due to the decrease in precaution spending is unlikely to drown out the direct decrease in theft caused by a decline in prices associated with a more owner-friendly rule.

The fact that owner-friendly dispute resolution rules may decrease theft does not indicate whether they are socially efficient; for that analysis, one must estimate the aggregate costs across society of the behavioral responses described above. The simplest analysis would compare two extreme legal rules: one that always returned a disputed work to the original owner (call it Rule O), and one that never did (Rule P). From what we've already said above, the costs associated with moving from Rule O to Rule P would be increased precautions and increased theft costs; the benefits would be the elimination of search costs (there being no reason for an owner to recover his work) and concealment costs (there being no reason for the subsequent purchaser to prevent the original owner's discovery of the work).

We expect the search and concealment costs under Rule O (and thus the cost savings associated with choosing Rule P) to be low relative to their benefits. Beyond the small

cost of registering the lost work (about \$65 for posting notice of a stolen work on the Art Loss Register), the marginal productivity of search is likely to diminish very rapidly. Except in extreme circumstances, dispossessed owners are more likely to “wait-and-see” if a work resurfaces than to expend further resources searching fruitlessly. Secondly, if a purchaser has acted in good faith, he will think it quite unlikely that the work has a defective title and will lower his concealment expenditures (by displaying the work) accordingly. On the other hand, the likely benefits of choosing Rule O versus Rule P, lower precautions and lower theft costs aggregated across society, are not insignificant. We noted earlier that owners always have an incentive to protect their art, but that incentive is certainly greater when there is *no possibility* of recovery from innocent purchasers, as under Rule P. And unlike expenditures on search costs, there is no obvious drop-off in the return available for precautionary expenditures.

All this yields a conclusion that rules that favor owners over innocent purchasers are more socially efficient than those that favor innocent purchasers. Such a finding is consistent with the practice of most societies around the world [see [Levmore \(1987\)](#)]. The question then becomes why statutes of limitation should exist at all, or for that matter, any limitations on the rights of owners. Most likely, there is a turning point where the error costs associated with enforcing dated causes of action overcome the benefits of an owner-friendly system.<sup>32</sup> However, when error is unlikely, for example, when a plaintiff can present reliable records or testimony evidencing his ownership of a work, we should expect courts to stretch statutes of limitation for the owner’s benefit. The cases we discuss below confirm this intuition.

### 5.3. *The cases*

In *O’Keeffe v. Snyder* (416 A.2d 862 (N.J. 1980)), the artist Georgia O’Keeffe sued a good faith purchaser for the return of paintings she contended had been stolen from her in 1946. O’Keeffe had waited until 1972 to report the disappearance of the paintings, learned of the location of the paintings in 1975, and discovered the identity of their current possessor (Snyder) in early 1976. When Snyder rejected her demand to return the paintings, O’Keeffe sued for their return.

The key issue in the case was whether a six-year statute of limitations period barred O’Keeffe’s action. In remanding the case for further considerations, the court held that the lower court should apply the “discovery rule” to determine when the six-year limitation period began to run. Under the discovery rule, the limitation period does not start until the injured party discovers, or by exercise of reasonable diligence and intelligence should have discovered the whereabouts of the missing painting.

The discovery rule is roughly consistent with our conclusions above. According to the rule, ownership rights will not be diminished by the passage of time so long as an

<sup>32</sup> Since owners (except for the artists) are themselves buyers, there is a point where statutes of limitation are in their interests as well. This idea echoes in the field of copyright law, where even copyright owners would, *ex ante*, prefer a finite to an infinite period of copyright protection. See [Landes and Posner \(1989\)](#).



owner takes reasonable efforts to search for the work. Until the identity of the current possessor is known, the marginal productivity of search expenditures is likely low; thus only the bare minimum of expenditures should be required of an owner. Today, these expenditures would involve reporting the missing painting to the Art Loss Register and (possibly) checking recent art publications, such as a catalogue raisonne, that contain information on the artist.

An alternative approach, known as the “demand and refusal” rule, avoids the administrative difficulty of determining the reasonableness of an owner’s search efforts. Under this rule, the statute of limitations does not begin to run until the current possessor rejects the original owner’s demand for a work’s return. Of course, this cannot happen until the original owner discovers the location of the work. *DeWeerth v. Baldinger* (836 F.2d 103 (2nd Cir. 1987)), which involved a dispute over a Monet landscape stolen from a German castle in 1945, illustrates the rule’s application. Baldinger acquired the painting in 1957 from the Wildenstein Gallery, which had previously purchased it from a Swiss dealer. The painting hung in Baldinger’s Park Avenue apartment for twenty-five years before DeWeerth’s nephew discovered its location from an entry in a catalogue raisonne. DeWeerth demanded the return of the painting even though New York had a three year statute of limitation. The Second Circuit applied a demand and refusal rule as New York law, distinguishing it from New Jersey’s discovery rule in the O’Keeffe case. The court however imposed on owners a duty to employ reasonable diligence in searching for the work, in addition to a duty not to unreasonably delay one’s demand for the work once it is found. It found against the plaintiff on the grounds that her search efforts (which were non-existent between 1957 until 1981) were insufficient. In *Solomon R. Guggenheim Foundation v. Lubell* (569 N.E.2d 426 (N.Y. 1991)), the highest state appellate court in New York explicitly rejected the duty of reasonable diligence expressed in *DeWeerth*. By its own account, the Guggenheim museum had lost track of a Chagall gouache “sometime in the late 1960s” (Id. at 427). After confirming it missing between 1969 and 1970, the Guggenheim notified no museums, galleries, artistic or law enforcement organizations of its loss. In 1985, the museum learned of the work’s location and demanded its return. When the then possessor refused to turn it over, the museum sued and won. The court noted both the administrative complexity of a requirement of reasonable diligence and as well the ambiguity in what amount of diligence would be reasonable.

The *Guggenheim* case illustrates another interesting point. The museum argued that publicizing the work’s disappearance would have “expose[d] gaps in security” (perhaps leading to more thefts) and “pushe[d] . . . the missing painting further underground”. The Guggenheim’s position hints of a collective action problem. All art owners would prefer a world in which stolen works were well-publicized, but no owner has the private incentive to publicize its own loss. One way to address this coordination problem is to lower the costs and raise the effectiveness of publicizing lost works. This is accomplished by lost art registries, such as the Art Loss Register (ALR), mentioned above.

Disputes over the ownership of Holocaust Art also raise a rich set of issues. One well-publicized dispute concerned two Egon Schiele paintings that had been “Aryanized” by

the Nazis in 1938 (*People v. Museum of Modern Art (In re Grand Jury Subpoena Duces Tecum*, 93 N.Y.2d 729 (N.Y. 1999)); *United States v. Portrait of Wally*, 2002 U.S. Dist. LEXIS 6445 (S.D.N.Y. 2002)). The State of New York moved to seize the works while they were on loan to the Museum of Modern Art. MoMA joined the party that had lent the works (the Leopold Foundation of Vienna) in resisting the seizure, arguing that the threat of such state action would discourage foreign collectors from loaning their works to the museum – thus preventing the museum from “mak[ing] works of art from all over the world available to New Yorkers” (*People v. Museum of Modern Art*, 1999, p. 742). The museum prevailed in the state case but was forced to forfeit one of the works in later civil proceedings initiated by the United States.

Fifty-seven delegations participated in the Washington Conference on Holocaust-Era Assets held at the U.S. State Department November 30–December 3, 1998 [[United States Department of State \(1998\)](#)]. The parties agreed upon eleven non-binding principles to assist in resolving issues relating to Aryanized art – including commitments to identify and publicize stolen works, to open archives to researchers, to establish registries, and to expedite the fair and just resolution of disputes. In “establishing that a work had been confiscated by the Nazis”, the delegations promised to consider the “unavoidable gaps or ambiguities in the provenance in light of . . . the circumstances of the Holocaust era” (*Ibid.*). Nations were further encouraged to develop alternative dispute resolution mechanisms, including mediation and arbitration. Today, most art museums in the United States post notice on their websites of works acquired after World War II that have gaps in their provenances in the 1930s and 1940s. The postings should aid owners and their heirs in identifying stolen works and facilitate their return.

## **6. Art market: Contract and authenticity issues**

### *6.1. Introduction*

The market attaches great value to the identity of the author of a work of fine art. If doubt develops that a work attributed to Rembrandt is not authentic, its market value will plunge. Indeed, if the work turns out to be a modern forgery, it may be worthless. Conversely, if scholars now believe that Rembrandt painted a work formerly attributed to a lesser artist living at the time of Rembrandt, its value will skyrocket. In each case, the physical object stays the same, yet changes in attribution substantially affect the value. One observes a similar phenomenon in the market for collectibles. Here it is not authorship but rather a work’s provenance or history that matters. For example, the dress worn by Marilyn Monroe when she sang “Happy Birthday” to President Kennedy at Madison Square Garden recently sold for \$1.15 million at auction. But the “same” dress without the Monroe/Kennedy association would sell for very little at a resale shop. To take another example, an imitation pearl necklace belonging to Jackie Kennedy Onassis sold for \$211,500 at Sotheby’s, yet an identical necklace could be bought for several

hundred dollars at a jewelry store.<sup>33</sup> These examples illustrate that the market value of a work of art or collectible consists of two components: the value of the physical object itself plus the value of intangibles embodied in the object, such as the work's author and provenance.<sup>34</sup> And it is the value of the intangibles rather than the physical object that primarily determines the value of the work. Understanding why this is so is beyond the scope of the paper.<sup>35</sup> Here we consider legal issues that arise in connection with authenticity disputes.

At the outset, it is useful to distinguish between two categories of authenticity disputes. One involves fraud in which the offending party (e.g., an artist or dealer) attempts to pass off fake or copy as an original work.<sup>36</sup> Successful art fraud exploits imperfect information and the premiums the market assigns to original works. Like theft, such fraud is a socially harmful activity because its cost (the time and effort in perpetrating the fraud) exceeds its benefit (the value of the fake); it is thus subject to criminal and civil penalties. The second category involves genuine uncertainty because of imperfect information. Here the question of authorship can only be answered with a probability, rather than a certainty, because of incomplete documentation on the history of the work. Fraud is not involved but questions arise concerning the interpretation of the contract between the parties and the costs and benefits of acquiring authorship information prior to the sale of the work. We focus below on this type of dispute.

## 6.2. *Authenticity and the law*

Parties to authenticity disputes employ a variety of methods to attack or defend the authorship of works of visual art. At one end of the spectrum is the "high-handed, note-the-weak-contours-on-the-vase-at-left ... old-fashioned connoisseurship" [Gopnik (1997, p. 36)]. At the other are highly technical scientific techniques such as X-radiography, pigment analysis, and canvas thread counting [see Schwartz (1995)].<sup>37</sup>

<sup>33</sup> Provenance also matters for works of art (even if authenticity is not an issue) because of the absence of clear standards for what constitutes good art. If an important collector has owned the work or the work has been exhibited at prestigious museums, a buyer will have greater confidence in the work's quality, which in turn will increase the value of the work.

<sup>34</sup> Intangible factors also include the intellectual property rights of copyright and trademark. Because these rights are often more valuable than the physical object itself, the value of an artwork will be substantially affected by whether these rights are sold along with the object or retained by the author. In this paper, we put to one side the intangible factors of intellectual property rights and instead focus on authorship and provenance questions.

<sup>35</sup> William Grampp (1989, Chapter 10) systematically addresses this question in *Pricing the Priceless*. See also Becker, Landes and Murphy (2000) (applying the notion of consumption externalities to explain why originals sell for much more than copies and why authorship and provenance changes can lead to large differences in price).

<sup>36</sup> Fakes are often not copies. A skillful forger may create a fake by creating a work that attempts to fill in a gap in the original artist's work. To be sure, the forger attempts to pass off the fake as a work by the original artist but the work itself need not be a copy nor closely related to a work executed by the original artist.

<sup>37</sup> X-radiography sends x-rays through a painting and onto a film to reveal overpainted areas of the work not visible to the naked eye. In autoradiography, the painting itself is made lightly radioactive. Different

Lying somewhere between the two, the provenance of a work will often weigh heavily on authenticity assessments. Scientific examination can establish the age and condition of a work; iconographic and stylistic comparison can place it in a national or regional school; art-historical scholarship can relate it to a given studio or master; connoisseurship can assign to it a certain level of quality by defined standards; the history of reception can tell us which attributions are controversial and which not; and provenance documentation (in particular, a clear ownership history without gaps that traces the work back to the artist's studio) reduces the likelihood that the work is a forgery or incorrectly attributed to a great artist.

In *Greenberg Gallery, Inc. v. Bauman* (817 F.Supp. 167 (D.D.C. 1993)), a group of four galleries claimed that a mobile they had jointly purchased for \$500,000 was not the Andrew Calder work "Rio Nero" they had thought it to be. They sought rescission of the sale on theories of fraud, breach of express warranty, and mutual mistake of fact. The previous owner had inherited the work from her father who had acquired it from the Perls Galleries who had represented Calder. The work's provenance, therefore, was impeccable.

The plaintiff's expert Klaus Perls, who had seen and photographed the actual Rio Nero twenty-three years before and had sold the work to the defendant's father, asserted that the work was "an exact copy" of the original (Id. at 171).<sup>38</sup> But comparing the questioned work to the archival photo, Perls argued that "the relative length of the rods was not the same" as in the original. Perls's reputation as a Calder expert was such that, according to other witnesses, his denial of the work's authenticity would significantly reduce or even destroy the work's value in the art market (Id. at 174, n.8). Defendants presented expert testimony questioning Perls's authentication methodology. Archival photos were unreliable, they argued, "since the lighting, camera angle, distance, and movement affect the shape of the work" (Id. at 172). Moreover, the defendant argued (plausibly) that any problems with the work were the result of damage in handling and lack of expertise in setting up the work – both brought about by the plaintiff's actions. The court found that the plaintiffs failed to prove that the work was not the original Rio Nero; in part, because it doubted the existence of another "'authentic' Rio Nero (whereabouts unknown)" (Id. at 175). In short, the record, in spite of Perls's testimony, was "too inconclusive to support [plaintiffs'] preponderance burden" (Id. at 175).

The case raises the difficult question of whether a court should substitute its judgment for that of the market. The *Greenberg* court was clear: "This is not the market . . . but

combinations of elements emit different levels of radiation onto films that are laid upon the work. In infrared reflectography, a film is bathed in light just beyond the range normally visible to the eye. The work is then videotaped with a camera designed to pick up carbon levels in the underdrawing. Pigment analysis samples a cross-section of a painting. While it captures all the materials in the painting in the proper sequence, it is destructive to the work and can only sample a small area. Finally, dendrochronology and canvas thread counting attempt to date the material on which a painting was created.

<sup>38</sup> Presumably, plaintiff's position was that the true Rio Nero had been sold to defendant's father, after which point the supposed copy was created.

a court of law, in which the trier of fact must make a decision based upon a preponderance of the evidence” (Id. at 174). The court thus treated the issue of authenticity as susceptible to a “yes” or “no” answer, rather than as an analysis yielding results along a continuum. The latter approach would focus a court’s inquiry on whether doubts about authorship caused a material decline in the work’s market value – analogous to how commercial contracts are scrutinized; whether the seller had failed to disclose material information about authenticity that he possessed (or should have reasonably possessed) at the time of sale; and whether the buyer’s behavior may have caused the decline in market value by altering or damaging the work. In *Greenberg*, the decline in the mobile’s value occurred only after the plaintiffs – knowledgeable art dealers – had accepted delivery and then shipped the work to different locations, which, as noted, may have damaged the work and led to doubts about its authenticity. In the absence of fraud, such inspection and definitive acceptance is usually sufficient to limit the seller’s liability. Thus the court may have reached the right decision for the wrong reasons.<sup>39</sup>

Consider an extreme version of the Calder case: both buyer and seller initially believe that a work is by X; yet after the sale it becomes clear that the work is not by X. Suppose further that the change in authorship leads to a substantial change in the value of the work (either up, if the work turns out to be by a more famous artist than X, or down, if the reverse is true). Should the buyer or seller bear the risk of mutual mistake? The standard law and economics response allocates the risk to the party who is able to bear the risk, or acquire information about it, most cheaply. This implies that in art transactions between sophisticated dealers and unsophisticated buyers, the seller should bear any authenticity risk that leads to a material decrease (or increase) in value, unless in the bargain the seller makes clear that there is some doubt as to whom the artist is. In the *Greenberg* case, where sophisticated parties were on both sides, the optimal rule may be the opposite because that would avoid expensive litigation in all cases except where the seller drafts an express warranty. Consider a different example. Suppose an unsophisticated person discovers a painting in his attic and sells it for \$60 to an art dealer who believes it is by a famous artist. The work turns out to be worth \$1 million. Can the seller claim he was duped and rescind the sale? The answer is no and this makes good economic sense. First, if a court held the sale invalid, it would mean that buyers would have to disclose private information about the work’s value. Such disclosure would probably defeat the transaction and, in general, discourage knowledgeable persons from searching for valuable art. Second, a seller can protect himself from selling a work that later turns out to be very valuable by hiring an appraiser. And, if the appraiser undervalues the work in his review, the former owner may have an action in negligence against him.

<sup>39</sup> Fraud requires knowing or intentional deceit (a so-called scienter requirement). Since the *Greenberg* court found insufficient evidence to doubt the work’s authenticity, it never reached the question of whether defendants had knowingly or intentionally misled the plaintiffs. We doubt that the record would have supported such a conclusion.

Rather than leave the rule-setting entirely to the courts, auction houses draft extensive sales documents to define the circumstances when authenticity disputes will justify rescission. In *De Balkany v. Christie Manson and Woods* (Queen's Bench Division, 11 January 1995), the buyer of what was described as an Egon Schiele painting in the auction catalogue claimed that the work was a forgery within the meaning of Christie's terms and conditions of sale. Although Schiele originally painted the work, the buyer claimed it was no longer a Schiele because 94% of the painting had been painted over by an unknown party after the original paint had flaked off. In response, Christie's argued that the work was an authentic Schiele and the overpainting was a question of the condition of the work, which the contract did not guarantee. The Christie's contract provided that works found to be forgeries could be returned within five years of a sale unless

(1) the catalogue description at the date of the auction was in accordance with the then generally accepted opinion of scholars . . . ; or (2) it can be established that the Lot is a Forgery only by means of a scientific process not generally accepted for use until after publication of the catalogue . . . or by means of a process which at the date of the auction was unreasonably expensive or impractical or likely to have caused damage to the Lot; (Id.).

The court found that the painting was a forgery (not, as Christie's claimed, an authentic Schiele in bad condition) and that Christie's could not avail itself of exceptions (1) or (2) above. Moreover, since the restorer had added Schiele's initials to the lower corners of the work, the court held that the overpainting was "made with an intention to deceive" and was not accurately described in the auction catalog as being "by" Egon Schiele.

The court found that Christie's could not avail itself of the exception clause above for two reasons. First, Christie's had not relied upon the "generally accepted opinion of scholars" in attributing the work to Schiele. Second, it had failed to investigate reasonably the extent of the work's overpainting, and any careful viewing of the painting would have led Christie's to describe the work as a painting "attributed to" Schiele, rather than painted by him (Id.). The court also noted that Christie's actions would have supported a tort claim for negligence since the painting's value had declined over 90% from its purchase price, an injury the buyer would not have suffered were it not for Christie's failure to take reasonable precautions in identifying the work.

Courts also become embroiled in authenticity disputes when an owner of a work claims to have been injured by another party's denial of the work's authenticity. In *Hahn v. Duveen* (234 N.Y.S. 185 (N.Y.Sup. 1929)), the plaintiff sued the well-known art dealer Sir Joseph Duveen for stating that a painting owned by the plaintiff was not a true da Vinci, but rather a copy of one displayed in the Louvre. In effect, the plaintiff claimed that Duveen had disparaged its product by making a false claim about its authenticity. To succeed in its lawsuit, the plaintiff had to show that Duveen's comment was false, misleading, and material, and that it had caused the work's value to decline. In spite of

a parade of connoisseurs (including Bernard Berenson) in near unanimous support of Duveen's position, Duveen lost the initial case and settled out of court before a retrial.

## 7. Estate issues

### 7.1. Introduction

This section examines two problems concerning the judicial treatment of art-rich estates. The first considers how courts value art-rich estates for tax and related purposes; the second concerns when trustees of charitable trusts comprising valuable art (e.g., the Barnes Foundation) should be allowed to depart from terms of the trust when it becomes impracticable, impossible or illegal to carry out these terms.

### 7.2. Estate valuation

When Andy Warhol died unexpectedly in 1987, his estate included a substantial body of his works – around 4000 paintings, 5000 drawings, 19,000 prints and 66,000 photographs. In litigation that followed, a court was asked to assess the monetary or “fair market” value of these works (*In re Determination of Legal Fees Payable by the Estate of Warhol*, 1994 N.Y. Misc. LEXIS 687 (N.Y. Misc. 1994)).<sup>40</sup> The court followed what has become a standard methodology – by first estimating the market prices of the individual items, usually from data on recent auction and private sales, and then by applying a “blockage discount” to the individual items or groups of items to arrive at an overall value. The rationale for the blockage discount is the belief that a large number of an artist's works could not be sold immediately without significantly depressing the market for the works. In determining the size of the blockage discount, courts consider such factors as the type of work; the number of works in the estate relative to the number sold in recent years; the artist's reputation; the likely future appreciation in prices; and the time and cost necessary to liquidate the estate in small lots. So, in the Warhol litigation, the court initially valued Warhol's art at \$506 million and then applied blockage discounts ranging from 20 to 35 percent to reach a final value of \$391 million (*Id.*). In another well-known case, the court valued the estate of the artist Georgia O'Keeffe at \$36 million after applying an average blockage discount of 50 percent to over 400 works initially valued at \$73 million (*Estate of O'Keeffe v. C.I.R.*, T.C. Memo. 1992-210 (U.S. Tax Ct. 1992)). And in the case of the sculptor David Smith, the individual items were initially valued at \$4.3 million and \$2.7 million after discounting (*Estate of Smith v. Commissioner*, 57 T.C. 650 (U.S. Tax Ct. 1972)).

<sup>40</sup> The case involved a dispute between the lawyer (Hayes) for the executor of the estate and the Andy Warhol Foundation for the Visual Arts (the main beneficiary of the estate). Hayes's contract called for a fee based on roughly 2 percent of the value of the estate. Hayes valued the estate at over \$700 million while the Foundation valued the estate at about \$100 million.

Valuing the estate of an artist who has retained a significant number of his own works raises interesting economic questions concerning the pricing of a durable good. Like other durable goods, art has a long life, yields a (non-monetary) return each period, requires carrying costs, and often has significant resale value. Art is also a risky asset. Not only are there wide swings in overall art prices but also the risks are even greater for individual artists who may be in fashion one period and out the next.

These risks justify employing a heightened discount rate to calculate the value today of the estimated proceeds from future sales. Why must future sales be estimated? Because, as the reasoning goes, a sale of all of an artist's works today will saturate the market for the works and will not maximize the value of the estate. A quick sale will not allow the estate time to promote the late artist's work, to locate new buyers, or to persuade collectors to increase their holdings of a single artist. Courts thus assume that value-maximizing estate will liquidate their holdings over a number of years. The problem is that by applying a heightened discount rate (in excess of the works' estimated rate of appreciation) to these future sales, the present value of future sales will be less than the estimated value of the works today – thus, the “blockage discount”.

This justification, however, is at odds with economic theory. In a rational expectations model, issues of the expected appreciation of art, the risk of holding an undiversified portfolio of art, projections on the number of future sales, aesthetic returns from owning art and discount rates are tangential to estimating the value of an artist's estate. As it turns out, the problem is much simpler than might appear. All one requires for valuation is current prices on actual sales, even assuming that it takes many years to liquidate the holdings in the estate.<sup>41</sup> Current prices capture all relevant information about future prices, risk and discount rates.

First consider the value of a single work, say a Warhol painting. Let  $P_0$  denote the monetary value of the work today. It follows that in equilibrium,

$$P_0 = P_n e^{-rn} \tag{1}$$

where  $P_n$  is the price of the work in  $n$  years, and  $r$  denotes the market discount rate. The discount rate is a composite of several factors as in

$$r = i + m - a \tag{2}$$

where  $i$  equals the one-period risk adjusted discount rate,  $m$  is the insurance, storage and other carrying costs (expressed as a percentage of the value of the work), and  $a$  is the aesthetic dividend per period from holding the work. Observe that Equation (1) is an equilibrium condition.<sup>42</sup> It is the outcome of individuals bidding for works in the art

<sup>41</sup> We are indebted to Casey Mulligan and William Grampp for helpful discussions on this point.

<sup>42</sup> If individuals receive different benefits from owning art, then the individual with the highest net benefit determines  $P_0$  in Equation (1) in the case of a single unique work. If we switch to valuing a large number of equivalent works that are fixed in supply, the net benefit of the marginal individual (the party just induced to buy the work) determines  $P_0$ .



market today and in the future. In equilibrium, the expected rate of growth of prices will equal the discount rate  $r$ . To see this, suppose  $P_0$  is less than  $P_n e^{-rn}$  say because  $P_0$  is expected to grow at a faster rate than  $r$ . Then buyers will bid up the current price until Equation (1) holds. Alternatively, if today's price exceeds the discounted value of the future resale price, today's owner will attempt to sell the work driving its price down until Equation (1) holds.

Observe that if art is a very risky investment,  $i$  and hence  $r$  will tend to be relatively high. The expected equilibrium rate of growth of art will also be high; and, taking as given the expected future price of the work, today's price will be lower the greater the risk of holding art. In short, today's price incorporates information about expected future prices and the discount rate. To continue our example, imagine that Warhol's estate contains a large number of his paintings. Assume further that we can convert these works into equivalent quality units. That is, let the price of a Warhol painting depend on various characteristics such as size, subject matter, date of execution, exhibition history, type of materials, etc. Suppose one can estimate a hedonic index relating these characteristics to prices. We can then take a group of diverse Warhol pictures and convert them into a stock of equivalent Warhol paintings. Let  $K$  denote the physical stock of these homogeneous paintings (or, equivalently, the stock of a durable good) and assume the estate plans to sell some paintings today, some next year and so on until no paintings are left in the estate.

Without any loss of generality assume that the estate plans to sell its entire stock of paintings within  $n$  years ( $n$  may be a very large number but knowing  $n$  is not essential to the solution below). Assume further that  $K$  is large relative to the total stock of Warhol paintings (which must be greater than  $K$  since Warhol sold many paintings before his death). The assumption that the estate holds a relatively large number of works raises two related questions. One is whether the estate's decision to vary the number of paintings it sells in any year will affect price. The other is whether the estate may face credibility or commitment problems because prospective buyers might be worried that if the estate sells a few paintings today at a high price, it will then dump the rest at lower prices in the future. As it turns out, none of these details matter to the problem at hand.

In valuing the estate, we assume its objective is to maximize the present value  $V$  of its sales of works over time. This holds whether the beneficiaries of the estate are individuals or a charitable foundation created by the artist. Let

$$V = \int_0^n k(t)P(t, k(t), Z)e^{-rt} dt \quad (3)$$

where  $k(t)$  denotes the number of works sold each period and  $P(t, k(t), Z)$  denotes the expected future price in period  $t$  which, in turn, may depend on the number of works sold in that period plus other factors such as the total stock of Warhol works and the number of works by other artists that are substitutable for Warhol's paintings. These factors are exogenous to the estate and represented by the variable  $Z$ .

Let  $V^*$  denote the maximum value of  $V$ . To achieve  $V^*$ , the estate might dispose of a few paintings each year, sell its entire holdings this year or hold off selling any works for

a number of years.<sup>43</sup> Again it turns out that we need not consider the details of how the estate disposes of its works in order to estimate  $V^*$ . Recall the equilibrium condition in Equation (1) relating the current to future price of a painting. This condition must hold for all  $t$  periods. That is, today's price must equal the discounted price one year from today, two years from today and so forth. Substituting Equation (1) into Equation (3) and noting that

$$K = \int k(t) dt$$

yields

$$V^* = P_0 \int_0^n k(t) dt = P_0 K. \quad (4)$$

In other words, the value of the works in the estate is simply equal to today's price times the total stock of equivalent works. Since in equilibrium future prices are expected to grow at a rate of  $r$  per period, we only need to observe a small amount of art auctioned off today ( $P_0$ ) in order to value the proceeds the estate will receive from future sales. This is a remarkable result because we do not need to know the details of how the estate disposes of its holdings in order to estimate its value.

We note several additional points.

1. In an important sense, nothing fundamentally changes when the artist dies. To be sure, no new works will be created and that may alter current and future prices. But the same number of works that existed the day before the artist's death exists after his death. Moreover, the concentration of holdings is the same. Before his death, the artist held  $K$ . Now the estate holds  $K$ . Hence any market power the estate might be able to exercise because  $K$  is relatively large could have been exercised by the artist himself.<sup>44</sup> Thus, as a first approximation,  $V^*$  is the same whether the artist or his estate holds  $K$ . We say "as a first approximation" because supply is fixed after the artist's death so prices may increase. A possible offset, however, is that the dealers may have less incentive to promote the artist's works after his death. Prices may also fall because promotional efforts by the artist himself will end. These are relatively minor points and, in any event, will be captured (in an expected sense) by current prices following the artist's death.

<sup>43</sup> The typical argument advanced against selling off works immediately is that such a "fire sale" would significantly depress prices. But the problem turns out to be more subtle. On the one hand, inter-temporal substitution may limit or eliminate the negative effects of selling a large number of works in any period. On the other, a "fire sale" by the estate may signal that the estate has private information that Warhol's standing in the art market is likely to diminish in the future and this, in turn, may lead collectors to revise downward their estimates of Warhol's stature. And since art prices are largely determined by demand (since supply is fixed), small shifts in demand may induce large changes in prices. For the moment we put off consideration of these issues.

<sup>44</sup> One difference, of course, is that the living artist also controls the rate at which the stock expands whereas the estate holds a fixed stock.

The essential point is that the artist's death does not fundamentally alter the valuation question.

2. We have been able to determine the value of the estate without making any assumption about the number of works that the estate will sell each period. All we require is the assumption that the estate's objective is to maximize the value of its assets. From this it follows that it will dispose of works in a way to maintain the equality between current prices and the present value of future sales. Given this outcome, we can use today's prices from a small number of sales (relative to  $K$ ) to value a large number of future sales.

3. Imagine that the estate must sell a large number of works today to satisfy tax obligations. Does this "forced sale" lower the value of the estate? To see that it doesn't, consider what would happen if selling a large number of works depressed current prices. Current prices would be less than the discounted value of future prices. Hence a buyer (or buyers) would step in and bid up today's price to restore the equilibrium between current and future prices. The buyer would probably be a dealer who would acquire the works to sell in the future.<sup>45</sup>

4. As mentioned earlier, if the estate sells a large number of works immediately, this may signal their belief that an artist's works are overvalued and future prices are likely to be lower than Equation (1) implies. Assuming that other participants in the art market believe the estate has some informational advantage evaluating the artist's future reputation, the net effect of the estate dumping works on the market will be that both present and expected future prices will fall. The equilibrium relationship between current and future prices will still be maintained though at lower prices. The estate may try to conceal these sales by selling through third parties. But this is not likely to be successful. Combining points (3) and (4) implies that dumping a large number of works on the market lowers price because demand shifts downward with a fixed supply. Note also that price is not declining because of a movement along the demand curve for the sale of works changes the identity of owners but not total supply.

### 7.3. Estate management: The doctrine of *cy pres* and administrative deviation

Albert C. Barnes, a self-made chemist who died in 1951, assembled one of the world's greatest art collections. The collection, now the property of a charitable trust (the Barnes Foundation), contains over 700 Impressionist works, including 181 Renoirs, 69 Cezannes, 60 Matisses, 44 Picassos, 18 Rousseaus and 14 Modiglianis. Barnes reportedly detested the elitism of the museum culture in high-society Philadelphia. Thus

<sup>45</sup> This raises the issue of whether the appropriate prices in Equations (1)–(4) are dealer or wholesale (often equated with auction) prices or a blend of the two. This turns out not to be a problem provided one is consistent. In a competitive art market, the difference between retail and wholesale reflects the cost of selling goods. These costs may be substantial because of the time and effort required to persuade buyers to purchase a work. Whether one uses wholesale or dealer prices depend on whether one wants to know the net or gross value of the estate. In the former case, one uses wholesale (auction) prices; in the latter, retail prices.

in setting up the Foundation, Barnes stipulated that the Foundation was to be an educational institution and not a public museum; public access was to be allowed only on Saturdays between September and June; no paintings were to be removed from the gallery nor moved from where he had placed them (the particular arrangement was part of his philosophy of art appreciation); no entrance fees could be charged; no loans or sales of paintings were to be allowed; and all of the Foundation's funds were to be invested in low-yielding Government securities. Most importantly, Barnes specifically provided that in no event should these terms be modified. If the Foundation did not have the resources to continue, Barnes specified that the art works were to be donated to a group of museums and other institutions [Abbinante (1997)].

Notwithstanding Barnes's testamentary clarity, as the financial situation of the Foundation has become increasingly strained in recent years, courts have sanctioned numerous deviations from these specific terms under the legal doctrines of *cy pres* and "administrative deviation". These doctrines are typically applied when the terms of a charitable trust become impossible or impractical to carry out: administrative deviation usually concerns minor changes regarding how the trust is managed (e.g., number of trustees, tuition charge, or mortgaging of property) whereas *cy pres* concerns more major changes necessary to carry out the general purpose of charitable foundation (e.g., redirecting foundation funds towards basic medical research when the foundation was set up to provide research funds to combat a particular disease that has been eliminated).

In 1995, the Montgomery County Court of Common Pleas and the Pennsylvania Superior Court granted the trustees of the Foundation greater flexibility in the management of Foundation funds. The courts allowed the Foundation to open the collection to public viewing three-and-a-half days a week, to charge an admission fee of five dollars, and to use the gallery for fundraising purposes. Perhaps most remarkably, the courts allowed the Foundation to conduct a world tour of the collection to raise money for the improvement of the Foundation's facilities. All of these modifications, against Barnes' specified intent for the collection, were defended as administrative in nature.

More recently, several foundations have offered to donate \$150 million to the Barnes Foundation if it moves the entire collection to a new facility in the museum district in Downtown Philadelphia. The Foundation insisted and the court subsequently agreed that such funds were necessary for the Foundation's continued viability, and that moving the collection was consistent with Barnes' desire that the collection be displayed so as to promote educational goals.

What is puzzling about the recent Barnes decision is that Barnes explicitly provided for the contingency that if the Foundation failed, its works were to be distributed to various museums. Rather than disposing of the collection as Barnes had provided, the courts have supported the Foundation in altering the terms of the charitable trust to bring about the outcome that Barnes had most explicitly sought to avoid – the creation of a new museum in Philadelphia to house his art. There is an economic downside when a court departs from the text of a will or a trust indenture. It undermines the incentive for charitable giving and the establishment of charitable foundations to promote the educational goals of the donor. To be sure, it might appear socially beneficial to depart from

the terms of the trust when circumstances arise that were not anticipated by the donor. But this comes at a price – namely, the reduced incentive to set up a charitable foundation in the first place. In some cases, the benefit from altering the trust may outweigh the cost, particularly, if the donor has not made any provision for dealing with these unanticipated changes. Then there is a strong argument that these changes would have to be approved by the donor if “transaction costs” had not made it impossible to discuss these modifications with him. But in the Barnes case, the argument for modifying the trust is much weaker; Barnes considered and provided for the contingency that the Foundation would no longer have the resources to continue its educational mission.

We wonder whether an alternative approach may better serve the interests of present and future generations. The value today of the right to control what happens 50 or 100 years in the future is likely a small fraction of purchase price for art. It may also be a relatively small fraction of the incentive to engage in charitable giving. Were the law to offer the right to control testamentary estates for only 50 or 100 years (rather than in perpetuity), and were courts to scrupulously respect those conditions for that term, testators and courts would have greater clarity regarding the proper time for pragmatism. Some courts may still be persuaded to allow significant deviations over the course of that term; but no longer would slippery-slope arguments for administrative deviation – i.e., that deviation is appropriate since the current situation may eventually prove unsustainable – hold water. Improved certainty may encourage charitable giving beyond today’s baseline.

## References

- Abbinante, C. (1997). “Protecting ‘donor intent’ in charitable foundations: Wayward trusteeship and the Barnes Foundation”. *University of Pennsylvania Law Review* 145, 665–709.
- Ames, E.K. (1993). “Beyond *Rogers v. Koons*: A fair use standard for appropriation”. *Columbia Law Review* 93, 1473–1526.
- Becker, G.S., Landes, W.M., Murphy, K.M. (2000). “The social market for the great masters and other collectibles”. In: Becker, G.S., Murphy, K.M. (Eds.), *Social Economics: Market Behavior in a Social Environment*. Belknap Press, Cambridge, pp. 74–83.
- Chiang, T.J., Posner, R.A. (2003). “Censorship and free speech: First amendment issues in the visual arts”. Unpublished manuscript.
- Ginsburg, J.C. (1995). “Exploiting the artist’s commercial identity: The merchandizing of art images”. *Columbia-VLA Journal of Law & the Arts* 19, 1–22.
- Ginsburgh, V. (1996). “On the economic consequences of resale rights on art”. Unpublished manuscript.
- Ginsburgh, V., (2005). “The economic consequences of *Droit de Suite* in the European Union”.
- Gopnik, A. (1997). “Doubting Vincent”. *New Yorker* (July 28), pp. 36–37.
- Grampp, W.D. (1989). *Pricing the Priceless: Art, Artists, and Economics*. Basic Books, New York.
- Hansmann, H., Santilli, M. (1997). “Authors’ and artists’ moral rights: A comparative legal and economic analysis”. *Journal of Legal Studies* 26, 95–143.
- King, S. (1997). “Warhol estate sued over Jackie photo”. *Art in America* 85 (2), 27.
- Landes, W.M. (2000). “Copyright, borrowed images and appropriation art: An economic approach”. *George Mason Law Review* 9, 1–24.
- Landes, W.M. (2001). “What has the visual artist’s rights act of 1990 accomplished?” *Journal of Cultural Economics* 25, 283–306.

- Landes, W.M. (2003). "Copyright". In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Northampton, pp. 132–142.
- Landes, W.M., Posner, R.A. (1989). "An economic analysis of copyright law". *Journal of Legal Studies* 18, 325–363.
- Landes, W.M., Posner, R.A. (1996). "The economics of legal disputes over the ownership of works of art and other collectibles". In: Ginsburgh, V., Menger, P.-M. (Eds.), *Essays in the Economics of the Arts*. Elsevier Science, Amsterdam.
- Landes, W.M., Posner, R.A. (2003). *The Economic Structure of Intellectual Property Law*. Belknap Press, Cambridge.
- Levin, K. (1988). *Beyond Modernism: Essays on Art from the '70s and '80s*. Harper & Row, New York.
- Levmore, S. (1987). "Variety and uniformity in the treatment of the good-faith purchaser". *Journal of Legal Studies* 16, 43–65.
- Lichtman, D.G. (2003). "Copyright as a rule of evidence". *Duke Law Journal* 52, 683–743.
- McCartney, B.T. (1998). "'Creepings' and 'glimmerings' of the moral rights of artists in American copyright law". *UCLA Entertainment Law Review* 6, 35–72.
- Norman, G. (1996). "The power of borrowed images". *Art & Antiques* (March 1996), pp. 123–125.
- Robinson, W. (1983). "Art and the law: 'Moral rights' comes to New York". *Art in America* 71 (9), 9–13.
- Rubinstein, R. (1994). "Abstraction in a changing environment". *Art in America* 82 (10), 102–109.
- Schwartz, G. (1995). "Truth in labeling". *Art in America* 83 (12), 50–58.
- The Art Newspaper (March 1998). "How to be duly diligent".
- The Art Newspaper (January 2000). "Artists oppose resale rights".
- U.S. Copyright Office (1996). *Waiver of Moral Rights in Visual Artworks: Final Report of the Register of Copyrights*, March 1, 1996. U.S. Copyright Office, Washington.
- United States Department of State, *Washington Conference on Holocaust-Era Assets* (1998). *Washington Conference Principles on Nazi-Confiscated Art*. U.S. Department of State, Washington. [http://www.state.gov/www/regions/eur/981203\\_heac\\_art\\_princ.html](http://www.state.gov/www/regions/eur/981203_heac_art_princ.html)
- Wall Street Journal (April 6, 2000). "The art market: Artists' royalties", p. A20.

## COPIES OF ARTWORKS: THE CASE OF PAINTINGS AND PRINTS\*

FRANÇOISE BENHAMOU

*Université de Rouen and Centre d'Économie de la Sorbonne, Université de Paris I, France*

VICTOR GINSBURGH

*ECARES, Université Libre de Bruxelles and CORE, Université catholique de Louvain, Belgium*

### Contents

Abstract	254
Keywords	254
1. Introduction	255
2. Forgeries, copies, and reproductions. Definitions and boundaries	256
2.1. Copies and fakes	256
2.2. Copies and originals	257
2.3. Copies and reproductions	258
3. Artists and their patrons	259
3.1. The demand for copies	260
3.2. A change in the status of copies	261
4. Markets for prints	263
5. Markets for paintings	265
6. Copies and art historians. A renewal of interest?	270
7. Regulating the market for copies	273
7.1. An unstable regulatory framework in museums. The case of the Louvre	273
7.2. The legal definition of originality	275
7.3. Public and non-public domain	277
7.4. Copies as copyrightable goods?	278

\* The chapter freely draws on our previous paper Benhamou and Ginsburgh (2001), "Is there a market for copies", *Journal of Arts Management, Law and Society* 32, 37–55. We are grateful to M. Aarts, Me Cornette de St. Cyr, M. Cornu, D. Delamarre, M. Fouado-Otsuka, Me Lelurier, T. Lenain, H. Loyrette, M. Melot, O. Meslay, D. Schulmann, B. Steyaert, G. Touzenis, H. Verschuur and S. Weyers for very useful conversations and comments. Neil De Marchi's very careful comments on a preliminary version of the chapter at the Princeton Conference led to several changes. We also thank Alvin Huisman (2001) who collected data on auctions 1976–1999. The second author acknowledges financial support from Fonds National de la Recherche Scientifique.

*Handbook of the Economics of Art and Culture, Volume 1*

*Edited by Victor A. Ginsburgh and David Throsby*

*Copyright © 2006 Elsevier B.V. All rights reserved*

*DOI: 10.1016/S1574-0676(06)01008-8*

8. Concluding comments	279
Appendix A: Copies and originals sold at auction 1684–1999	280
A.1. Period 1684–1725	280
A.2. Period 1801–1810	280
A.3. Period 1890–1900	281
A.4. Period 1976–1999	281
References	281

## Abstract

In his essay on imitation in the arts, Adam Smith considers that the exact copy of an artwork always deserves less merit than the original. But the hierarchy between copies and originals has changed over time. So has the perception of copies by lawyers, philosophers, art historians and curators. The development of a market for copies is part of a wider contemporary questioning of the boundaries between originality and copy. We analyze whether and how the various actors in the art market (artists, collectors, lawyers, curators, art historians and philosophers) contribute to valuing and creating or, at times, to killing copies. Artists and collectors have never belittled copies. Art historians think that copies have an important role in preserving the memory of lost artworks, and in educating young artists, but nevertheless consider copies better left to the reserves of museums. Lawyers are ambivalent and judicial precedents bear testimony to the ambiguous legal status of copies. Contemporary art historians and art philosophers have influenced curators and museums to organize exhibitions that make use of copies, giving them a new life.

## Keywords

copies, fakes, multiples, prints, originality

*JEL classification:* D44, K30, Z11



“We use copies to certify originals, originals to certify copies, then we stand bewildered.”

Hillel Schwartz

## 1. Introduction

Research on copies is essentially focused on industrial activities (books, records, fashion, protection of patents) and on the incentives or disincentives to creativity resulting from copyright.<sup>1</sup> But copies are also linked to questions concerned with value, the allocation of property rights, and regulation, three central questions in economics.

In his essay on imitation in the arts, *Adam Smith (1795)* considers that the exact copy of an artwork always deserves less merit than the original.<sup>2</sup> But the hierarchy between copies and originals has changed over time. So has the perception of copies by lawyers, philosophers, art historians and curators. The observation of these changes can be used to analyze art tastes and practices. The development of a market for copies is part of a wider contemporary questioning the boundaries between originality and copy.

In this chapter, we analyze whether and how the various actors in the art market (artists, collectors, lawyers, curators, art historians and philosophers) contribute to valuing and creating or, at times, to killing copies. Artists and collectors have never belittled copies. Art historians think that copies have an important role in preserving the memory of lost artworks, and in educating young artists, but nevertheless consider copies better left to the reserves of museums. Lawyers are ambivalent and judicial precedents bear testimony to the ambiguous legal status of copies. Contemporary art historians and art philosophers have influenced curators and museums to organize exhibitions that make use of copies, giving them a new life.

In Section 2 we define copies, contrasting them with forgeries, and reproductions. Section 3 deals with the permanent role of copies over time. Sections 4 and 5 give some insights into the price of copies (relative to originals). Sections 6 and 7 are devoted to the changing views held by art historians, philosophers and law-makers. Section 8 concludes.

<sup>1</sup> See, for example, *Johnson (1985)*, *Grossman and Shapiro (1988a, 1988b)*.

<sup>2</sup> Smith is just stating a view that was common in his time (and even before).

## 2. Forgeries, copies, and reproductions. Definitions and boundaries

### 2.1. Copies and fakes

Copies or reproductions differ from fakes since only the latter are produced to deceive<sup>3</sup> but it is often difficult to detect whether a work was made with the intention to deceive. Hoving (1996, p. 32) considers fakes those thousands of Roman sculptures of the archaizing style, produced between 100 B.C. and 100 A.D., which are copies of Greek marbles from the sixth century B.C. The fact that even Greek sculptors produced copies long before the first century B.C. makes it highly doubtful that these were intentional forgeries.<sup>4</sup> Arnheim (1983) holds the view that we should be “grateful to get an idea of the lost Greek sculpture through Roman copies”.<sup>5</sup> After all, the celebrated *Venus of Milo* is also a Roman copy.<sup>6</sup>

Copies pay tribute to the original, recognize its value and draw their own value from it. Forgeries deny the aesthetic superiority of the original. Acknowledged copies are cheaper than originals. As long as they are not detected as such, fakes are as expensive as originals, and their number usually increases with the fame of the artist. Good forgers often have an extraordinary knowledge of the work of an artist and of what art history has to say about him, as was the case with Van Meegeren, the forger of Vermeer. According to Werness (1983, pp. 33–34), “the excitement with which [Van Meegeren’s] *Emmaus* was received was partly due to this very practice [of sifting] through the art historical literature. Again, the resemblance to Caravaggio’s painting of the same subject ‘proved’ Vermeer at least knew the painting and had possibly traveled to Italy”. In short, Van Meegeren “proved” what Vermeer experts wanted to hear. Forgers exploit assumed “holes” in the work of an artist and fill them.<sup>7</sup> Catalogues raisonnés are a tool for limiting the production of fakes. They began to appear during the nineteenth century, to “close” the oeuvre of an artist, at the very moment when the standard practice of copying in the same medium started to compete with other means such as photography [Castelnuovo (1987)].

<sup>3</sup> It is well known and documented that even masters seem to have produced forgeries. Lucky he who, today, is the owner of the “Roman” sleeping cupid carved by Michelangelo. See Hoving (1996, p. 55).

<sup>4</sup> Chamoux et al. (1973) give the example of the Thessalian prince Daochos who, in 335–330 B.C., preferred to order marble copies of existing bronze sculptures, rather than original marbles.

<sup>5</sup> See the Exhibition Catalog *La fascination de l’Antique, Rome 1700–1770*, Lyon: Musée de la Civilisation Gallo-Romaine, 1999.

<sup>6</sup> See Hol (2004).

<sup>7</sup> Recall also the enormous influence the Ossianic epics, attributed to the third-century Celtic bard Ossian, had on late eighteenth-century European literature. The poems were later found to be a forgery due to Macpherson, an eighteenth-century poet. This does not prevent the *Encyclopedia Britannica* to sum up Macpherson’s work as follows: “The varied sources of his work and its worthlessness as a transcript of actual Celtic poems do not alter the fact that he produced a work of art which did more than any single work to bring about the romantic movement in European, and especially in German literature. Herder and Goethe were among its profound admirers.” See Koestler (1989, pp. 402–403).

## 2.2. Copies and originals

There is an imperceptible transition from “original” to “copy”, and copies of paintings may be considered as satisfying substitutes for originals. Multiples cause even greater confusion than paintings. Man Ray photographs shot, developed, and printed by him in one or several copies are considered “vintage” photographs. Such works cease to be called “vintage” if Man Ray did not print the photographs himself. The case is much more ambiguous if the photographs are printed by another photographer, under Man Ray’s supervision, or without such supervision, but with his permission, or by Man Ray himself, but twenty years after the picture was shot. A print of a negative by Man Ray that the artist did not find interesting enough to print is not original. A photograph printed in 1990 by a collector who happened to possess a genuine negative by Man Ray is not a Man Ray. The one printed by the Centre Pompidou for an exhibition devoted to the artist is not an original either. Therefore, for photographs, the definition of originals and copies is rather arbitrary. The same applies to engravings and lithographs. These are obviously copies<sup>8</sup> because, like photographs, they are not unique.<sup>9</sup> Such prints are obviously “right” if the artist did the original engraving on the copper plate or the drawing on the stone, produced each copy in his atelier or on his own printing machine, and signed each copy. If all of these characteristics apply, but the print is not signed,<sup>10</sup> or if the artist had the prints done by someone else but verified and signed each one, the definition of originality becomes as ambiguous as in the case of photographs.<sup>11</sup> When the artist only did the drawing and had it copied on a copper plate by a good professional engraver, the engraving is no longer an original. The market considers that when the artist does nothing at all, but just signs the copies, as Salvador Dali did, the engraving is not an original.<sup>12</sup>

The Rodin museum in Paris is authorized to produce up to eight copies of every piece of plaster that was left by Rodin in his atelier after his death but that he did not wish to

<sup>8</sup> This vocabulary is not unanimously accepted. Griffiths, the keeper of the Department of Prints and Drawings at the British Museum, suggests that “the word [copy] is dangerously ambiguous . . . [and that] the term impression should be used of a print. The term copy in prints should be strictly reserved for a redrawing of an original by another hand. If done by the original designer, such a copy is referred to as a replica”. See Griffiths (1996, p. 139).

<sup>9</sup> For multiples, the definition of original and copy is complex, leading Melot (1985) to formulate “the curious theorem that for objects of art, multiples can also be unique”.

<sup>10</sup> Note that signing of prints was occasional during the eighteenth century and became more systematic after 1850 only. See Griffiths (1996, pp. 152–153). Signature is not considered necessary for prints in French decrees that define works of art. See Melot (1995).

<sup>11</sup> This is obviously a widely accepted practice for lithographs: “Once the drawing is finished the artist’s task is done. The rest is the province of the printer, whose operations are complicated enough to make it unusual for the artist to do his own printing”. See Griffiths (1996, p. 102).

<sup>12</sup> Imagination is without limits. Architect Richard Meier’s luxuriously designed New York apartments are now being sold as limited editions, numbered and signed by the designer. *The Art Newspaper* 146, April 2004, p. 2.

be cast in bronze. The museum sells them as genuine Rodins. The same is true for works by Arp that can be reproduced by the Fondation Arp in Rolandseck. Jewels drawn by Verdura fifty years ago are produced today and sold as genuine Verduras. Living artists may be very opposed to such actions. Donald Judd, for example, rejected a copy of one of his installations, though its owner had given permission to a Los Angeles gallery to erect a copy in an exhibition they were mounting of Judd's works.<sup>13</sup>

Posthumous editions of prints, modestly called "later states", are sold under the names of the artists who created the originals.<sup>14</sup> An unsigned poster is not an original work. But posters by celebrated artists are almost considered artwork. The Musée d'Ixelles in Brussels claims to be the only museum to possess the complete collection of Lautrec's posters and exhibits all of them in one of the best rooms of the museum. As we shall see later, the price of prints by Rembrandt himself is not more than three times larger than the price of posthumous prints, even from a transformed original plate.

These cases show that the definition is ambiguous, and essentially depends on norms and conventions. In some cases a copy is accepted as "right", even if it is copied long after the artist died. In other cases arbitrary barriers help building a separation between originals and copies.<sup>15</sup>

### 2.3. Copies and reproductions

There are also significant differences between copies and reproductions. Copies are made to reproduce as perfectly as possible the original work, but the copier may also be tempted to copy because he thinks that he can mimic the process of creation itself. Copiers use, in most cases, the same technique as the original artist, though there are many instances of drawings or paintings made after sculptures, and vice-versa.<sup>16</sup> The objective of reproductions is to give by any possible technical means, including digitalization, the very feeling and illusion of the original with little or no artistic intention. Copies can only be imperfect, because there is mediation and there may be interpretation. Van Gogh's Milletts can hardly be considered perfect copies. Reproductions

<sup>13</sup> *The Art Newspaper* 143, January 2004, p. 32. The author of the article (A.C. Grayling from Birkbeck College) wonders why "questions about the nature of art [are] still so intimately connected with questions of authorship in this age where the artist's hand is deliberately kept out of the work".

<sup>14</sup> It is interesting to note that Lessing (1983, p. 73) describes "modern prints from old litho stones [as] forgeries, though, assuming no deception is involved, forgeries of a peculiarly amoral, non-offensive sort".

<sup>15</sup> Alan Peacock made us aware of a contemporary Scottish artist, Jack Vettriano, who sells at auction at very high prices. One of his paintings, *Mad Dogs*, was sold for £140,000 on May 28, 2004. He became famous, at least in Scotland, through prints and postcards. This is what *The Scotsman*, 29 May 2004 writes about him: "Few artists have divided the art world as deeply as Vettriano. He is regularly slated by art critics who deem his work lacking in detail, subtlety and painterly technique. In the early 1990s, his work was turned down by the Royal Scottish Academy and the Scottish Arts Council, and he is still not represented in any major public collection. However, prints and posters have made him the best-selling artist in Britain, earning an estimated £500,000 a year in royalties." No need for originals!

<sup>16</sup> Henry Moore, for instance, produced sculptures "copied" from Masaccio's frescoes of the Carmine in Florence. See *Maison* (1960).

should be perfect and not aim at interpreting, although, as noted by Focillon (1919), photographs of the Mona Lisa can be unfaithful.

An interesting issue is coming up with color photographs and videos that both have a finite life and may fade. Research is going on at the Cesar Foundation in Basel to try to reverse this deterioration. The suggested solution consists in storing photographs in digital form and reprinting them in such a way that the original color is restored when the original fades. This is a perfect solution, except that it is an infringement of the 1976 American Copyright Act which protects against duplication “in any tangible medium of expression now known or later developed”.<sup>17</sup>

It is, however, doubtful that the difference between the terms “copy” and “reproduction” should be taken too seriously. They are very often used interchangeably. The *Oxford Advanced Learner’s Dictionary*, for instance, defines “to reproduce” as “to make a copy of a picture”, and in their *Dictionnaire de la Peinture*, Laclotte and Cuzin (1991) define copies as “imitations or reproductions” of a work, but consider repetition, replica, reduction, and facsimile as different categories of copies. The blurred boundaries between copies and other kinds of reproductions justify the relatively broad definition of copies that we adopt in our study of prices, which includes “copies”, “attributions”, “manner of”, “after”, “school of”, etc. These various qualifications are used in the special sales of copies organized by Christie’s in Amsterdam.<sup>18</sup>

### 3. Artists and their patrons

The boundaries between copies, reproductions and originals were far from being obvious during the Roman Empire. As noted earlier, the Roman Empire was full (and fond) of copies. According to Holtzmann (1996, pp. 846–849), copying was essential to artistic activity at that time, and was considered a way of recognizing the artistic talent of a master. The idea that the concrete reference to a model confers value to creation became even more important during the Middle Ages, as exemplified by the history of illuminated manuscripts.

During the Renaissance, copies and replicas (that is, copies by an artist of his own work) were produced in large numbers as well: imitation became almost a necessary activity. Michelangelo substituted his own drawings for older ones. Hans Holbein copied Flemish painters, but hid the fact [Martens (2001)]. Rubens modified old drawings. Pieter Bruegel left only 45 paintings, but these became so famous that they generated a large number of copies. We know of 123 versions of *Winter Landscape with Birdtrap*. The eldest son of Bruegel, Pieter the Younger, was very fond of exploiting this vein, almost inventing the idea of multiples; he produced several dozens of *Returning from*

<sup>17</sup> See *The Art Newspaper* 139, September 2003, p. 21.

<sup>18</sup> Christie’s October 7, 1995 sales catalog of copies in Amsterdam gives a glossary of words they use, and which have different meanings: replica, variant, version, copy, quotation, pastiche, paraphrase, parody, and persiflage.

*the Country Fair* and *25 St. John the Baptist Preaching* [Francastel (1995)]. Hoving (1996, p. 57) notes that inventories made after El Greco's death "list as many as five or six versions of the most noteworthy originals, all made in differing sizes, all of which the master could not have painted himself. They sold and still sell today as originals". Martens (2004) suggests that while Italian art had its theorists (Vasari, of course, during the 16th century, but also, Ghiberti, Alberti and Filarete, during the 15th century) the Flemish used copies to transmit their knowledge and discoveries.

All masters had ateliers where pupils specialized in certain items and worked for the master. According to de Saint-Simon (1996, pp. 339–344), in Rigaud's atelier, artisans produced portraits the prices of which varied with dimensions and according to how much had been corrected by the master.<sup>19</sup>

Copying went on during the 17th century. Rubens copied Leonardo, Michelangelo, Mantegna, and Raphaël. Greco copied Corregio, Michelangelo, and Titian.<sup>20</sup> Watteau copied Rubens.

Manet is said to have executed 400 copies from 290 sources. Van Gogh produced 520 copies [Schwartz (1996, p. 248)]. Bacon copied Velazquez and Van Gogh. Picasso copied Manet, Delacroix and others. Painters copy other painters, but they also copy their own works. Ingres copied himself, executing, for instance, *Paolo and Francesca* eighteen times, and writing very proudly that "the majority of those works of mine whose subjects I like have seemed worth the trouble of being perfected through repetition and retouching".<sup>21</sup> Gerôme, a "pompiers" painter successful during the 1850s, produces so many copies of his own work that Zola felt he had to write a satirical tract against him: "Mr. Gerôme . . . paints canvases and these are reproduced or printed in thousands of copies. The subject is everything, and the painting is nothing. Copies are worth more than the original".<sup>22</sup> Later in his life De Chirico used to copy his older style which had made him famous, but got into trouble because he antedated the paintings. Jasper Johns painted several versions of *Flag*, *Numbers*, and *Target*. And who knows how many *Marilyns*, *Maos*, and *Mona Lisas* Andy Warhol, the very king of copyists, did produce.

### 3.1. *The demand for copies*

Growing demand for paintings is probably at the root of the growing production of copies. During the Renaissance, art lovers welcomed exact copies that could be substituted for originals. Chamoux et al. (1973, p. 15) write that "the proof of talent is in the ability to produce a copy that can mislead the viewer". The Renaissance Italian poet

<sup>19</sup> Rembrandt's atelier differed from others and can be considered the first "academy" because he organized sessions during which he asked his pupils to draw and paint from living models.

<sup>20</sup> See, for example, Alpers (1988) on Rembrandt, or Montias (1982) on the atmosphere in the Low Countries in general.

<sup>21</sup> Ingres, *Ecrits sur l'art*, quoted by Schwartz (1996, p. 248).

<sup>22</sup> Quoted by Lafont-Couturier (1998, p. 37).

Petrarch purchased copies when he could not find the original of a painting he wanted for his collection [Dubus (1992)].

De Marchi and Van Miegroet (1996, p. 62) note that during the Renaissance, the “terminology original/copy was more fluid, [and that] it is not clear from observed practices that transactors in paintings (makers, dealers, buyers) thought in terms of a sharp distinction between the two. Confusion about the economic status of the original was, moreover, deliberately maintained by the workshops of leading masters in the fifteenth and sixteenth centuries”. However, as Martens (2001) points out, originality was the valued default option since without specific instructions to produce a copy, the painter was expected to be creative.

In the early 17th century, collecting became a fashionable activity among kings, noblemen, and even simple connoisseurs. Some collectors patiently assembled carefully chosen works; others took shortcuts. For example, the Duke of Lerma acquired 1431 paintings between 1599 and 1606; to achieve this, he bought series of works, such as 240 portraits of popes and 153 heads of Roman emperors [Brown (1995, p. 111)]. Others purchased copies which, though “recognized as inferior to the original, [were] considered as valuable record[s] of admirable composition[s] or invention[s]. Even so discerning a connoisseur as Charles I sent the copyist Michael Cross to Spain to copy the works by Titian” [Brown (1995, p. 111)]. Louis XIV bought copies. One of the Medici (Ottaviano) offered to Frederic the Second, duke of Mantua, a copy by del Sarto of a painting by Raphaël. When the king heard of the substitution, he became enthusiastic about the gift produced by a painter who was able to reproduce with such perfection the work of a great master. Colbert asked French artists to make as many copies as possible of paintings they could see in Rome.<sup>23</sup> Montias (1996, p. 24) estimates that “the proportion of copies among landscapes and still-lives in randomly selected inventories [in Amsterdam, during the years 1650 to 1669] may [have been] as high as one half to three quarters”. The demand for copies had a twofold purpose: training for artists and complementing collections. Creators and artists, as well as art amateurs and collectors, considered copies as valuable works of art. They were a way for artists (and connoisseurs) to get close to works that they would otherwise be unable to see, and are still a means for artists to train. Ayrton (1960, pp. 16–17) reminds us that “until recently even the wealthiest, most traveled and most studious of artists (for example, a Dürer or a Rubens) could only see in his entire lifetime the quantity of material now displayed to the casual visitor who spends one day in the Louvre or the Metropolitan museum”.

### 3.2. *A change in the status of copies*

It is not easy to pin down even the approximate time at which the quest for “originals only” decreased the relative prices of copies. In the mid-18th century, imitation loses ground.<sup>24</sup> The emergence of public museums in the late 18th century certainly

<sup>23</sup> Letter of 23 July 1672, quoted in Karpinski (1989, p. 105).

<sup>24</sup> This is particularly so in literature. See Mortier (1982).

played a role, as did at the same time, Winckelmann (1768) who was the first to point out that most of the antique marbles were Roman copies.<sup>25</sup> The romantic belief in the isolated artist with genius, which became dominant during the nineteenth century, also contributed to the decline in the appreciation of copies.

The decline of the status of copies is also the consequence of technical progress (photographs as substitutes for copies), and of legal intervention (the extension of property right laws). The invention of color photography obviated one of the major functions of copying: Copies progressively ceased to be a means for artists and connoisseurs to get close to works that they would otherwise be unable to see. The quality of mechanical reproductions improved so much that most customers found them good enough and stopped ordering manual copies,<sup>26</sup> except to complement collections.<sup>27</sup>

In the late nineteenth century, a *Musée des Copies*, also called *Musée Européen*, opened in Paris. A document introduced in the French Parliament clearly states the reasons for this: “Nobody thinks that only originals should be exhibited. If this were the case, there would exist only eight to ten cities with museums exhibiting original paintings, and only there would it be possible to educate the public”.<sup>28</sup> Charles Blanc, then Directeur des Beaux-Arts, commissioned copies from living French artists and tried to gather older copies exhibited in other museums. Blanc’s successor, Philippe de Chennevières, decided to close the museum a mere nine months after its opening; its exhibits (paintings and sculptures) were returned to the Ecole des Beaux-Arts, where they are probably still used for teaching and copying.<sup>29</sup> Those who were opposed to the museum suggested opening a photographic library showing expensive “hand made” copies. According to Chamoux et al. (1973, p. 29), this is the time at which forgeries started proliferating, and experts and “attributors” emerged.

But copies were still traded during the end of the 19th century. A lag is often observed between new ideas and their effect on markets. In the late 20th century, copies seem to raise renewed interest. Museums organize exhibitions of copies and use copies even in exhibitions devoted to originals. This development had some influence on the art market, which in turn had effects on copies.

<sup>25</sup> See also Ginzburg (1989).

<sup>26</sup> Ayrton (1960, p. 7) shares this viewpoint and adds that “the status of copies has fallen in proportion to the rise of reproduction”, as the quality of reproductions has increased. Copies were less useful to keep a trace of works of art whose state was deteriorating. The political function of copying disappeared during the same period: Copies were no longer useful for displaying the portraits of kings and emperors in public spaces.

<sup>27</sup> The German collector Fiedrich von Schack, for example, used to purchase masterpieces as well as copies between 1860 and 1880. He bought 55 copies of Venetian paintings in order to complement his collection.

<sup>28</sup> See *Annales de l’Assemblée Nationale* XIV, 11 November–21 December 1872, pp. 479–481.

<sup>29</sup> Details can be found in Chamoux et al. (1973) and in Cuzin (1993).



#### 4. Markets for prints

Prints are an excellent example of works that exist in multiple, but identical, “copies”. Prices give an insight on how markets create artificial scarcity [Melot (1973)] and generate a hierarchy that depends on the degree of “originality”.

It is useful to recall that prints are produced from engraved plates (copper or other metals).<sup>30</sup> Such plates may be reworked by the artist, leading to prints produced from each reworked plate, which differ from each other, though often not greatly, since corrections on copper plates and lithographic stones are difficult. Therefore, several consecutive “states” of the same work, printed from plates that have undergone several transformations, may coexist on the market. Some artists produce their own prints, others do not.<sup>31</sup> Some plates may survive the master and may be used, without transformation, to print so-called “late prints”. These are identical to the ones that are produced from the last state of the plate by the artist himself, but differ since he did not print them. Finally, some plates may be reworked after the artist’s death and lead to “posthumous” prints. In short, markets may be faced with three types of works:

- (a) original prints, produced by the artist during his lifetime, from plates that may have undergone transformations (states);
- (b) late prints, produced after the artist’s death, using his plates in the (last) state in which he left them; and
- (c) posthumous prints, also produced after the artist’s death, but from a plate that was transformed after his death.

Intuitively, there is less and less originality when going from (a) to (c). Type (a) prints can be considered to be fully by the master himself. Type (b) prints differ from the first ones, since they were not printed during the master’s lifetime, but use the original plate. Finally, type (c) prints are printed from a transformed plate, and not by the master. One may thus expect prices to decrease between (a), (b) and (c).

Lazzaro (2006) collected 4700 observations of Rembrandt prints sold at auction between 1985 and 1998, and runs regressions of prices on dummy variables describing the “state” of each print sold,<sup>32</sup> controlling for many other characteristics, including degree of rarity, technique used (etching, drypoint, etc.), subject matter (portraits, landscapes, etc.), experts’ appraisal (very fine, good, etc.), quality of impression, quality of paper (including special features such as watermark, etc.), conservation status, as described in the sales catalog (quality of margins, stains, other defects, etc.), salesroom, year of sale. Her observations include original prints by Rembrandt, late prints and posthumous prints. She also includes dummy variables which describe whether a specific state, original or not, was followed by other states, and if so, by which types of states (other

<sup>30</sup> Or drawings on a specific surface such as a lithographic stone. We simplify very much here. See Griffiths (1996) for details.

<sup>31</sup> Modern techniques, such as lithography, are very elaborate and it has become unusual for the artist to do his own printing, though he may supervise it. See Griffiths (1996, p. 102).

<sup>32</sup> Salesrooms consider states to be an important characteristic and document these in their sales catalogs.

Table 1  
Prices of prints by Rembrandt (second and further posthumous state = 100)

The print sold is a	Only one state by R. exists	Two states by R. exist	Three states by R. exist
First state by R., possibly followed by original states only	319	327	n
First state by R., followed by late prints	321	632*	n
First state by R., followed by posthumous states	319	537	799*
Second state by R., possibly followed by original states only	–	336	209*
Second state by R., followed by late prints	–	344	626*
Second state, followed by posthumous states	–	284	422
Third state by R., possibly followed by original states only	–	–	351
Third state by R., followed by late prints	–	–	219
Third state by R., followed by posthumous states	–	–	268
Late print, possibly followed by late prints only	180*	n	n
Late prints, followed by posthumous states	80	146	150
Posthumous state no. 1 (only one state exists)	141	160	167
Posthumous state (other than first)	100	100	100
No. of observations	1673	1633	818

Notes: “n” means that there were not enough observations to include a dummy in the regression; \* means that the regression coefficient is based on less than 30 observations. Results on the many other control variables are not reported. They are detailed in Lazzaro (2006).

original, late, posthumous). Since the results in which she included all the 4700 observations were difficult to interpret, she runs three different regressions, according to how many states of each specific work Rembrandt produced during his lifetime.

Her results, summarized in Table 1, show that original prints by Rembrandt are only two to three times more expensive than late prints by other printers. Thus a late print from an untouched plate does quite well, though it can be considered a copy. But what is more surprising is that prints obtained from plates that have been reworked after Rembrandt’s death fetch prices that are similar to those of late prints. It therefore seems that copies are still considered worthy, maybe because Rembrandt’s hand is “present”

in each print, making the distinction between originals and “copies” rather fuzzy. This is not the case for copies of paintings: the hand of the master is not present.<sup>33</sup>

It is also worth pointing out that prices of a state by Rembrandt do not seem to be influenced very much by the events that followed. Consider, for instance, the case where only one state by Rembrandt exists, and the following three possibilities: only original states exist, late prints followed original ones (the plate was not reworked), late prints and posthumous states printed from the reworked plate followed. One could assume that if late, and even worse, if posthumous prints exist, the price of original prints would be affected. By and large, this does not seem to be the case, and results from effects that, according to Lazzaro, go in opposite directions. Late and posthumous prints suggest a good and thus much demanded work, with the consequence of increasing the price of originals. But they may also depress the price, since they affect rarity, and make the sorting between originals and late or posthumous prints more difficult.

## 5. Markets for paintings

The case of paintings is quite different. The separation between copies and originals is clearer, though incomplete. De Marchi and Van Miegroet (1996) analyze the market for originals and copies in the Netherlands during the seventeenth century. Their research is based on a few cases for which they could compare the value of a specific original with the value of the copy, often made by the painter who produced the original. They found that originals are roughly three times more expensive than copies even if produced by the same painter, as was often the case in the past. They attribute this difference to the innovative character of the original, which is no longer present in the copy. It is very difficult to gather good data on transactions of copies, and we know of no other case than the one by De Marchi and Van Miegroet who were able to collect prices of paired works.

We collected data from auctions (often the only part of the market for which observations are available) organized during four periods separated from each other by some hundred years: 1684–1725 [mainly Amsterdam, Hoet (1752)], 1801–1810 [France, Peronnet and Fredericksen (1998)], 1890–1900 [Europe, but mainly France, Mireur (1901–1912)] and 1976–1999 (World, *The Art Sales Index*).<sup>34</sup> The data at hand do not make it possible to compare the prices of originals with those of copies made after these originals. But the number of observations is hopefully large enough to make comparisons of average prices between originals and copies meaningful.

<sup>33</sup> While art galleries and salesrooms do sell late prints by Rembrandt, they are more careful with photographs and avoid selling non-vintages photographs, though the “hand of the artist” may be considered to be as much (or little) present as for late prints. Another circumstance may have played a role also. The market for photographs started in the early 1980s, while art markets for paintings and sculpture were getting highly speculative. It may thus have been necessary to build trust for a newly developing medium, and impose the notion of “vintage photographs”.

<sup>34</sup> See Appendix A for a short description of the data.

Unfortunately, salesrooms are quite evasive in describing such works. Authors of the copies are usually not known, and only rarely are they by a “master”. The date at which the copy was made is very seldom given, which is of course unfortunate since a copy of Rubens made in 1650 even by an unknown artist, is worth more than a copy made in 1850. Even nowadays, copies sell at auction, but salesrooms, at least Christie’s and Sotheby’s, draw a distinction between those produced before and after 1800. The former are usually offered at regular Old Master sales and are therefore almost considered Old Masters. Later copies are also auctioned by international salesrooms, but during non-specialized sales of *varia*.<sup>35</sup> Contemporary copies of Old Masters are usually not sold at auction but by specialized galleries. There is thus a distinction to be drawn, not only between old and more recent copies after Old Masters, but also between copies after Old Masters and copies after artists who were active later.

Table 2 reports on our findings for the four periods. It provides the number of copies,<sup>36</sup> and the average relative price of copies.<sup>37</sup> Number of copies and relative prices are disaggregated by “type” of copy, period and country in which the painter after whom the copy was made was active.<sup>38</sup> Even if it is far from obvious that the data are homogeneous over time, we can draw some conclusions. Copies were reasonably well-priced (a copy was worth 30 to 50 percent of an original) until 1900, but there is a very sharp drop sometime afterwards, since the best copies (that is, those that are worthy enough to be sold at auction) are worth 35 times less than originals in the late 20th century.

Neil De Marchi (2004) suggests that this relative price effect for Old Masters may be the outcome of two factors. Due to their short supply (compounded by the increasing wealth of buyers) Old Masters paintings may have undergone an absolute increase of prices. Simultaneously, the technical possibilities that made detection of copies much easier may have prevented collectors to sell works that they had acquired as originals in the past, so that only the ones that were known to be copies would reach the market (an Akerlof effect) and depress prices of copies. From there follows an obvious decrease of the relative prices of copies with respect to originals.

<sup>35</sup> Christie’s considers that “the irruption of photography and graphical reproduction techniques in the beginning of World War II, has made all of this obsolete. For these reasons, we do not include copies made after 1940 in our sales”. See Christie’s press release for its 29 September 1998 Amsterdam sale. According to Ms. Aarts from Christie’s Amsterdam, this decision is also made to avoid selling copies that could have been considered as fakes at a certain moment, without the salesroom being aware of it.

<sup>36</sup> Replicas that salesrooms as well as art historians consider as originals, are not included.

<sup>37</sup> The relative price of a copy is the ratio between its hammer price and the median price of originals by the painter who is copied. The median price was chosen since in some cases, the number of originals was small. The median also excludes extreme prices which can just be due to mistakes.

<sup>38</sup> For the first three periods, the search was made on copies, and not on painters. Given the nature of the most recent database (Art Sales Index), the search was made on painters for whom it was known that many copies existed (Boucher, Hals, Raphael, Rembrandt, Rubens, Van Dyck and Watteau). For these data, it was impossible to separate copies according to whether they were “attributions”, “school of”, “copies” and “other”.

Table 2  
 Characteristics of the samples of copies 1684–1999 (number of copies and relative prices)

Description	1684–1725		1801–1810		1890–1900		1976–1999	
	No.	Price	No.	Price	No.	Price	No.	Price
Attributions	6	0.269	148	0.845	93	0.521	n.a.	n.a.
School of	0	–	176	0.591	74	0.290	n.a.	n.a.
Copies	73	0.280	396	0.572	4	0.095	n.a.	n.a.
Other	33	0.307	129	0.430	4	0.241	n.a.	n.a.
Original painted								
Before 1600	33	0.346	199	0.838	50	0.676	570	0.005
1600–1700	79	0.264	611	0.552	85	0.331	1087	0.028
After 1700	0	–	39	0.176	40	0.235	149	0.099
Country of artist “copied”								
Italy	25	0.487	313	0.711	27	0.261	570	0.005
Low Countries <sup>a</sup>	76	0.220	352	0.548	84	0.476	1087	0.028
France	9	0.256	176	0.495	50	0.461	149	0.099
Other <sup>b</sup>	2	0.545	8	1.000 <sup>c</sup>	14	0.089	0	–
Italy before 1600	20	0.471	188	0.793	24	0.276	570	0.005
Italy 1600–1700	5	0.552	119	0.595	0	–	0	–
Italy after 1700	0	–	6	0.435	3	0.137	0	–
Low C. before 1600	11	0.063	3	3.189 <sup>c</sup>	21	1.241 <sup>c</sup>	0	–
Low C. 1600–1700	65	0.243	349	0.526	63	0.221	1087	0.028
Low C. after 1700	0	–	0	–	0	–	0	–
France before 1600	0	–	0	–	3	0.129	0	–
France 1600–1700	9	0.256	143	0.580	12	1.150 <sup>c</sup>	0	–
France after 1700	0	–	33	0.129	35	0.254	149	0.099
Other before 1600	2	0.545	8	1.000 <sup>c</sup>	2	0.364	0	–
Other 1600–1700	0	–	0	–	10	0.039	0	–
Other after 1700	0	–	0	–	2	0.059	0	–
Total	112	0.288	849	0.602	175	0.408	1806	0.028

<sup>a</sup>Flanders and Holland.

<sup>b</sup>Germany and Spain.

<sup>c</sup>Copies at least as expensive as originals.

Table 3 gives the results of regressing (the log of) relative prices of copies on type of copy: attribution, school of, copy and other, controlling for other characteristics (country of origin of the artist copied, and the time during which this artist was active). All variables appearing in the right-hand side of the equations are dummies. Here, we are

Table 3  
Estimation results

Description	1684–1725		1801–1810		1890–1900	
	Coeff.	St. error	Coeff.	St. error	Coeff.	St. error
Attributions	0.000	–	0.000	–	0.000	–
School of	–	–	–0.307*	0.139	–0.288	0.244
Copies	–0.165	0.421	–0.451*	0.121	–1.126	0.724
Other	–0.143	0.442	–0.454*	0.150	0.153	0.705
Original painted						
Before 1600	0.000	–	0.000	–	0.000	–
1600–1700	0.884*	0.277	–0.019	0.142	–0.893*	0.300
After 1700	–	–	–1.238*	0.253	–1.513*	0.407
Country of artist “copied”						
Low countries <sup>a</sup>	0.000	–	0.000	–	0.000	–
Italy	1.623*	0.296	0.689*	0.130	–0.545	0.377
France	0.203	0.354	0.334*	0.124	0.930*	0.357
Other <sup>b</sup>	1.625*	0.740	0.923*	0.474	–0.911*	0.402
Intercept	–2.684*	0.483	–1.295*	0.182	–1.172*	0.305
R-squared	0.239		0.105		0.165	
No. of observations	112		849		175	

The dependent variable is the relative price of the “copy”.

<sup>a</sup>Flanders and Holland.

<sup>b</sup>Germany and Spain.

\*Coefficient is significantly different from 0 at the 5% probability level.

essentially interested in the coefficients picked by the “type of copy” dummies, and expect that attributions (normalized to zero) will fetch the largest relative price, and will be followed by “school of”, “copy” and finally, “other”. As expected, attributions fetch higher relative prices than other types of copies in each time period, but significantly so only between 1801 and 1810. Prices for attributions are relatively close to those of originals, since often, a mere missing signature raises doubt. The drop in the relative price for copies after Italian, Dutch and Flemish masters may be due to a change in tastes, but more careful research may be needed to confirm this conclusion.<sup>39</sup>

<sup>39</sup> It would be interesting to check whether relative prices changed significantly over the three first periods. This could be tested by running a regression in which the three sets of data are pooled, dummies for periods are introduced, and a test is run which checks whether the coefficients picked by these period dummies are significantly different or not. Unfortunately, there are two difficulties here. First, definitions of types of copies

Table 4  
Christie's Amsterdam specialized sales of copies

	1994	1995	1996	1997	1998
Total number of works	50	52	51	71	73
Number of works unsold	8	9	11	1	11
Number of works sold at a price above the upper limit of the pre-sale estimate	15	17	15	47	23

Notes: in the 1994, 1995 and 1996 sales, paintings were followed by other works; in 1997 and 1998, the sale consisted of paintings only.

Since 1994, Christie's organizes special sales of copies in Amsterdam,<sup>40</sup> at which copies after Old Master pictures are auctioned. The first three sales consisted of some fifty paintings and drawings, followed by a mixed bag of various objects (glass, ceramics, silver, jewelry, furniture and other works of art) obviously designed to fill the session. These sales were organized on Saturdays to attract "young" buyers who were interested in art but unable to afford originals. Christie's then changed the concept, and since 1997 the sales consist of paintings only and are held during the week, like most other sales. This change suggests that both supply and demand increased and that markets have become more active. To check for the validity of this assumption, we looked at some characteristics of the sales, such as the number of unsold works, the number of works sold at more than the upper limit of the pre-sale estimate range, and prices.<sup>41</sup> Those data may be considered as an indication of tensions on the market. Table 4 summarizes our findings, and shows that there is an upward trend.<sup>42</sup>

Most interesting is that, according to Christie's Amsterdam specialists, these sales are the firm's (that is, the Amsterdam subsidiary) most internationally attended auctions. For example, the ten most expensive items sold at the 1998 auction went to American, Italian, Belgian, German, and Dutch traders and to German, Taiwanese, Dutch, and Irish collectors.<sup>43</sup> Of course, Christie's holds only one such specialized sale per year,

may have changed over time and second, later periods may include old copies and more recent copies, that fetch different prices. Since copies are only rarely dated, or attributed to specific artists (and salesrooms do not find it worth to spend time on finding the date), we could not introduce production dates among the explanatory variables.

<sup>40</sup> It is not clear whether the sale is organized in Amsterdam to keep London and New York "clean", or whether this is due to many copies being after Dutch and Flemish artists. Note that such copies are also sold in Christie's South Kensington rooms, but together with sales of other objects.

<sup>41</sup> Increasing average prices do not necessarily indicate that prices increase, since the increase could simply be due to larger dimensions, or better quality. This is implicitly taken into account in the pre-sale estimates by the salesroom. Therefore, works that are sold at more than the pre-sale estimate are, we think, a better indicator of increasing demand.

<sup>42</sup> Moulin (1992) reports on a large increase of the demand for copies by the new Canadian bourgeoisie.

<sup>43</sup> According to the after-sale press release.

but as was already mentioned, every Old Master sale, whether in London, New York, or Amsterdam (by both Christie's and Sotheby's), contains works that are "attributed to", from the "circle of", in the "manner of", "after", "with signature", "with seal", or by a "follower of".<sup>44</sup>

Contemporary copies, including after Old Masters, are produced and traded by specialized galleries. Donors often have copies made of the works they offer to museums. Bikauer, an American collector, offered a very large canvas by Delaroche (a French painter [1797–1856]) to the Louvre and ordered a copy of the painting to keep as a "souvenir" of the original one. The owners of the château de Vaux-le-Vicomte near Paris had copies made of most of the paintings that were once hanging in the castle but were sold, lost, or stolen. Guardis and Canaletto were also copied and sold by the Parisian Galerie Delamarre to complement incomplete collections elsewhere.

All these signs bear testimony to the constitution and vitality of a market for copies, though as we shall see in the next section, lawyers make their status ambiguous. The law is, as it should be, opposed to fakes, but is unclear about copies. Legal decisions are often inconsistent and scare away buyers and sellers.<sup>45</sup> Curators and art historians are taking timid steps to recognize the value of copies. Although they still do not admit that copies can be included in the collection of a museum, they organize exhibitions devoted to copies or in which both originals and copies are displayed. Of course, the accent is still put on originals. In this respect, the case of the *Madonna of the Pinks* is remarkable. Until its attribution to Raphael in 1991, the painting was worth a very small amount, possibly as little as £8000.<sup>46</sup> It was paid £22 million by the National Gallery in 2004.

## 6. Copies and art historians. A renewal of interest?

The attitude of art historians towards copies has also undergone changes. Contemporary art historians consider copies as one of the means to study tastes and norms over time.

<sup>44</sup> This is extracted from Christie's sale catalogs, which describe the various cautious denominations as follows. "Attributed to" means "In our opinion probably a work by the artist or maker in whole or part". "Circle of" means "In our opinion a work of the period of the artist or maker and showing his influence". "Manner of" means "In our opinion a work executed in the style of the artist or maker but of a later date". "After" means "In our opinion a copy of any date of a work of the artist or maker". "With signature/seal" means "Has signature/seal which in our opinion is not that of the artist". Finally, "follower of" means "In our opinion a work executed in the artist's style but not necessarily by one of his pupils". Very similar words and descriptions appear in Sotheby's sale catalogs. Note that in 1882 already, Burckhardt had suggested such distinctions, based on the works by Rubens and his workshop. See Burckhardt (1882).

<sup>45</sup> Justice Laddie (1996), one of the best known British experts on copyright law holds the view that "these detailed and pedantic exceptions to copyright protection are not only difficult to understand in some cases, but they also reinforce the perception that virtually all reproductions of copyright works, no matter how innocuous, are infringements".

<sup>46</sup> See *The Art Newspaper*, September 2003, p. 9.



In particular, they suggest that in the past copies may have been used as “textbooks”, especially in Northern Europe [Martens (2004)]. Lobstein (2002) studies paintings that were copied in France at the end of the 19th century, and uses the information to understand how collections were constructed.<sup>47</sup> The large number of copies after Goya in Aquitaine in the middle of the 19th century is closely linked to the influence of Spanish culture in this southern region of France. This was also a time at which the French State ordered copies of Spanish masters (Murillo, Ribera, Velasquez and others). Fashion changed after 1870 as shown by copies after Italian painters who have crowded out Spanish artists.

Contemporary art historians are also less fascinated by the romantic idea that has given so much importance to the concept of originality, and take into account the legacy left by Marcel Duchamp or Andy Warhol, that originality is nonsense. Warhol tried to turn things around with his prints of the Mona Lisa, claiming that it is the multiplicity of reproductions that made a star out of her. After all, everybody “knows” a well-known movie star, though very few know her personally. She only exists through the attention given to her in the media. Therefore, and contrary to the predictions made by Benjamin (1971), copies and reproductions do not destroy the aura of the original, but contribute to its value [Hughes and Ranfft (1997) and Seguy Duclot (1998)]. This is probably what Warhol meant with his silkscreen *Thirty Are Better Than One*. Copies and reproductions are the best testimony to the value of originals [Belting (1998)], and a work that does not inspire copies is a dead work [Pastoureau (1991)].

Art philosophers are often even more positive and consider that the borderline between copies and originals should be revised. Caillois' (1975) praise of the creative nature of copies is becoming more and more influential. Goodman (1983, p. 100) prefers a good copy of Lastman by Rembrandt to the original Lastman.<sup>48</sup> Meiland (1983, pp. 122–123) claims that “if what a great work of art does is to present us with a new vision of the world, then an exact copy can perform exactly the same function”. Originality, he writes, should not be praised *per se*, but matters only if the work is good. He goes as far as writing that “though there are many situations in which an original has or would have greater total aesthetic value than an exact copy . . . there may also be situations in which an exact copy has at least as much aesthetic value – primary and derivative – as the original. And this should not be surprising since an exact copy expresses exactly the same work of art as the original” [Meiland (1983, p. 130)]. Therefore, it does not matter whether the original of a self-portrait by Rembrandt hangs in The Hague, as was thought until recently, or in Nuremberg, as was discovered a few years ago.<sup>49</sup> Nor does it if the Louvre *Madona of the Rocks* was painted by Leonardo before or after the one that hangs in the National Gallery.

Curators who devote more and more exhibition space to copies also share this view. The Paris *Musée des Copies* opened in 1872 failed, but a “museum” of copies recently

<sup>47</sup> He especially stresses the role of the Galerie espagnole in the Musée du Louvre.

<sup>48</sup> See also Foucart (1973) who discusses copies that Jordaens made after paintings by Rubens.

<sup>49</sup> See *Le Monde*, 10 December 1998.

opened in Naruto, Japan. Its more than one thousand exhibits consist of copies of some of the most celebrated European works of art, including life-size reproductions of the Sistine Chapel, Giotto's frescoes in Padova, and Pompeian villas. The copies were produced with authorizations from important museums, including the Metropolitan, the Louvre, and the National Gallery. Whether the final products are copies or reproductions is unclear. The technology used is kept secret. It starts from slides copied onto enameled panels,<sup>50</sup> during the last stage, some depth and relief are added, which according to the curator adds to the impression that it is the "real thing".<sup>51</sup> Lascaux consists of "five hundred tons of modeled concrete reproducing every little bump and hollow of the original, with a precision of a few centimeters . . . [and including] the patina of time".<sup>52</sup> And visitors come in droves to see the copy, since the original is reserved for a happy few. Schwartz (1996, p. 249) notes that the Diaspora Museum in Tel Aviv (Beth Hatfoutsoth) which illustrates Jewish life across centuries "is all copies and makes no bones about it".

The Georges de La Tour exhibition in Paris in 1997–1998 displayed many copies, because originals are lost. One of the rooms was devoted to eight copies of de La Tour's *Saint Sebastien soigné par Irène*, of which no original is available. The Millet–Van Gogh exhibition in Paris in 1998, made clear that "copies" of Millet by Van Gogh may even become more valuable and expensive than originals.<sup>53</sup>

Borges (1944) provides a beautiful illustration of this paradoxical situation. In his short story *Pierre Menard*, he "describes two fragments of works, one of which is part

<sup>50</sup> Perhaps the technology is not that new after all. Stendhal in his *Voyage en France*, published in 1837, writes: "J'ai vu chez M. Bonnard plusieurs magnifiques tableaux en porcelaine de M. Constantin. Dans deux siècles, on ne connaîtra plus les fresques de Raphaël que par les tableaux de M. Constantin". Constantin's short-lived success was due to his talent as a copyist. He even required that the copies be lit so as to reproduce the conditions in which the original is exhibited.

<sup>51</sup> See its flier *What Is the Otsuka Museum?* Naruto: Otsuka Museum of Arts, 1998.

<sup>52</sup> Delluc and Delluc (1984), quoted by Schwartz (1996, p. 249).

<sup>53</sup> In recent years, such exhibitions were organized in Dresden in 1970 (Exhibition Catalog by W. Schmidt et al., *Dialogue-Kopie, Variation und Metamorphose alter Kunst in Graphik und Zeichnung vom 15. Jahrhundert bis zur Gegenwart*, Dresden: Kupferstichkabinett, 1970), in Münster in 1976 (Exhibition Catalog by G. Langemeyer and R. Schleier, *Bilder nach Bildern, Druckgraphie und die Vermittlung von Kunst*, Münster: Westfälisches Landesmuseum für Kunst und Kulturgeschichte, 1976), at Sotheby's London in 1977 (Exhibition Catalog by B. Nicolson, *Art into Art, Works of Art as a Source of Inspiration*, London: Sotheby's), in Vienna in 1980 (Exhibition Catalog by H. Hutter, *Kopie-Replik-Paraphrase*, Vienna: Academie der Bildende Kunst, 1980), in New York in 1988 (Exhibition Catalog by E. Haverkamp Begemann, *Creative Copies, Interpretative Drawings from Michelangelo to Picasso*, New York: The Drawing Center, 1988), at the National Library in Paris in 1991 (Exhibition Catalog by M. Pastoreau, *Vrai ou faux? Copier, imiter falsifier*, Paris: Bibliothèque Nationale, 1991), at the Louvre in 1993 (Exhibition Catalog *Copier et créer. De Turner à Picasso: 300 oeuvres inspirées par les maîtres du Louvre*, Paris: Musée du Louvre, 1993), in Delft in 1995 (Exhibition Catalog *Schone Kopieën, gerestaureerde Kopieën naar Oude Meesters door Paul Tetar van Elven*, Delft: Museum van Elven), in Geneva in 1997 (Exhibition Catalog *L'art d'imiter. Falsifications, manipulations, pastiches. Images de la Renaissance italienne*, Genève: Musée d'Art et d'Histoire de Genève, 1997), in Paris in 2000 (Exhibition Catalog *Monter/Sampler*, Paris: RMN, 1990) and in Paris in 2002 (Exhibition Catalog *Manet-Velazquez, la manière espagnole au 19ème siècle*, Paris: RMN, 2002).

of *Don Quixote* by Cervantes, and the other, like it in every graphic respect – like it, indeed, as much as two copies of the fragment by Cervantes could be – which happens to be by Pierre Menard<sup>54</sup> and not by Cervantes”. Nevertheless, “Borges tells us that the *Quixote* by Menard is infinitely more subtle than that of Cervantes, while that of Cervantes is immeasurably more coarse than its counterpart even though every word contained in the Menard version can be found in Cervantes”.<sup>55</sup> To make things even more difficult, Danto adds that “it is a fact that the two works identified by Borges, that of Cervantes and that of Menard, would generate classes of indiscernible copies, the one class copies of the work of Cervantes, the other copies of that of Menard: but these would be copies of different, even importantly different, works, though nothing would be easier than to mistake a copy of Cervantes for a copy of Menard”.

## 7. Regulating the market for copies

Copies are often considered “the first degree of fakes” [Chamoux et al. (1973)]. One of the main reasons for their low status is their possible illegal use. A well-known case illustrates this assumption. Nelson Rockefeller decided to have copies made of 118 works from his private collection and of works he had donated to museums. Copies of Rodin were stamped on their base with a copyright symbol, a copy number, and a date. In 1980, a gallery offered one as an authentic work: the stamps had been covered with shoe polish!<sup>56</sup> Accordingly, as Merryman and Elsen (1987, p. 61) point out, “exact reproductions of works of art, like exact reproductions of money, postage stamps, or other objects whose value inheres in their authenticity, easily lend themselves to misuse by the unscrupulous, who may misrepresent them as originals rather than as reproductions . . . Exact reproductions of works of art devalue original works by creating confusion between originals and reproductions. The more exact the reproduction, the greater the potential confusion and the consequent devaluing effect”.

Regulations by museums and intellectual property laws organize and limit the production of copies. The question is whether this helps clarifying their status or whether, on the contrary, regulatory and legal restrictions increase the degree of uncertainty.<sup>57</sup>

### 7.1. An unstable regulatory framework in museums. The case of the Louvre

The Louvre began to organize (and limit) copying at a very early stage. When the museum was created in November 1793 (as the Musée National des Arts), its 538 paintings

<sup>54</sup> Pierre Menard is supposed to have written his *Quixote* in 1899, while Cervantes’ *Quixote* was written in the 16th century.

<sup>55</sup> Danto (1981, pp. 33–36).

<sup>56</sup> See Merryman and Elsen (1987).

<sup>57</sup> See also Chapter 7 by Landes and Levine in this volume, especially the sections on copyright and trademark issues, and on authenticity and the law.

were accessible to artists during five, then seven days out of ten (the “revolutionary decade”). The general public came to the museum to order copies [Schaer (1993)], and copying was free. A first restriction was imposed in 1797: Copyists had to register and a master or a school had to provide a certificate.<sup>58</sup> The only constraint was that the works could not be moved while copying was going on. The 1824 regulation did not change much, with the exception of the days reserved for artists (Tuesday to Friday) and some rules concerning silence and health issues. Important sales of copies were organized during this period: in 1809, for example, Le Brun organized the sale of the Van Hoorn collection, an important Dutch collector of copies.

The 1848 regulation was still favorable to copyists, who could work every day. The crowd of artists executing copies to order became so large that the Louvre was forced to adopt the rule that only one copyist was allowed to copy a specific painting. In 1865 a new concern appeared in the “Règlement des musées impériaux”, concerning copyists who used to work in the Musée du Luxembourg, dedicated to contemporary art. They had to apply for an authorization from the Sénateur Surintendant des Beaux-Arts (and not from the artist of the original painting). An important sale of copies painted by Colin, a well-known copyist at the time, was organized in Paris in 1860. It contained copies of paintings by 71 artists.

Additional restrictions were imposed in 1872, 1875, 1893, 1903, and 1907, but with no important consequences. In 1908, the concern for living artists became stronger: Copyists of works “whose rights to be reproduced [were] ‘reserved’ could do so but not sell their work without being authorized”.<sup>59</sup> In 1913 a rule stipulated that there could be no more than one copyist in some of the rooms (especially the Dutch Cabinets). In 1927 and 1932, new restrictions appeared that could be interpreted as the first attempt to limit the number and even the value of copies: A copyist was not allowed to work during more than two months on the same painting; copyists could not negotiate the sale of their work in the museum; copies had to be stamped before the copyist could move them from the museum.<sup>60</sup> In 1946, the Règlement des musées nationaux added that copies could not be of the same size as originals, whatever the date of the original.<sup>61</sup> Copying in museums became therefore a totally regulated activity.

Several European museums followed with similar rules. Between 1852 and 1856, the Rijksmuseum in Amsterdam allowed copying sessions from Monday to Wednesday only.<sup>62</sup>

In the beginning of the 20th century, all the elements that draw a clear distinction between originals and copies and suggest a definitive hierarchy were there. This slow

<sup>58</sup> See *Rapport du Conservatoire du Musée national des Arts au Comité d'instruction publique*, Paris: Musée du Louvre, 1797.

<sup>59</sup> Article 8 concerning the Musée du Luxembourg.

<sup>60</sup> See *Règlement des conditions de travail des copistes et photographes dans les salles et galeries des musées nationaux*, Paris: Archives du Musée du Louvre, 1932.

<sup>61</sup> This is still so nowadays.

<sup>62</sup> See the Introduction to the *Christie's Amsterdam sale Catalog of copies*, 4 October 1994.

but inevitable construction of rules went together with the development of a speculative modern art market, and an increasingly elaborated legislative framework to protect artists' intellectual property rights. However, the distinction between copies and originals is not always as clear as it should be.

## 7.2. The legal definition of originality

A few years ago Christie's sold a painting by Egon Schiele that the buyer later found to be heavily restored. A judgment directed the auction house to reimburse the buyer because, with 90 percent of the work repainted, it could no longer be considered by Schiele. The judgment did not discuss the fact that the original painting was by Schiele, nor did it discuss at what point the restoration rendered the work not by Schiele. One can wonder whether the court's judgment would have been different if only 80 percent had been repainted. Meanwhile, a completely repainted Barnett Newman *Who Is Afraid of Red, Yellow and Blue?* which was badly damaged by a vandal, hangs in the Stedelijk Museum in Amsterdam as if nothing had ever happened. The painting is hardly an original by the artist, but is displayed as if it were an original.<sup>63</sup>

According to Mund (1983, p. 128), copies may be exact (similar to the original, but produced at a later date) or interpretative (freely transforming the original work, and adopting the style of the copyist). This twofold definition is consistent with the legislation that defines counterfeiting as a non-authorized borrowing of part or the entire original work. In *Steinberg vs. Columbia Pictures Industries, Inc.* (663 F. Supp.706, S.D.N.Y. 1987), Saul Steinberg, the well-known *New Yorker* cartoonist, sued the producers, promoters, distributors and advertisers of the movie *Moscow on the Hudson*. He alleged that the promotional poster for *Moscow* infringed his copyright on an illustration that he had produced for *The New Yorker*. The court recognized that the original and the copy were "substantially similar", though not all the details were identical. As Stanton (2002) notes, it is now admitted that "the copying need not be of every detail so long as the copy is substantially similar to the copyrighted work". Substantial similarity does not require identity, and "duplication or near identity is not necessary to establish infringement".<sup>64</sup> As Judge Learned Hand wrote for the *Krofft* case, "no plagiarist can excuse the wrong by showing how much of his work he did not pirate".

Originality is the central notion that separates originals from copies, but defining what is lost or gained in duplicating is not easy.<sup>65</sup> Legal texts on the issue are elusive:

<sup>63</sup> Some years ago, one of us walked into the room where it is exhibited, and asked the warden whether he knew when the painting came back to Amsterdam after its restoration in the United States. The answer was that it had never left the room where it hangs.

<sup>64</sup> Stanton (2002) also cites other such cases: *Comptone Co. v. Rayex Corp.*, 251 F.2d 487, 488 (2d Cir. 1958); *Krofft*, 562 F.2d at 1167; *Sheldon v. Metro-Goldwyn Pictures Corp.*, 81 F.2d 49, 56 (2d Cir.); *cert. denied*, 298 U.S. 669, 80 L. Ed. 1392, 56 S. Ct. 835 (1936).

<sup>65</sup> For Arnheim (1983, p. 237), "it is not sensible to accept only original works as art and dismiss all reproductions as nonart".

“All that is needed [...] is that the ‘author’ contributed something more than a ‘merely trivial’ variation, something recognizably ‘his own’”.<sup>66</sup>

In the case of multiples, the notion of originality is bound to remain confused. Artists such as Whistler, Degas, or Picasso signed their prints as evidence of their authenticity. Merryman and Elsen (1987, p. 526) distinguish between

- (a) prints by well-known artists that are merely reproductions of original works, even though they may be well printed on fine paper in limited editions and signed by artists; and
- (b) lithographs, etchings, woodcuts, or other print media that are a primary form of art-making as is the case with Jasper Johns.

Originality and uniqueness are often associated with scarcity,<sup>67</sup> obtained by limiting the number of copies.<sup>68</sup> In the case of Rodin’s work, for example, a decree sets the number of authorized *post-mortem* originals [Gautier (1991)]. By analogy, uniqueness is, still today, a central argument.

Legal precedents reveal large variations in the definition of copy. First, the law permits copying of manner, technique, and style. Desbois (1978) observes that one “can oppose to counterfeiting a slavish copy or adaptation, but the imitation of genre, or artist’s manner, cannot be prosecuted”.<sup>69</sup>

Second, originality is frequently associated with quality. In the case of a painting by Braque, reproduced as a print, a Paris court detected the “skill and sensitivity of the engraver” and authorized copying without compensating the heirs of the artist (December 19, 1992). In the American *Hand of God* case, in which an artist had produced a scale reduction of a sculpture by Rodin (entitled *Hand of God*), a district court found that the originality was due to the fact that “it takes an extremely skilled sculptor many hours working directly in front of the original” to make an exact copy.<sup>70</sup> This case resembles the one brought by the buyer of a Dürer against an art dealer during the 1650s, when the buyer realized that the Dürer was a copy by Luca Giordano, who had hidden his own signature on the back of the wooden panel. The quality of the copy was so good that Giordano won the case.<sup>71</sup>

Third, a copy may include the signature of the original artist. In 1992, legal action was brought against the Parisian Galerie Delamarre for reproducing and selling copies

<sup>66</sup> Merryman and Elsen (1987, pp. 183–184).

<sup>67</sup> Melot (1994) devotes a whole chapter to the different techniques used to create scarcity, such as numbering of the prints, adding colors, or the autograph signature.

<sup>68</sup> Degas usually controlled the production of prints, making each of them unique. He refused to produce industrial reproductions, though he had a real passion for engravings, prints, and so on. Melot (1994) explains how Degas or Pissarro tried to make each print unique: “The very definition of a work of art can only be in opposition with a series. Quality is incompatible with quantity”.

<sup>69</sup> See also Colombet (1997).

<sup>70</sup> *Alva Studios, Inc. vs. Winninger*, quoted in Merryman and Elsen (1987, p. 184).

<sup>71</sup> This is also reminiscent of the story told by Koestler (1989, p. 402) of the competition held to choose among a dozen guests imitating Charlie Chaplin, to decide who came the closest to the original: “Chaplin himself happened to be among them – and got only the third prize”.

of masterpieces that included the painter's signature. Whereas the initial judges found the gallery guilty of infringement on the grounds that the right to copy does not include the right to copy the author's signature, the French High Court rejected the judgment (June 11, 1997, confirmed by the Paris Court of Appeals, October 13, 1999). The court argued that "reproducing, on a copy of a work in the public domain, the signature of the author does not infringe upon the author's moral rights when there is no risk of confusion between the original and the copy".<sup>72</sup> The court added that there was no risk of confusion since the dimensions of the painting were different and the word "copy" was marked indelibly on the back and the edges of the canvas.

Fourth, the definition of originality varies according to which judicial authority or administration is in charge of the evaluation. Post-mortem works produced by the Musée Rodin are considered originals by the French fiscal administration but as non-originals by intellectual property laws, and are, therefore not liable to resale rights.<sup>73</sup>

### 7.3. Public and non-public domain

A work enters the public domain seventy years after the death of its author and becomes freely available for copying, advertising, and publishing.<sup>74</sup> Copiers of a work that belongs to the public domain are not subject to copyright. Other works occupy a legal gray area: neither the American nor the French law clearly defines the terms "copy" and "reproduction". A copy is assimilated to a photograph; the signature is an element of the painting, and changing the size is a way of escaping counterfeiting laws.

There is no definition of a "work of art" in the law, either. For inheritance taxes on works of art, the French administration bases the levy on the opinion of an ad hoc committee of experts.<sup>75</sup> Until 1994, the French definition of "original print" included the terms "executed by the artist" and excluded mechanical procedures; under such a definition, Rauschenberg's or Ernst's best engravings would not be originals [Melot (1973)]. The American Print Council excludes post-mortem editions (but the date of the edition is sometimes rather difficult to discover) and defines three conditions for a print to be original:

<sup>72</sup> *Le Journal des Arts*, 95, December 18, 1999, p. 44.

<sup>73</sup> *Le Journal des Arts*, 5, July 1994.

<sup>74</sup> Note that not even this view is universally accepted, as is illustrated by the conflict between the Bridgeman Art Library (which grants licenses to reproduce works of art belonging to European and American museums) and the Canadian firm Corel (publisher of Professional Photos CD-Rom Masters, containing reproductions made without the consent of Bridgeman). The lawyer of Corel holds the opinion that "Bridgeman's ektachromes are not originals and are therefore, not submitted to copyright" to which the lawyer of Bridgeman reacts by claiming that "the skills of the photographer who creates a large ektachrome and digitalizes it possess all the characteristics of a creation on a new support". See *Le Journal des Arts* 68, 9 October 1998. The case was judged in a New York court. The federal judge accepted Corel's arguments and ruled that the photographs were "slavish copies" that required both skill and effort but were not original enough to benefit from copyright protection (*Bridgeman Art Library, LTD. v. Corel Corporation*, 36F. Supp. 2d191, S.D.N.Y. 1999).

<sup>75</sup> Bourdon, Pontier and Ricci (1996).



- (a) the artist alone has created the master image in or upon the plate, stone, wood-block or other material, for the purpose of making the print,
- (b) the print is made from the announced material by the artist or according to his directions, and
- (c) the finished print is approved by the artist.<sup>76</sup>

In the European Union, a law passed in 1994 defines works of art to which a lower rate of value-added tax is applicable. Under this law, for a photograph to be a work of art, it must be made by the artist, printed by him or under his supervision, signed and numbered with a limit of thirty prints, regardless of dimension or medium. More generally, European regulations define the two characteristics that define a work of art: It has to be unique (or exist in a limited number of copies) and it has to be produced by the artist himself or under his strict supervision [Melot (1995)].

#### 7.4. Copies as copyrightable goods?

Copies are both subproducts and works by themselves. A copy remains a “second best”, an imperfect substitute for the original, but may also be considered a work of art, as are the copies of Old Masters auctioned by Christie’s or Sotheby’s. That is why the law protects copies<sup>77</sup> and recognizes their status as works of art. A copy may itself be copyrightable if it contains “an original contribution not present in the underlying work” [Merryman and Elsen (1987, p. 184)]. The law states that originality – the very foundation of the notion of a work of art – results only from the execution. Legally, a copy is thus an original work of art, because the personality of the copying artist leaves traces, even if the copy is close to the original. This can be considered the “legal transposition” of a remark by Theo Van Gogh to his brother Vincent, who, while at the St. Remy hospital in 1890, made copies of paintings by Millet: “Your copies of Millet are perhaps the best you have ever produced” (letter 848/733), and “copies as you do them are not copies anymore” (letter 840/T23).<sup>78</sup> One of the legal consequences is that the material execution itself justifies copyright royalties for copyists. There is no difference between copying nature – landscapes, still lifes or portraits without specific additions – and copying preexisting works of art.

Therefore, for the lawyer, the personality of the creator is the main criterion for the justification of copyright royalties. The following case illustrates this position. Some time ago, the owner of a castle in France ordered sculptures for the façade of his castle. When the chateau was opened to the public, the artist asked to be protected by a copyright. The Court of Appeals refused to compensate the sculptor, considering that “the sculptures resulted from a mere repetition and accumulation of decorative designs” and because “the quality of the execution was weak and lacking any global vision”

<sup>76</sup> See *What Is an Original Print?* London: Lumley-Cazalet, 1967, quoted by Melot (1973).

<sup>77</sup> One of the consequences is the difficulty for an artist to copy freely and resell his own work, once the “original” has been sold. This is what the law defines as self-plagiarizing.

<sup>78</sup> Van Gogh also produced copies after Rembrandt, Delacroix, and Gauguin. See Van Tilborgh (1998).



[Colombet (1997)]. The French High Court (Cour de Cassation) reviewed this judgment (November 9, 1993) and determined that the traces of the hand of the copyist were sufficient for the sculptures to be works of art. In this case the copies are seen as originals, and are entitled to royalties. These cases show that judges are supposed to assess the “originality of a copy”, but judgments are not always consistent, and do hardly provide a stable economic foundation to a copyright for copies.

Nevertheless, if originality is assimilated to invention – as opposed to imitation, which characterizes copies [Holtzmann (1996)] – then originality becomes a risky business and should be rewarded by a copyright in the same way as a patent rewards invention. As Meyer (1983) points out, “the great artist has dared to risk failure in order to reveal a new aspect of the universe for us”. This is also what De Marchi and Van Miegroet (1996) found for the 17th century.

## 8. Concluding comments

Copies have always been present, often only marginally but sometimes at the core of the art market in the past. Copies served to convey information to artists and collectors. When this function disappeared, the relative price of copies decreased, copies became aesthetically less valued, but the market did not disappear. Many signs point to their recent revival. Copies are substituted for originals that could be destroyed by bad weather and pollution: 230 sculptures of the façade of the Louvre are replaced by copies. So are the bronze horses atop St. Mark’s in Venice, and the statue of Marcus Aurelius in the Campidoglio in Rome.<sup>79</sup> The Lascaux cave that tourists can visit is a copy. Copies are used to restore originals,<sup>80</sup> and artists keep producing them. There are permanently 84 copyists in the Louvre. Half of them are artists, 25% art students and 25% amateurs. As noted by Muller (1989, p. 147), “copies augment the small and often inaccessible supply of excellent pictures”. Galleries (Troubetskoy in Paris or True Fakes Ltd. in New York) sell freshly-made copies. Museums of copies are being born, and art historians or museum keepers display copies for the purpose of illustrating concepts. There are now specialized auctions at which only copies are sold. The market wants copies, old and new; after Old Masters or Impressionists, after Modern and Contemporary painters.

<sup>79</sup> Sénéchal (1998) discusses the example of the paintings that were hanging in St. Peter’s in Rome and were suffering from humidity. They were replaced by copies on canvas during the 18th century, and later, by mosaic copies.

<sup>80</sup> Copies have contributed in preserving the appearance of lost originals. For example, during a restoration of a painting, a *repentir* by Titian was discovered and authenticated by its copy [Volle (1998)]. Likewise, the restoration of Veronese’s *Pilgrims of Emmaüs* was made possible thanks to the copy that showed that two columns were present in the original painting. There are many cases in which the original is known through copies only. Titian’s *Martyrdom of St. Peter*, lost during the fire that destroyed the San Giovanni e Paolo church in Venice in 1866 is known through a copy by a pupil of Ingres. The practice to replace paintings in churches by copies became frequent during the 18th century.

The market wants exact copies and also interpretative copies. To celebrate the Millennium, Neil Mac Gregor, director of London's National Gallery, chose well-known living painters (Lucian Freud, David Hockney and Jasper Johns, among others) and asked them to produce an interpretative copy of works kept in the museum.

Reviving copies is harmless as long as copies are not sold as originals. Preventing this from happening is made easier since there exist modern techniques for marking artworks as copies. Attention to this issue could prove more fruitful than the current haggling over whether a photograph of the Mona Lisa gives copyright to the Louvre or to the photographer, though the painting belongs to the public domain for centuries. Copying seems to be unavoidable with the rise of digitization that brings about new ways of producing copies, since the virtual is also naturally multiple [Queau (1998)]. Digitization leads to new ways of making copies, authorizing duplication in endless numbers, as well as manipulations and interpretations. This may eventually prove to be very positive by making works much better known and even lead art lovers to order old-fashioned handmade copies.

## Appendix A: Copies and originals sold at auction 1684–1999

This appendix lists painters whose copies (and originals) were sold at different times, and on whom our calculations are based. The first and second numbers that appear after each name refer to number of copies and originals sold during the period.

### A.1. Period 1684–1725

Bambouts (2, 12), Breugel (10, 62), Brouwer (2, 23), Caravaggio (2, 4), Douw (5, 27), Hondokoeter (2, 16), Jordaens (2, 13), Lairese (2, 48), Netscher (2, 20), Ostade (3, 56), Poelenburg (7, 71), Poussin (7, 16), Raphael (9, 4), Reni (3, 11), Rottenhamer (3, 10), Rubens (7, 61), Terburgh (2, 5), Tiziano (6, 23), Van der Werf (2, 13), van Schalke (2, 10), Wouverman (4, 95).

Source: Hoet (1752).

### A.2. Period 1801–1810

Albani (12, 12), Barocci (6, 18), Bassano (7, 9), Berchem (23, 77), Both (7, 32), Bourdon (6, 103), Brouwer (7, 26), Brueghel Jan the Eldest (4, 61), Caracci (31, 3), Caravaggio (9, 25), De Champaigne (12, 48), Corregio (30, 16), Cuyp (11, 64), Domenico (22, 31), Dou (16, 34), Dujardin (13, 38), Dürer (7, 15), Greuze (8, 32), Guercino (13, 40), Laer (6, 18), Le Brun (12, 27), Le Sueur (20, 41), Lingelbach (3, 32), Lorrain, Claude (8, 34), Maratti (9, 45), Metsu (3, 38), Michelangelo (10, 1), Mignard (6, 25), Murillo (11, 46), Ostade (16, 5), Ostade, Isaac (3, 50), Panini (6, 47), Pietro da Cortona (5, 25), Poelenburgh (7, 80), Potter, Paulus (11, 15), Potter, Pieter (1, 3), Poussin (36, 38), Raffaello (79, 8), Rembrandt (27, 65), Reni (43, 45), Ruisdael (4, 14), Rubens

(34, 49), Santerre (4, 19), Solario (5, 8), Subleyras (4, 17), Teniers (50, 210), Terboch (3, 33), Tintoretto (4, 29), Tiziano (20, 34), Valentin de Boulogne (3, 12), Van Dyck (10, 12), Vouet (6, 12), Wouwerman Phillips (9, 70), Wouwerman (14, 7).

Source: [Peronnet and Fredericksen \(1998\)](#).

### A.3. Period 1890–1900

Bellini (4, 8), Boticelli (2, 8), Boucher (12, 26), Campana (2, 3), Chardin (2, 29), David (6, 18), Eyck (1, 3), Gainsborough (2, 50), Giotto (4, 4), Hall (4, 15), Holbein (2, 27), Lancret (5, 16), Mignard (12, 29), Murillo (10, 22), Ostade, Adrien (2, 47), Ostade, Isaac (5, 18), Oudry (6, 23), Porbus (11, 28), Potter (2, 12), Raphaël Sanzio (13, 10), Rembrandt (18, 60), Rubens (18, 71), Tiepolo (3, 37), Van Dyck (7, 50), Van der Heyden (3, 11), Van Orley–Barent Van Brussel (6, 7).

Source: [Mireur \(1901\)](#).

### A.4. Period 1976–1999

Boucher (111, 266), Hals (11, 9), Raphaël Sanzio (570, 11), Rembrandt (173, 55), Rubens (593, 108), Van Dyck (310, 48), Watteau (38, 70).

Source: [Huisman \(2001\)](#).

## References

- Alpers, S. (1988). *Rembrandt's Enterprise*. The University of Chicago Press, Chicago.
- Arnheim, R. (1983). "On duplication". In: Dutton, D. (Ed.), *The Forger's Art*. University of California Press, Berkeley.
- Ayrton, M. (1960). "Introduction". In: Maison, K.E. (Ed.), *Themes and Variations, Five Centuries of Master Copies and Interpretations*. Thames and Hudson, London.
- Belting, H. (1998). "Le fétiche d'art au XXème siècle. Le cas de la Joconde". *Diogenes* 183, 79–102.
- Benhamou, F., Ginsburgh, V. (2001). "Is there a market for copies?". *Journal of Arts Management, Law and Society* 32, 37–55.
- Benjamin, W. (1971). "L'oeuvre d'art à l'ère de sa reproductibilité technique". In: *Oeuvres*, vol. 2. Denoël, Paris.
- Borges J. (1944). "Pierre Menard, autor del Quijote". In: *Ficciones*. Translated in Borges, J., *Oeuvres Complètes*, vol. 1. Gallimard, Bibliothèque de la Pleiade, Paris.
- Bourdon, J., Pontier, J.M., Ricci, J.C. (1996). *Droit de la culture*. Dalloz, Paris.
- Brown, J. (1995). *Kings and Connoisseurs*. Yale University Press, New Haven.
- Burckhardt, J. (1882, 1998). "De l'authenticité des tableaux anciens". In: *Leçons sur l'art occidental*. Fernand Hazan, Paris.
- Caillois, R. (1975). *L'apostat*. Gallimard, Paris.
- Castelnuovo E. (1987). "Attribution". In: *Encyclopedia Universalis*.
- Chamoux, F., Chastel, A., Gauthier, M.M., Grodecki, L., Foucart, A., Schnapper, B., Velmans, T. (1973). "Copies, répliques, faux". *Revue de l'Art* 21, 5–31.
- Colombet, C. (1997). *Propriété littéraire et artistique et droits voisins*. Dalloz, Paris.
- Cuzin, J.-P. (1993). *Copier et créer. De Turner & Picasso: 300 oeuvres inspirées par les maîtres du Louvre*. Réunion des Musées Nationaux, Paris.

- Danto, A. (1981). "Content and causation". In: Danto, A. (Ed.), *The Transfiguration of the Commonplace*. Harvard University Press, Cambridge, MA.
- Delluc, B., Delluc, G. (1984). "Lascaux II: A faithful copy". *Antiquity* 58, 194–196.
- De Marchi, N. (2004). Private communication.
- De Marchi, N., Van Miegroet, H. (1996). "Pricing invention: "Original", "copy", and the valuation of art in early modern Netherlandish markets". In: Ginsburgh, V., Menger, P.-M. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier, Amsterdam.
- de Saint-Simon, C. (1996). *Mémoires*, vol. 1. Gallimard, Paris.
- Desbois, H. (1978). *Le droit d'auteur en France*. Dalloz, Paris.
- Dubus, P. (1992). "Portraits de collectionneurs. A propos de quelques collectionneurs vénitiens au XVIème siècle". In: Dorleac, B. (Ed.), *Le commerce de l'art de la Renaissance à nos jours*. Editions de la Manufacture, Paris.
- Focillon, H. (1919). *Technique et sentiment. Etudes sur l'art moderne*. H. Laurens, Paris.
- Foucart, L. (1973). "Rubens: Copies, répliques, pastiches". *Revue de l'Art* 21, 48–55.
- Francastel, P. (1995). *Bruegel*. Fernand Hazan, Paris.
- Gautier, P.Y. (1991). *Propriété littéraire et artistique*. Presses Universitaires de France, Paris.
- Ginzburg, C. (1989). *Mythes, emblèmes, traces. Morphologie et histoire*. Flammarion, Paris.
- Goodman, N. (1983). "Art and authenticity". In: Dutton, D. (Ed.), *The Forger's Art. Forgery and the Philosophy of Art*. University of California Press, Berkeley.
- Griffiths, A. (1996). *Prints and Printmaking*. British Museum Press, London.
- Grossman, G.M., Shapiro, C. (1988a). "Foreign counterfeiting of status goods". *The Quarterly Journal of Economics* 103, 79–100.
- Grossman, G.M., Shapiro, C. (1988b). "Counterfeit-product trade". *American Economic Review* 78, 59–75.
- Hoet, G. (1752). *Catalogus of Naamlijst van Schilderijen met derzelver prijzen*. Pieter Geraard van Baaten, 's Gravenhage.
- Hol B. (2004). "Reproduction des oeuvres d'art. Les copies dans la sculpture antique". In: *Encyclopedia Universalis*. Online.
- Holtzmann B. (1996). "Reproduction des oeuvres d'art". In: *Encyclopedia Universalis*.
- Hoving, T. (1996). *False Impressions*. Simon and Schuster, New York.
- Hughes, A., Ranfft, X. (Eds.) (1997). *Sculpture and its Reproductions*. Reaktion Books, London.
- Huisman, A. (2001). *Les copies d'oeuvres plastiques*. Manuscript. Université Libre de Bruxelles, Brussels.
- Johnson, W.R. (1985). "The economics of copying". *Journal of Political Economy* 93, 158–174.
- Karpinski, C. (1989). "The print in thrall to its original: A historiographic perspective". In: Preciado, K. (Ed.), *Retaining the Original: Multiple Originals, Copies, and Reproductions*. National Gallery of Art, Washington, DC.
- Koestler, A. (1989). *The Act of Creation*. Arkana, London.
- Lacotte, M., Cuzin, J.-P. (1991). *Dictionnaire de la peinture*. Larousse, Paris.
- Laddie, J. (1996). "Copyright: Over-strength, over-regulated, over-rated?" *EIPR* 5.
- Lafont-Couturier, H. (1998). *Gerôme*. Herscher, Paris.
- Lazzaro, E. (2006). "Assessing quality in cultural goods: The hedonic value of originality in Rembrandt's prints". *Journal of Cultural Economics* 30, 15–40.
- Lessing, A. (1983). "What is wrong with a forgery?" In: Dutton, D. (Ed.), *The Forger's Art. Forgery and the Philosophy of Art*. University of California Press, Berkeley.
- Lobstein, D. (2002). "Les commandes de copies de peintures espagnoles pour l'Etat français, 1840–1880". In: Lacambre, G., Tinterow, G. (Eds.), *Manet Vélazquez. La manière espagnole au XIXème siècle*. Réunion des Musées Nationaux, Paris.
- Maison, K.E. (1960). *Themes and Variations. Five Centuries of Master Copies and Interpretations*. Thames and Hudson, London.
- Martens, D. (2001). "Die Würde des Kopisten oder Hans Holbein. Die Auseinandersetzung mit einem flämischen Vorbild". *Städel Jahrbuch* 18, 165–182.
- Martens D. (2004). "Du Saint Luc peignant la Vierge à la copie des maîtres: la norme en acte dans la peinture flamande des XVème et XVIème siècles". Manuscript. Université Libre de Bruxelles, Brussels.

- Meiland, J.W. (1983). "Originals, copies, and aesthetic value". In: Dutton, D. (Ed.), *The Forger's art. Forgery and the Philosophy of Art*. University of California Press, Berkeley.
- Melot, M. (1973). "Les frontières de l'originalité et les problèmes de l'estampe contemporaine". *Revue de l'Art* 21, 108–118.
- Melot, M. (1985). "La notion d'originalité et son importance dans la définition des objets d'art". In: Moulin, R. (Ed.), *Sociologie de l'Art*. La Documentation Française, Paris.
- Melot, M. (1994). *L'estampe impressionniste*. Flammarion, Paris.
- Melot M. (1995). "La signature des estampes". Paper presented at a conference on *L'écriture du nom propre*. Bibliothèque Nationale de France, Paris.
- Merryman, J.H., Elsen, A.E. (1987). *Law, Ethics and the Visual Arts*. University of Pennsylvania Press, Philadelphia.
- Meyer, L.B. (1983). "Forgery and anthropology of art". In: Dutton, D. (Ed.), *The Forger's Art. Forgery and the Philosophy of Art*. University of California Press, Berkeley.
- Mireur, H. (1901–1912). *Dictionnaire des ventes d'art faites en France et à l'étranger pendant les XVIIIe et XIXe siècles: Tableaux, dessins, estampes, aquarelles, miniatures, pastels, gouaches, sépias, fusains, émaux, éventails peints et vitraux, 7 volumes*. L. Soullié, Paris.
- Montias, J.M. (1982). *Artists and Artisans in Delft. A Socio-Economic Study of the Seventeenth Century*. Princeton University Press, Princeton, NJ.
- Montias, J.M. (1996). "Quantitative methods in the analysis of 17th century Dutch inventories". In: Ginsburgh, V., Menger, P.-M. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier, Amsterdam.
- Mortier, R. (1982). *L'originalité. Une nouvelle catégorie esthétique au siècle des Lumières*. Librairie Droz, Genève.
- Moulin, R. (1992). *L'artiste, l'institution et le marché*. Flammarion, Paris.
- Muller, J.M. (1989). "Measures of authenticity: The detection of copies in the early literature on connoisseurship". In: Preciado, K. (Ed.), *Retaining the Original: Multiple Originals, Copies, and Reproductions*. National Gallery of Art, Washington, DC.
- Mund, H. (1983). "Approche d'une terminologie relative à l'étude de la copie". *Annales d'Histoire et d'Archéologie* 3, 19–31.
- Pastoureau, M. (1991). *Vrai ou faux? Copier, imiter, falsifier*. Bibliothèque Nationale, Paris.
- Peronnet, B., Fredericksen, B.B. (Eds.) (1998). *Répertoire des tableaux vendus en France au XIXe siècle, 1801–1810, Los Angeles: The Provenance Index of the J. Paul Getty Trust, 2 vol.* Getty Museum, Los Angeles.
- Queau, P. (1998). "Multiplicités virtuelles". *Diogène* 183, 102–112.
- Schaer, R. (1993). *L'invention des musées*. Gallimard et Réunion des Musées Nationaux, Paris.
- Schwartz, H. (1996). *The Culture of the Copy*. Zone Books, New York.
- Seguy Duclot, A. (1998). *Définir l'art*. Odile Jacob, Paris.
- Sénéchal, P. (1998). "Objets uniques/objets multiples. Autour de la copie. Introduction". *Diogène* 183, 2.
- Smith, A. (1795, 1982). *Essays on Philosophical Subjects*. Liberty Classics, Indianapolis. The Glasgow edition of the works and correspondence of Adam Smith.
- Stanton L.L. (2002), *Comments on Steinberg vs. Columbia Pictures Industries, Inc.* [www.law.cornell.edu](http://www.law.cornell.edu).
- Van Tilborgh, L. (1998). "Van Gogh disciple de Millet". In: *Van Gogh Millet. Réunion des Musées Nationaux, Paris*.
- Volle N. (1998). "La restauration des peintures et la copie". *Conférence De l'original à la copie, Paris, Cité de la Musique, 19–20 November 1998*.
- Werness, H.B. (1983). "Han Van Meegeren fecit". In: Dutton, D. (Ed.), *The Forger's Art. Forgery and the Philosophy of Art*. University of California Press, Berkeley.
- Winckelmann, J.J. (1768, 1952–1957). "Briefe, 4 volumes". De Gruyter, Berlin. Edited by H. Diepolder and W. Reehm.

## COPYRIGHT, ART AND INTERNET: BLESSING THE CURSE?\*

PATRICK LEGROS

*ECARES, Université Libre de Bruxelles and CEPR, Belgium*

### Contents

Abstract	286
Keywords	286
1. Introduction	287
2. The example of MP3s and software	290
3. A market for works of art	296
3.1. Appropriability: Competition	299
3.2. Appropriability: Market power	301
4. Issues and conclusion	304
References	306

\* Many thanks to Victor Ginsburgh for his patient editorial work and for convincing me that economists should be interested in arts. I benefited from comments by William Baumol and Andrew Newman.

**Abstract**

The new technologies of digitalization and the Internet threaten the market positions of artists and intermediaries. Artists because the technology of production of works may be readily accessible and craftsmanship may no longer be a defining characteristic of art. Intermediaries because their rents are linked to entry barriers in the distribution market. This curse of new technologies may be a blessing in disguise since it also increases the possibilities of production, of distribution and the emergence of new works of art. The system of intellectual protection gives market power to artists and the economic literature has analyzed the tradeoff between the dynamic inefficiency generated by this market power and the need to preserve the incentives for creation. We review this literature and some of its recent applications to artistic, and more generally intellectual, creation. Even if artists can capture perfectly the market value of the future home production by consumers, they may favor a strong copyright regime that prevents consumers from using their home production. Intermediaries and artists may want to limit competition in order to increase the rents brought by the indivisibility of creative ideas. The preferences of artists for strong or weaker form (e.g., licensing of rights for home production) of copyright may be related to their creativity.

**Keywords**

copyright, appropriability, licensing, intellectual protection, Internet

*JEL classification:* D2, D13, D45, O3

Beauty, however, in its general aspect, is the inseparable characteristic of the idea when it has become known. In other words, everything is beautiful in which an idea is revealed; for to be beautiful means no more than clearly to express an idea.

Schopenhauer (2004-eBook edition)

In the last analysis, the artist may shout from all the rooftops that he is a genius: he will have to wait for the verdict of the spectator in order that his declarations take a social value and that, finally, posterity includes him in the primers of Artist History.

Marcel Duchamp (1966)

## 1. Introduction

At the time photography was invented, the technology was expensive, difficult to use and required specialized skills and craftsmanship. Because many painters at the time were doing portraits, they saw the danger of the new technology for their activity. The folk history credits the painter Paul Delaroche to have said after seeing the Daguerreotype “from today, painting is dead”.<sup>1</sup> Other artists embraced the opportunity to use the new medium and indeed, a movement developed quickly that defined photography as art. When George Eastman invented the “clic-clac Kodak” in 1868, photography became widely accessible; while there were some issues of craftsmanship, the act of taking a picture became trivial enough that it took effort by photographic artists to preserve their identity. Some have argued that the *pictorial movement* emerged in response to this democratization of the access to the technology. The recent emergence of digital photography has made the marginal cost of taking and viewing pictures rather trivial; despite this democratization of the technology, photography as an art form is still alive, present in large museums and taught in art departments in prestigious universities.

Before the Gutenberg press, the church had a monopoly on the stock of original writings and monks were the main artisans for reproducing these works, often by using techniques and crafts that required years of training. The Gutenberg press rendered this craftsmanship unnecessary for copying or for production of new books.<sup>2</sup>

The emergence of new technologies is therefore both a blessing and a curse for art. The blessing is that more opportunities for artistic creation are available. The curse is that more people have access to it. It is a curse because issues of craftsmanship tend to be less important, and a work of art may now have to be distinguished from its look

<sup>1</sup> Some historians of photography like Robert Leggat (1999) claim that in fact Delaroche was a supporter of photography. He had been commissioned by the French government to present a report on the Daguerreotype where he wrote “Daguerre’s process completely satisfies all the demands of art, carrying essential principles of art to such perfection that it must become a subject of observation and study even to the most accomplished painters.”

<sup>2</sup> See Chapter 8 by Benhamou and Ginsburgh in this volume.



alike by another dimension than craftsmanship. That aesthetics or craftsmanship is not a necessary characteristic of a work of art is well espoused by philosophers and artists.<sup>3</sup> Duchamp's Readymades are an extreme illustration of this since common objects like a urinal, a bicycle wheel or a snow shovel can become works of art. When art is not necessarily linked to aesthetics, the definition of art becomes somewhat of a challenge.

While economists are ill equipped to philosophize on art, they and scientists in general are also in the business of creating ideas, and transmitting these ideas to peers and the public at large. In a modest, or probably immodest way, this activity is sometimes compared to that of artists. As producers, most of us recognize the difficulties to come up with truly original ideas, to write these ideas in a way that will be transparent and convincing to our readers. As consumers we sit through seminars, read working papers or published papers and we hope that these activities will give us some insights into the message that the author wants to convey. Sometimes, we enjoy a speaker's charisma, sometimes we enjoy the writing of a paper, but eventually we are interested in the underlying idea. We train students for many years hoping that one day they will be able to read and understand our papers or that they will themselves be able to contribute to the production of ideas. In the process, we try to make a living and get credit for our contributions to the field.

Scientists are different from artists however, both on the demand side and the supply side for their creations. On the demand side their "natural" markets differ: new working papers are in general consumed by scientists and researchers in the field, but new paintings are bought and appreciated by non-painters. On the supply side, there is an intentionality behind scientific production: scientists try to communicate precise ideas, results and they follow well established methodologies for doing so; there is not necessarily intentionality for artistic creation. As [Duchamp \(1959\)](#) notes

In the creative act, the artist goes from intention to realization through a chain of totally subjective reactions. His struggle toward the realization is a series of efforts, pains, satisfaction, refusals, decisions, which also cannot and must not be fully self-conscious, at least on the esthetic plane.

Hence, while both scientific and artistic communications are imperfect and require interpretation, in art this interpretation is complicated by the heterogeneity of the potential consumers and the sometimes fuzzy knowledge that the artist has about his or her motives for creation. [Danto \(1986\)](#) views this as the main hurdle in defining art:

There are two sorts of mistakes the concept of art gives rise to, one of which is philosophical and the other merely critical. The first is to interpret something which is not in candidacy for art, and the second consists in giving the wrong interpretation of the right sort of thing.

Artists and philosophers refer to a "missing dimension" – the "idea" in Schopenhauer's opening quote – to explain the imperfect communication between the artist and

<sup>3</sup> I refer the reader to [Chapter 5](#) by Roger McCain in this volume and to [Danto \(1986\)](#).

the difficulty for outsiders to interpret artistic creation. Duchamp (1966) has even defined an “art coefficient” to capture the level of imperfection in artistic communication

The result of this struggle is a difference between the intention and its realization, a difference which the artist is not aware of. Consequently, in the chain of reactions accompanying the creative act, a link is missing. This gap, representing the inability of the artist to express fully his intention, this difference between what he intended to realize and did realize, is the personal ‘art coefficient’ contained in the work. In other words, the personal ‘art coefficient’ is like an arithmetical relation between the unexpressed but intended and the unintentionally expressed.

Because interpretations become so important, some have argued that they are themselves works of art [Danto (1986)]; we should not be surprised since common parlance refers to singers or pianists as *interpreters*!<sup>4</sup> Achieving the interpretation of a work of art requires knowledge, appreciation of the historical context in which the work was created. And here the Internet’s curse that more people use the same technology as the artists can be also a blessing because Internet may facilitate the diffusion of the knowledge needed for interpretation.

Hence, markets in which ideas are valued, either directly, like in research, because they contribute to the stock of knowledge and facilitate the production of new ideas, or indirectly, like in art, because they change our perception of the world, have two characteristics that distinguish them from other markets:<sup>5</sup>

- H. The consumer values the work of art, that is the physical or digital good by which the “idea” is embodied, both for its aesthetics and for the message it brings about the idea;
- I. The transmission of the idea from the creator to the consumer is not immediate, is subject to noise both during the process of creation and during the process of interpretation; furthermore interpretation requires in general effort on the part of the consumer.

There are other characteristics that are not specific to art markets. Many of them are covered in this Handbook: the rise of superstars, the use of auctions as allocation mechanisms, the two-sided nature of the market and the gatekeeper role of intermediaries, the difficulty to specify complete contracts. The Internet and the possibility of digitalization of works of art have magnified some of these aspects, in particular the possibility to distribute at a rapid pace digital works, to reproduce or even modify existing works, and for artists to bypass current gatekeepers.

<sup>4</sup> There is a famous quote on the pianist Glenn Gould: at the end of a dispute with a fellow pianist about a piece by Bach, Gould is credited to have said “Ok you will do it your way, I will do it *his* way”.

<sup>5</sup> See also Chapter 5 by Roger McCain and Chapter 16 by Tony Bryant and David Throsby in this volume. Properties H and I are actually reasonable assumptions for most processes of communication; language is inherently ambiguous and skilled speakers use rhetoric to present facts or events in a new light. Despite the importance of information in today’s economic analysis, most of it continues to assume that communication is a frictionless process, not subject to interpretation, interference or noise. Recent exceptions are Legros and Newman (1999, 2002), Dewatripont and Tirole (2005).

This chapter is by necessity incomplete. As George Stigler had noted for the theory of regulation in the 70s, the proper time to survey a literature is after the subject is developed and a consensus is reached in the field. The theory of regulation did not achieve this in the 70s. We are far from this goal for art and Internet; the Internet is in its teens, and research on its economic and cultural effects is still in its infancy.<sup>6</sup> I take the view here that it is because the Internet makes entry easier for artists and consumers that economists should be interested in its effects. I will therefore focus on the tradeoffs between the blessing (a larger market and more possibilities for creating works of art) and the curse (more competition) of the Internet for art.

Most of the discussion in the media and in the academic literature on this topic has been about the *appropriability* by artists and intermediaries of the revenues generated from creative ideas. The curse of the Internet and the new technologies is the ease with which some digital works can be replicated and distributed. Many like the monks a long time ago or the painters more recently fear that this will eliminate incentives for creation or distribution of art. The leading example has been copyright infringements for digital music and software. I review this case in Section 2. Music is obviously only used as an example of art form making use of the Internet; digital painting, poetry, photography, video are other prominent examples of works of art being distributed or created on the Internet. The issues of copyright protection, incentives for creation, the tradeoff between market expansion versus competition effects are common to all these works of art. In Section 3, I will then take a more abstract approach and use a model based on Boldrin and Levine (2002a, 2002b) and Quah (2002a) to argue that preferences of different participants in the market for strong copyright may have little to do with social efficiency, including incentive provision for new creative ideas. I continue in Section 4 with a faster tour on other issues linked to the market expansion effect of the Internet and a conclusion.

## 2. The example of MP3s and software

Traditional gatekeepers, the music majors, are fighting for strict copyright laws, by which consumers are not allowed to replicate, distribute, and much less modify a work, unless it is for “fair use”. This fight takes the form of legal battles, like against NAPSTER and other peer-to-peer networks where individuals exchange freely MP3s or videos. The software and music industries claim large losses from copyright infringement – or piracy as it is called in the media.<sup>7</sup> Some economists question the magnitude

<sup>6</sup> See however Brynjolfsson (2002), Kahin and Varian (2000), Peitz and Waelbroeck (2003b), Quah (2002b, 2003).

<sup>7</sup> Most of the empirical work is preliminary and the magnitude of the measured effect of piracy sometimes contradictory. Peitz and Waelbroeck (2003a) find a decrease of 1.76% in CD sales worldwide and that MP3 account for 25% of the decrease in CD sales in the top ten markets between 2000–2001. Hui and Png (2003) find that CD loss may be 15% higher than the 1999 industry estimate (software and music industry estimated

of these losses arguing that if copyright laws were perfectly enforced, many users would decide not to consume [Gayer and Shy (2001a, 2001b)], or that downturns in CD sales correspond to normal cycles or substitution to other types of entertainment [Liebowitz (2002)].<sup>8</sup>

The legal battle is on-going. For instance, the maker of iMesh file-sharing software agreed to pay \$4.1 million to the recording industry for copyright infringement. On August 20, 2004, a U.S. federal appeals court ruled that makers of two leading file-sharing programs (Grokster and StreamCast) are not legally liable for copyright infringement by the users of their software. This may force the music industry to fight directly against the users. And they have some successes on this front; for instance, on August 26, 2004, the FBI seized for the first time software and computers of users of a P2P network in Texas. According to an article of August 25, 2004 in the *Financial Times*, record companies have also launched legal actions against karaoke bars in big Chinese cities, demanding damages for the infringement of their licensing rights.<sup>9</sup> But many see this legal fight as hopeless.<sup>10</sup>

The industry is also fighting in more intrusive forms. For instance, by uploading bad quality MP3s on peer-to-peer networks in order to decrease the benefits of using these networks, or even – as in a recent proposal – by using software agents to sabotage the computers of the users of these networks [Corbett (2003)]. Given the wealth of information stored on computers and their other uses, the social cost of this interference may be a magnitude greater than the cost to the music industry of copyright infringement.

In parallel, the music industry uses the Internet to develop new distribution systems; the entry of Apple Computer on the online market (iTunes), where songs are sold for 99¢, has been quickly followed by other recording companies like Sony but also non-recording companies like Wal-Mart Stores, or firms engaged in delivering media software like RealNetworks and Microsoft.<sup>11</sup> MP3s have given rise to complementary hardware manufactured by some of the very firms – like Sony – fighting copyright infringement.<sup>12</sup> Software solutions, The Digital Rights Management (DRM),

its loss from piracy at \$16 billion for 1999). It is not clear how these studies control for substitution effects with other types of entertainment or correct for business cycles.

<sup>8</sup> A 2002 study by Jupiter Media Metrix found that people who download intensively from the Internet has 75% chances to have spent more on CDs than others.

<sup>9</sup> Background information, new legal actions and history of legal cases are available, for instance, on the site of the Recording Industry Association of America, [www.riaa.com](http://www.riaa.com).

<sup>10</sup> For instance, following NAPSTER's legal death, the number of P2P networks has multiplied.

<sup>11</sup> By all accounts, this market is quite small (accounting for less than 2% of all music sales in the U.S.) but is quite competitive; for instance, RealNetworks cut recently the price of its downloads to 49¢ in anticipation of Microsoft's entry (New York Times, August 30, 2004).

<sup>12</sup> This has led to proposals to tax hardware, like MP3 players or blank CDs and DVDs in order to capture some of the losses from copyright infringement (for instance, a tax of the order of 60¢ is levied in France that is put in a fund for artists). Gayer and Shy (2001a, 2001b) offer a model along these lines; they show within their model that such taxation will be ineffective.

limit copying and sharing by users.<sup>13</sup> A literature has developed the idea that illegal copying may increase the demand for complementary software or hardware<sup>14</sup> or that the industry may in fact benefit from copying because of network effects.<sup>15</sup> While suggestive, the effects are probably weak if we refer to a “revealed preference argument”: the resources spent by recording firms in fighting piracy.

Artists are quite divided on the issue of strong copyright. Some artists have organized themselves to fight against copyright infringement (e.g., *Artists coalition against piracy*, <http://welcome.to/acap>, signatories include Elton John). Others embrace the new technology. Some see an opportunity to become known and eventually sign a contract with one of the majors (examples abound of unknown artists who were able to sign their first major contract after having distributed for free their work).<sup>16</sup> This phenomenon is not limited to new artists, since well established artists, and sometimes stars, also distribute their work on the Internet under weak copyright form.<sup>17</sup>

That established firms and some artists want to use the legal system and the new technologies at their advantage, preserve their rents, is easy to understand. That the legal system should follow their lead is another issue. Proponents of strong copyright laws claim that without them there will be significantly less artistic creation or distribution. Economists are familiar with the precept that strong property rights are needed for inducing innovation: otherwise the innovator is not able to collect enough revenues ex-post to cover his ex-ante investment. There is then a tradeoff between ex-ante and ex-post efficiency [Nordhaus (1969)]: giving monopoly rights ex-post creates inefficiencies but is needed in order to provide (efficient) ex-ante incentives to invest.<sup>18</sup> This tradeoff is similar to the tradeoff in industries that exhibit increasing returns to scale.<sup>19</sup> firms cannot make positive profits when pricing at marginal cost and therefore pricing above

<sup>13</sup> Not surprisingly, DRM has become the target of hackers. See, for instance, DRM Watch Staff (2004) documenting recent successes in cracking DRM protection in Apple iTunes.

<sup>14</sup> Conner and Rumelt (1991), Shy and Thisse (1999) give theoretical arguments; Givon, Givon and Muller (1995) provide an empirical analysis of this effect and find that software piracy may boost demand for the legal software. The effect may also go in the other direction. For instance, in the January 7, 2005, issue of the *Financial Times*, Peter Jamieson, chairman of the British Phonographic Industry – representing leading labels and music distributors in England – is cited as noting a “strong move in the digital market” coinciding with strong sales of the Apple iPod and other MP3 players.

<sup>15</sup> Takeyama (1994, 1997), Shy and Thisse (1999).

<sup>16</sup> In France, the most famous example is the singer Lorie, who signed a contract with Sony Music under the label EGP after having been known through Internet downloads. In the US, the pop band Fisher signed a contract with Farmclub and Interscope after having had more than a million downloads of their album on MP3.com.

<sup>17</sup> David Bowie may have been the first to distribute the music of a new CD “Hours . . .” on the Internet before it was even made available in stores; see *The Economist*, September 9, 1999.

<sup>18</sup> See Landes and Posner (2004) and Scotchmer (2004) for state of the art economics approach to intellectual property protection.

<sup>19</sup> The parallel is not quite fair since as noted by Boldrin and Levine (2002a) the production of MP3s does not exhibit increasing returns: the cost of creation of the song by the artist is sunk rather than fixed at the time a CD or an MP3 is distributed.

marginal cost is needed. But the usual response to the problem in these industries is to *regulate* – or give subsidies to the industry – rather than to give total and uncontrolled monopoly power to one firm.<sup>20</sup>

While it may be difficult to assess the Nordhaus effect for artists, we can turn to recent studies on patents. [Lerner \(2003\)](#) analyzes patent protection over 150 years and 177 policy changes. He shows that stronger patent protection has few positive effects on patent applications by applicants from the country in which the change was made (he found in fact a negative relationship after correcting for aggregate effects), which suggests that the Nordhaus effect is not operative. Cross-section effects are consistent with the theory, however, since there is a more important effect of shifting to stricter patent policies when starting from weaker patent protection or greater development.<sup>21</sup>

Closer to our topic, two studies analyze the effect of the change in patent protection for software in the 1980's. [Bessen and Maskin \(2000\)](#) present a theoretical argument to show that if innovation is sequential and complementary, patent protection may reduce overall innovation and that the more competitive the market, the more inefficient patenting is.<sup>22</sup> Using as a natural experiment the extension of patent protection to software in the 1980's, Bessen and Maskin show that while the “Nordhaus effect” would suggest that R&D intensity and productivity should have increased among patenting firms, they did not. [Bessen and Hunt \(2004\)](#) argue that the significant increase in software patenting after the 1980's (software patents representing now 15% of all patents) is mainly due to large manufacturing companies since only 5% of these patents belong to software publishers. This suggests a strategic motive for patenting and Bessen and Hunt find evidence that software patents and R&D at the firm level are substitutes. [The methodology and results in [Bessen and Hunt \(2004\)](#) are criticized by [Hahn and Wallsten \(2003\)](#).]

One interpretation of these two studies is that stricter property right laws rather than inducing more innovation may in fact enable firms with market power to substitute patenting for R&D in order to generate rents. By most accounts software is not art,<sup>23</sup> but these studies put into question the claim that strong copyright laws are needed for artistic production. They also suggest that the social problem of providing incentives

<sup>20</sup> [Romer \(2002\)](#) proposes government direct funding, [Shavell and van Ypersele \(2001\)](#) propose the use of rewards based on realized sales as alternative means for incentive provision.

<sup>21</sup> Lerner notes that he is not able to capture the potential impact of other policy tools used in parallel to patent policy, like offering prizes for discoveries.

<sup>22</sup> A key assumption of the model is that only firms that are active in the market can copy. See [Green and Scotchmer \(1995\)](#) for the question of surplus division when there is sequential innovation. A related issue is the possibility for artists to cooperate in the creation of works of art; the Internet allows the multiplication of such cooperative efforts – as the open source movement has illustrated. Economic issues linked to patenting in cooperative production are analyzed in [Scotchmer and Green \(1990\)](#).

<sup>23</sup> However, consider the experiment suggested by Jean-Luc Moulène, a French photographer: use a digital camera to take a picture of a green pad on a red background, and take another picture of the same pad on a background of a different color. The color of the pad will not be the same in the two pictures; this is because the digital process functions on the basis of harmonic equilibrium controlled by the software. This leads Moulène to conclude that the true creative process in digital picture taking is the software.

for innovation should not be confused with the protection of rents of intermediaries, or rents of established artists. And there are reasons to believe that the discussion in the literature and the media on the costs and benefits of piracy contributes to the confusion.

First, there is a focus on the profit losses of record companies or established artists. There are analyses of the potential loss of recording companies due to piracy, but I do not know of studies evaluating the minimum future revenue needed for artists to be induced to create. We have no appreciation of the cost of copyright infringement on artistic creation.

Second, most of the debate about copying and copyright infringement assumes that CDs and MP3s are the only works of art embodying the creative idea of the artist. However, different works of art change the proportion of revenues going to the artist versus the recording company, or outside financiers. It would be interesting to analyze whether MP3 copying has positive effects for attendance at public concerts, or for viewer interest in TV shows, video clips, since these are also significant sources of revenue for the artist.

Third, the possibility for the artist to appropriate the gains from online diffusion, whether legal or not, in a non-monetary way is generally ignored, at least in empirical work. For instance, since attendance at public concerts is a function of the artist's reputation, online distribution, whether controlled or not by the artist, will contribute to this reputation. Hence even if the artist cannot appropriate the monetary gains from the distribution of MP3s he or she may appropriate the reputational gains (that may eventually turn into monetary benefits). The "revealed preference" argument we used earlier that recording companies do not favor weak copyright cannot be applied to dismiss this possibility. Indeed, for public concerts, revenues are often captured directly by the author, while in the case of CD sales, a large proportion of the revenues is captured by the recording company. Hence there is no contradiction in having recording companies oppose weak copyright – to preserve their revenues from CD sales, while artists may favor weak copyright – in order to benefit from reputational effects of online diffusion, and increase their revenues from other works of art. This is an empirical issue.

Non-monetary incentives may matter as much as direct monetary incentives for artistic creation.<sup>24</sup> What is puzzling in a world where only monetary incentives are at stake, becomes relatively clear when other motives, like reputation building and career concerns complement monetary incentives. For instance, in their study of the open source software movement [Lerner and Tirole \(2002a, 2002b\)](#) point to career concerns as a motive for cooperation among developers. Their empirical analysis of 40,000 open source projects suggests that those that are consumer oriented have restrictive licenses, while those that are developers oriented (or commercial operating systems) have less restrictive licenses. These results are consistent with developers trading off the benefits of being recognized for having contributed significantly to a project (the career concern) versus the risk that someone will appropriate the collective work (that they dub "hijacking").

<sup>24</sup> See [Chapter 16](#) by Bryant and Throsby in this volume.



Recent initiatives for distribution of software and artistic or intellectual work weaken the copyright restrictions while also preventing (or trying to prevent) hijacking. The *GNU General Public License* allows users to copy, modify, create and sell derivative products on the condition that these derivatives acknowledge their origin – give credit to the previous code writers – and contain the same licensing terms. *The creative commons license* for intellectual work is in spirit similar but can be made more restrictive, for instance, concerning remixing or changes of original songs, or commercialization of derivative works.<sup>25</sup> The main restriction in both cases is the obligation to acknowledge the original creator in any derivative work or in any use of the original work. Violations of these licenses are not documented yet, which may indicate that the stakes involved by infringing on these licenses are for the moment weak.<sup>26</sup> These licenses show the possibility for artists to appropriate credit for their work – at least in the non-monetary sphere – even if it is made widely available and if the possibilities of replication or modification are not costly.

On Magnatune.com it is possible to download MP3s for direct consumption,<sup>27</sup> but also to buy different types of licenses: some are for corporate use, some are for remix and derivative works.<sup>28</sup> On Digital Art Auction (<http://www.digitalartauction.com>), the business model is similar to a subscription system. The artist auctions the master copy of his work of art. Bidders propose in the form of a pledge or bid a retail price they'd be willing to pay, and once the maximum revenue available from these bids meets the artist's requirements, the artist chooses the price, receives the revenue from all successful bidders who then receive a copy of the artwork. Another example of innovative financing is the "Bowie bonds". In 1997, David Bowie issued \$55 million worth of bonds that were bought by Prudential Insurance Co. The bonds were backed by future royalty payments on the publishing rights and master recordings of some of Bowie's tunes.<sup>29</sup>

Whether these alternative business models will succeed is still unclear, but they suggest that ease of copying and weak copyright may not prevent an artist from collecting

<sup>25</sup> See the description of these licenses for "artists" – musicians, writers, filmmakers, photographers and (!) scholars – at <http://creativecommons.org/learn/artistscorners/>. See also Lessig (2004).

<sup>26</sup> The technology also allows some protection against infringement. For instance, Commons licensed works have software tags attached to them and these tags will be present in derivative works. If the stakes for infringing are high enough someone will probably find a program to remove the tag [see Legros and Newman (1999) for related ideas in a contract environment].

<sup>27</sup> Interestingly the price is between \$5 and \$18 and the consumer chooses how much to pay; the average price paid is around \$8. This may comfort some economists' view, e.g., De Long and Froomkin (1999), that tomorrow's economy will be based on gift exchange or that business models will be similar to fund raising campaigns.

<sup>28</sup> Example: "This license permits you to use any number of audio samples from a single song by "American Baroque", to create a single song of your own. You can also make remixes or other derivative works. If you make several songs with our samples, you will need a separate license for each song you create (note that alternate versions [i.e. remixes] of your song are considered one song).

<sup>29</sup> The bonds had an average life of 10 years, and were priced to pay investors 7.9 percent interest (*Bloomberg News*, March 2, 1997).



revenues. This point has been made in the literature on copying [Besen and Kirby (1989), Liebowitz (1985), Johnson (1985), Novos and Waldman (1984)], and sharing [Ordober and Willig (1978), Bakos, Brynjolfsson and Lichtman (1999), Varian (2000), Klein, Lerner and Murphy (2002)].<sup>30</sup> Copying may be beneficial to producers because while it creates more competition, it also increases the willingness to pay of users who anticipate the benefits of copying and sharing. In these models the technologies of production and of distribution are fixed and exhibit increasing returns; it is then cumbersome to capture the market expansion effect of the Internet. A model developed by Boldrin and Levine (2002a) is better fit for capturing the two effects.<sup>31</sup> Legros (2005) uses their approach to formalize a market for intellectual creations and to analyze who among the market participants will “bless the curse”. I turn to this model now.

### 3. A market for works of art

There are two periods and a representative consumer with a subjective discount rate  $\delta \in (0, 1]$  and a concave and increasing utility  $u(c)$  for consumption (assume the Inada condition  $u'(0) = +\infty$ ). If the price of consumption is  $p$ , the consumer consumes  $p = u'(c)$ , and the demand function is given by the solution  $c = D(p)$  to this equation; the elasticity of demand is  $\varepsilon = -1/r(c)$ , where  $r(c) = -c \frac{u''}{u'}$  is the index of relative risk aversion; demand is elastic when  $r(c) \leq 1$ . I assume that  $u$  exhibits increasing relative risk aversion, that is

$$r(c) = -c \frac{u''(c)}{u'(c)} \text{ is increasing in } c. \quad (1)$$

In the first period an artist creates  $s_1$  works of art. These are distributed and sold by firms at zero cost; the asset price of works is  $q_1$  in the first period and therefore the artist has revenue  $q_1 s_1$ . Consumers choose how much to consume ( $c_1$ ) in the first period and pay a price  $p_1$  for consumption; the works  $s_1 - c_1$  that are unsold are used by firms to create in period 2 other works of art at a rate  $\beta$ . Consumers have access to a “home production” technology that transforms one work into  $\alpha$  units; hence home production does not conflict with the act of consumption: if  $c$  units are consumed, the consumer has utility  $u(c)$  and obtains  $\alpha c$  units in period 2. I will make the reasonable assumption that  $\beta \geq \alpha \geq 1$ .<sup>32</sup> When  $\alpha$  increases, consumers have access to a technology that becomes

<sup>30</sup> For instance, Varian (2000) analyzes the incentives of a producer to rent or to sell information goods when sharing is facilitated by the formation of “clubs”, that is groups of agents who agree to share any good purchased or rented by one of the members. Libraries, video clubs are such instances of clubs. He shows that profits can increase with sharing when the transaction cost of sharing is small and when the content is viewed a few times only. Varian focuses on monopoly providers while Ordober and Willig (1978) consider Ramsey prices.

<sup>31</sup> See also Hellwig and Irmen (2001).

<sup>32</sup> Note that there is limited rivalry in the sense that once a firm sells consumption flow  $c$ , it cannot use the  $c$  units to create copies. See Quah (2002a, 2002b) for a non-rivalry example where he assumes  $\alpha > \beta$  in order to capture the idea that reproduction is faster with dissemination.

similar to that of firms: at the time of the Daguerreotype,  $\alpha$  was small and probably equal to zero, at the time of digital photography,  $\alpha$  is large.

A strong copyright regime is one in which the firm can prevent, via legal or technical constraints, consumers to use their home technology. A weak copyright regime is one in which the firm cannot – or does not want to – prevent consumers from using that technology.

Note the two main assumptions until now:

- (i) The cost of innovation (coming up with the creative idea) is sunk rather than fixed, i.e., there are constant returns to scale rather than increasing returns in the production of works of art.
- (ii) Works of art are not divisible and the “creative idea” is embodied into a medium that can be replicated only if the medium is made available; moreover producing additional works of art on the basis of existing ones requires time and the rate of production is finite. While finite, this rate can be made arbitrary large; finiteness corresponds to the concept of *finite expansibility* in David (1992) [see also Quah (2002a, 2002b)].

Assumption (i) is a significant departure from the usual assumption of increasing returns made in the literature on innovation [see however Chapter 11 by Baumol in this volume and Sutton (1991)]. Assumption (ii) captures the view that the creative ideas, and later their interpretation, have to be embodied into a support before being replicated or used to produce other works of art. Since there must be a productive activity for replicating, this activity comes at a cost, modeled here in terms of time.

Following the discussion leading to conditions H and I, let us define a work of art by a two-dimensional point  $(x, y)$ .<sup>33</sup> The first component  $x$  is never observed and corresponds to the underlying creative idea. The second component  $y$  is the observable part and corresponds to the physical or digital properties of the work (e.g., painting, recording on a CD, working paper of an economic model). Interpreting a set of  $m$  works of art  $\{(x, y_i); i = 1, \dots, m\}$  that are known to be based on the same creative idea (e.g., Duchamp’s Readymades, or a research question in science) amounts to inferring the non-observable component  $x$  from the observable components  $\{y_i\}$ .

Artists are distinguished by the number of creative ideas that they have: an artist with a large (small) number of creative ideas will be able to create only a small (large) number of works of art per creative idea. Hence if John has  $n$  creative ideas  $(x = 1, \dots, n)$  and a production capacity of  $k$ , he will have  $s_1 = k/n$  works of art  $(x, y_{xi}), i = 1, \dots, k/n$ , produced on each of these ideas where I assume that the artist never produces twice the same work of art, that is  $y_{xi} \neq y_{\hat{x}j}$  for all  $xi \neq \hat{x}j$ . Interpretation is facilitated the larger  $s_1$ : observing more works of art that are distinct in their observable components improves interpretation and inference of the underlying idea. In the words of Duchamp (1966),  $s_1$  is an index of the “art coefficient” that can be associated with the production

<sup>33</sup> Our view that consumers value the characteristics of the object and their use for home production is in the spirit of the work of Lancaster (1966) or Becker (1976).

of the artist. Because  $s_1$  is also an index of the number of creative ideas, larger values of  $s_1$  are associated with less creative artists.

For low values of  $s_1$ , interpretation is difficult but once obtained, it is easy to create new works; as  $s_1$  increases, interpretation is easier, but it becomes more difficult to create new works. Therefore, the rate  $\beta(s_1)$  available to firms for producing new works is an inverted U-shaped technology: increasing for low values of  $s_1$  and decreasing for high values of  $s_1$ . I assume throughout that the replication rate in the home production technology is bounded above by the rate available to artists:  $\alpha \leq \beta(s_1)$ . Until necessary I simply write  $\beta$  instead of  $\beta(s_1)$ .

Once interpretation is achieved, technology (or craftsmanship) can be used to produce additional works based on the same creative idea  $x$ . Each such idea creates its own market, with consumers having utility  $u(c)$  for consuming  $c$  different works: a new work has value for consumers only in that its observable component is different from existing observable components. This is obviously an extreme view; it would be the case if aesthetics has no value for consumers who care only about understanding the creative idea. The qualitative results will not change as long as the value that consumers put on new works of art is larger than for exact copies.

Artists are “driven” in the sense that they always produce initially (that is in period 1) the maximum number of works.<sup>34</sup> The distribution of these works is done by firms which may decide not to distribute all works, to smooth consumption over periods or to keep prices high if they have monopoly power.

Assumptions (i) and (ii) are sufficient for the artist to appropriate positive revenues even if there is competition on the market for works of art, and even if the home technology becomes as good as the firm technology (that is  $\beta - \alpha$  gets small), or if the firm technology improves (that is  $\beta$  gets large). The assumption of a competitive market is obviously extreme: having a new idea suggests indeed that there are few immediate substitute ideas and works embodying a similar idea in the market. Since under conditions (i) and (ii) the inventor will not be expropriated, it is fair to wonder why the holder of such innovation would behave competitively. The competitive benchmark is however useful because if the artist can have revenues large enough to induce creative activity under competition, this makes the case for weak copyright laws even stronger when there is market power. But considering the monopoly assumption provides additional insights into the reasons for an artist or a firm to favor weak rather than strong copyright.

<sup>34</sup> While we allow for strategic release to the market by the firm distributing the works of art of the artist, strategic release by the artist himself may be important to model. For instance, Duchamp (1966) writes “I realized very soon the danger of repeating indiscriminately this form of expression and decided to limit the production of “Readymades” to a small number yearly. I was aware at that time, that for the spectator even more for the artist, art is a habit forming drug and I wanted to protect my “Readymades” against such a contamination. Another aspect of the “Readymade” is its lack of uniqueness . . . the replica of the “Readymade” delivering the same message, in fact nearly every one of the “Readymades” existing today is not an original in the conventional sense.”

### 3.1. Appropriability: Competition

The representative agent's welfare maximization problem is

$$\max_{c_1} u(c_1) + \delta u(\beta(s_1 - c_1) + \alpha c_1), \quad c_1 \in [0, s_1],$$

where  $s_1$  is the initial asset holding in the economy and  $s_2 = \beta(s_1 - c_1) + \alpha c_1$ , and  $\delta$  is the subjective discount rate of consumers.

There exists a unique rate of copying above which it is optimal for the consumer to consume all works at  $t = 1$ . At this cutoff value  $\hat{\alpha}(s_1)$  the marginal utility from consumption in the first period equals the discounted expected marginal utility in the second period, that is it solves the equation  $u'(s_1) = \delta(\beta - \alpha)u'(\alpha s_1)$  and satisfies  $\hat{\alpha}(s_1) \in (0, \beta)$ . For increasing relative risk aversion utility functions, this cutoff value  $\hat{\alpha}(s_1)$  is an increasing function of  $s_1$ , that is as the number of works available in the first period increases, the home technology must be good enough for the firm to sell all units in the first period. When  $\alpha < \hat{\alpha}(s_1)$  welfare maximization requires  $c_1^* < s_1$  solving  $u'(c_1) = \delta(\beta - \alpha)u'(\beta s_1 - (\beta - \alpha)c_1)$  while when  $\alpha > \hat{\alpha}(s_1)$  the solution is  $c_1^* = s_1$ . Hence, as  $\alpha$  increases, consumers are good substitute for firms for producing the works of art, and the opportunity cost of not consuming today increases.

On a competitive market, the first period asset value  $q_1$  of the works will be equal to the expected revenue from consumption in the two periods, and the artist will have revenue of  $q_1 s_1$ . Our previous discussion suggests that we may want to distinguish between monetary and non-monetary benefits of the artists, but for simplification, we do not. The business model of *Digital Art Auction* we described above fits a monetary interpretation since the artist gets revenues by having agents subscribe (pay the asset price) for obtaining copies from a master.

Asset prices are  $q_t$  and consumption prices are  $p_t$ . Feasibility conditions are  $c_t \leq s_t$ ,  $t = 1, 2$ . Asset prices satisfy  $q_1 s_1 = p_1 c_1 + p_2 c_2$  and  $q_2 s_2 = p_2 c_2$ . Because  $t = 2$  is the last period, there is no value of saving assets and  $c_2 = s_2$  and  $q_2 = p_2$ .

The competitive consumer's problem is then under prices  $p_t, q_t$ :

$$\begin{cases} \max_{c_1, c_2, s_2} u(c_1) + \delta u(c_2), \\ p_1 c_1 + q_2 s_2 \leq q_1 s_1, \\ p_2 c_2 \leq q_2 s_2 \end{cases} \quad (2)$$

for given prices  $p_t, q_t$ . For an interior solution we have

$$\frac{u'(c_1)}{\delta u'(c_2)} = \frac{p_1}{p_2}$$

and demand functions are defined implicitly by  $u'(c_1) = \mu p_1$ ,  $\delta u'(c_2) = \mu p_2$ , where  $\mu$  is the Lagrange multiplier of the first period budget constraint.

The firm's problem is

$$\begin{cases} \max_{y_1, y_2, s_2} q_1 s_1 = p_1 y_1 + p_2 y_2, & y_t \leq s_t, \\ s_2 = \beta(s_1 - y_1) + \alpha y_1, \end{cases}$$

where  $y_t$  is supply at  $t$ . Equilibrium conditions are that  $y_t = c_t$  for  $t = 1, 2$ . Profit maximization yields

$$c_1 \leq s_1 \quad \text{if} \quad \frac{p_1}{p_2} = \beta - \alpha.$$

The welfare maximizing allocation  $(c_1^*, c_2^*)$  is then decentralized by prices  $p_1, p_2$  satisfying<sup>35</sup>

$$\begin{aligned} p_1 &= u'(c_1^*), \\ p_2 &= \delta u'(c_2^*). \end{aligned}$$

If the rate of home production is low enough ( $\alpha \leq \hat{\alpha}(s_1)$ ) the market value of the initial assets is

$$\begin{aligned} q_1 s_1 &= p_1 c_1^* + p_2 c_2^* \\ &= u'(c_1^*) c_1^* + \frac{u'(c_1^*)}{\beta - \alpha} (\beta s_1 - (\beta - \alpha) c_1^*) \\ &= \frac{\beta}{\beta - \alpha} u'(c_1^*) s_1. \end{aligned} \quad (3)$$

While if the rate of home production is high enough ( $\alpha > \hat{\alpha}(s_1)$ ), all initial works are sold in the first period and the market value is

$$q_1 s_1 = (u'(s_1) + \delta \alpha u'(\alpha s_1)) s_1. \quad (4)$$

Therefore, the minimum revenue of the artist is bounded below by  $u'(s_1) s_1$ , independently of the rates  $\alpha$  and  $\beta$ . Moreover, this bound is independent of  $\delta$ ,  $\delta$  being also an index of appropriability of second period industry profits by the artist. This is the main result in [Boldrin and Levine \(2002a\)](#): as long as there is indivisibility in the provision of ideas, a competitive market will give a positive rent to the fixed factor (the artist). Hence, innovation is compatible with competition.

As  $\alpha$  increases, for given first period consumption, tomorrow's price will be lower since there will be more works available on the market; this is the curse or the competitive effect. As  $\alpha$  increases however, consumers value more first period consumption since increasing first period consumption does not go against smoothing intertemporal utility. The fact that home production improves means that there are more potential

<sup>35</sup> Indeed when  $\alpha \leq \hat{\alpha}(s_1)$ , in an interior solution, consumer's optimization yields  $u'(c_1)/\delta u'(c_2) = p_1/p_2 = u'(c_1^*)/\delta u'(c_2^*)$  which implies that  $c_t = c_t^*$  for  $t = 1, 2$ . At the price ratio  $p_1/p_2 = \beta - \alpha$ , the firm is indifferent between all combinations of  $y_1$  and  $y_2$  since its profit is  $p_2((\beta - \alpha)y_1 + y_2)$  while the resource constraint can be written  $(\beta - \alpha)y_1 + s_2 = \beta s_1$ ; since  $c_2 = s_2$  the result follows.

If  $\alpha > \hat{\alpha}(s_1)$ , the optimal allocation is  $c_1 = s_1$  and by definition  $\frac{u'(s_1)}{\delta u'(\alpha s_1)} > \beta - \alpha$ . Consumer maximization implies  $c_1 = s_1$ . Since  $u'(s_1) - \delta(\beta - \alpha)u'(\alpha s_1) > 0$ , firm's profit function  $u'(s_1)c_1 + \delta u'(\alpha s_1)(\beta s_1 - (\beta - \alpha)y_1)$  is increasing in  $y_1$  and therefore  $y_1 = s_1$  is optimum for the firm.

works of art available in the market; this is the blessing or the market expansion effect. The net effect of a small change in  $\alpha$  on the asset value is ambiguous and depends on the elasticity of demand. For a fixed value of  $\beta$ , it is clear from Equations (3) and (4) that the maximum value of the initial assets is attained at  $\alpha^*(s_1) > \hat{\alpha}(s_1)$ . This implies that all authors prefer a weak to a strong copyright regime; when  $\alpha$  is larger than  $\alpha^*(s_1)$ , however, authors may value the imposition of some restrictions on home production.

Consider the special case of constant demand elasticity, that is when the utility function is  $u(c) = c^{1-R}/(1-R)$ , where  $R \geq 0$ . Elasticity of demand is  $-1/R$  and demand is elastic when  $R \in (0, 1)$ . In this case the cutoff value  $\hat{\alpha}(s_1)$  is independent of  $s_1$ ;  $\hat{\alpha}(s_1) = \hat{\alpha}$  solving  $\hat{\alpha}^R = \delta(\beta - \hat{\alpha})$ . When  $\alpha \geq \hat{\alpha}$ ,  $c_1^* = s_1$ ,  $p_1 = s_1^{-R}$  and  $q_1 s_1 = s_1^{1-R}(1 + \delta\alpha^{1-R})$  which is increasing in  $\alpha$  since  $R \in (0, 1)$ ; therefore the market value of the assets increases as home production becomes as good as the firm's technology or as  $\beta$  increases. When  $\alpha \leq \hat{\alpha}$ ,

$$c_1 = \frac{\beta s_1}{\beta - \alpha + (\delta(\beta - \alpha))^{1/R}},$$

$$q_1 s_1 = \beta^{1-R} \left( \delta^{\frac{1}{R}} + (\beta - \alpha)^{\frac{R-1}{R}} \right)^R s_1,$$

and are increasing with  $\alpha$  (for a fixed  $\beta$ ) and increasing in  $\beta$  (for a fixed  $\alpha$ ). Therefore when demand has constant elasticity, the market value of the asset increases when home production improves, and in this case all authors strictly prefer the weak regime, independently of  $\alpha$ .

### 3.2. Appropriability: Market power

Consider now an extension of this basic model in which the artist, or the firm distributing his initial works, has market power. The market unfolds like in the competitive case. The monopoly firm decides to sell  $c_1$  works at time 1. Consumers become sellers at time 2 and behave competitively, that is take the price  $p_2$  as given; the quantity they will be able to offer on the market depends upon the copyright regime put in place.

The static profit function is  $\pi(c) = cu'(c)$  and the marginal profit function is  $\pi'(c) = u'(c)(1 - r(c))$ . Under Equation (1)  $\pi$  is single peaked. Let  $c^M$  be the unique maximizer of the profit function and write,

$$\pi^M \stackrel{\text{def}}{=} c^M u'(c^M).$$

I assume to simplify that the monopoly can commit to a production plan (or alternatively to prices).

In the *strong copyright regime*, the monopoly can prevent consumers to use their home technology. This could be because of a legal constraint or technical protection (like DRM); while there is some doubt on the ability to suppress *all* possibility of production by consumers, it seems hard to dispute the fact that this will lead to a lower value of  $\alpha$ . We consider here the extreme case where the strong copyright protection can be perfectly enforced and lead to  $\alpha = 0$ . Consumers solve Equation (2) and the

monopoly solves

$$\begin{cases} \max_{c_1, c_2, s_2} \pi(c_1) + \delta\pi(c_2), & c_2 \leq s_2, \\ s_2 = \beta(s_1 - c_1). \end{cases} \quad (5)$$

When the constraint  $c_2 \leq s_2$  binds – which is typically the case when  $s_1 < c^M$  – the solution to Equation (5) is  $\phi(s_1, \beta) = \max_{0 \leq c \leq s_1} \pi(c) + \delta\pi(\beta(s_1 - c))$ .

In the *weak copyright regime*, consumers can use their home production technology that enables them to create new works of art at a rate  $\alpha$ . Home production creates unwanted competition to the monopoly, and this effect is magnified when consumers behave competitively when selling their home production on the market. Because the production of consumers satisfies constant return to scale, it is convenient to think of a competitive firm having capacity  $\alpha c_1$  on the market in the second period, the monopoly firm having capacity  $\beta(s_1 - c_1)$ .

The representative consumer receives dividends from the monopoly firm and the competitive firm.<sup>36</sup> At time 1, the monopoly anticipates the behavior of the consumer and the competitive firm, and acts as a monopoly on the residual demand  $p_1 = u'(c_1)$ ,  $p_2 = \delta u'(c_2 + \alpha c_1)$ . The consumer takes as given the prices on the market for consumption and for the assets and has a consumption plan  $(c_1, c_2, s_2)$  solving Equation (2). Demand functions are given by  $p_1 = u'(c_1)$  and  $p_2 = \delta u'(c_2)$ . At the optimum both constraints bind and we have  $q_1 s_1 = p_1 c_1 + p_2 c_2$ . The competitive firm is present only at time 2, and its supply is  $\alpha c_1$ . Anticipating this second period supply, the monopoly solves then

$$\begin{cases} \max_{c_1, c_2, s_2} q_1 s_1 = \pi(c_1) + \delta\pi(c_2), & c_2 \leq s_2 + \alpha c_1, \quad c_2 \geq \alpha c_1, \\ s_2 = \beta(s_1 - c_1). \end{cases} \quad (6)$$

Note that the second period profit of the monopoly from the sale of consumption good is  $\delta(c_2 - \alpha c_1)u'(c_2)$  but that the total industry profit  $\delta c_2 u'(c_2)$  is incorporated in the initial asset value. The market internalizes the externality that period 1 consumption brings in terms of home production and the price of the assets  $q_1 s_1$  is strictly greater than the profit the monopoly firm makes from its sales.

There are therefore two effects of home production from the point of view of the monopoly. There is first the *market capacity expansion* effect illustrated by the constraint  $c_2 \leq s_2 + \alpha c_1$ , or  $c_2 \leq \beta(s_1 - c_1) + \alpha c_1$ : even if the monopoly sells all assets at time 1 ( $c_1 = s_1$ ), consumers can still consume in the second period since there is home production; this effect is positive since it is a source of extra profit. There is also the *competitive effect*, illustrated by the constraint  $c_2 \geq \alpha c_1$ , suggesting that the second period capacity may be too large from the point of view of profit maximization, i.e., the monopoly is now limited to choose second period prices greater than  $\delta u'(\alpha c_1)$ .

<sup>36</sup> There is a potential benefit for the monopoly to buy some of the works of the competitive firm at  $t = 2$ : doing so will increase second period profits since the monopoly would face a larger residual demand. However, the asset price  $\hat{q}_2$  should adjust. To simplify I ignore this possibility.

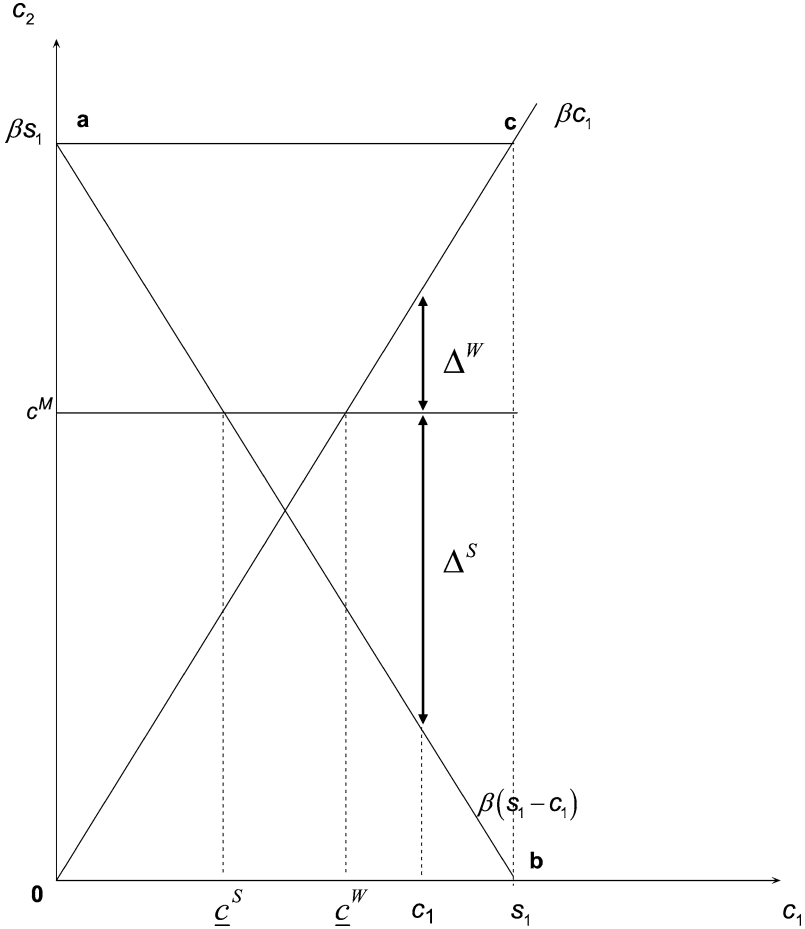


Figure 1. Case  $\alpha = \beta$ .

Figure 1 illustrates the market expansion and competition effects generated by the weak copyright regime in the case  $\alpha = \beta$ , that is when consumers have access to the same production technology as firms. With the strong copyright regime, the set of feasible second period industry sales ( $c_2$ ) is given by the triangle  $ab0$ , that is the area bounded by the production frontier  $\beta(s_1 - c_1)$  and the two axes. With the weak copyright regime the feasible set is given by the triangle  $ac0$ , that is the area bounded by the production frontier  $\beta s_1$ , the lower bound on second period consumption  $\beta c_1$  and the two axes.

Consider a value of  $c^M$  as in the figure, that is  $s_1 < 2c^M/\beta$ . In the strong copyright regime, first period consumption is greater than  $\underline{c}^S$ : if  $c_1 < \underline{c}^S$ , the monopoly will choose to set  $c_2 = c^M < \beta c_1$  which is not profit maximizing since by increasing



$c_1$  first period will increase without distorting second period profits. Similar reasoning shows that in the weak copyright regime first period consumption is greater than  $\underline{c}^W$ . It is then immediate that the strong copyright regime can dominate the weak copyright regime only if first period consumption is greater than  $\underline{c}^W$ . For any level of first period consumption  $c_1$  greater than  $\underline{c}^W$ , the strong regime puts an upper bound on second period consumption that is strictly lower than the monopoly quantity (underconsumption  $\Delta^W$ ) while the weak regime puts a lower bound on second period consumption that is strictly greater than the monopoly quantity (overconsumption  $\Delta^S$ ). As long as the profit function is symmetric around  $c^M$ , the resulting loss in profits is greater for  $\Delta^S$  than for  $\Delta^W$ . Hence, when  $s_1 < 2c^M/\beta$ , artists favor the weak copyright regime. Using a similar reasoning, when  $s_1 > 2c^M/\beta$ , artists favor the strong regime.

Since  $\beta(s_1)$  is U-shaped, both highly creative artists ( $s_1$  small) and poorly creative artists ( $s_1$  large) favor weak copyright laws; only “average” artists favor strong copyright laws. Highly and poorly creative artists are indirectly protected from the competitive effect: for highly creative artists interpretation is difficult while for poorly creative artists interpretation is easier but it is difficult to create new works of art.

This simple model therefore suggests that artists at the two extreme of the creativity scale benefit from market expansion: the most creative because market expansion facilitates interpretation of their work, the least creative because they have already “cornered” the market on their creative idea. This non-convexity resonates well with the variety of opinions of artists on the issue of copyright enforcement and on other protective measures such as resale rights.

#### 4. Issues and conclusion

Like photography and the invention of the printing press a few centuries ago, the new technologies of digitalization and the Internet threaten the market positions of artists and intermediaries. Artists because the technology of production of works may be readily accessible and craftsmanship may no longer be a defining characteristic of art. Intermediaries because their rents were linked to entry barriers in the distribution market. This curse of new technologies may be a blessing in disguise since it also increases the possibilities of production, of distribution and the emergence of new works of art.

Thinking of works of art as multi-dimensional goods with consumers valuing all dimensions but being able to observe only a subset of these permits a simple answer to Arrow’s (1962) problem: how could a creative idea yield revenues on a market if valuation requires disclosure of the idea and if the idea can be appropriated at no cost? As long as interpretation is needed for works of art, and consumers value this dimension, artists can obtain revenues from their creative ideas even if consumers have access to a production technology that makes some of the observable dimensions of the work easy to replicate or produce. Copyright may complement this effect, but the preferences of different participants in the market for strong copyright reflect a basic tradeoff between market expansion and competitive effects and have little to do with social efficiency,

including incentive provision for new creative ideas. Appropriability per-se is indeed not enough for market participants to favor weak copyright, that is to allow consumers to use to its full extent the new productive opportunities of the Internet. Intermediaries and artists may want to limit competition in order to increase the rents brought by the indivisibility of creative ideas.

Legal licenses like those proposed by Common Creative License facilitate the emergence of new business models that allow artists to bypass current gatekeepers while still providing appropriability, whether monetary or non-monetary. Whether or not the artist will obtain enough revenues to cover the costs of creation is really an empirical issue, but the current support of some artists for weak copyright is an indication that this is the case.

I warned the reader that this chapter will be incomplete. There are other effects of the Internet that deserve further study. Some are already analyzed in chapters of this handbook. Others are less studied and may prove important; one leading question is the relationship between ease of entry and the “quality” of the offerings on the art market.

In a world where information is complete, entry should lead to social gains; but this is not necessarily the case for the art market. For instance, there is a tension between the desire for a “global” presence and the desire to fit local tastes and culture. The media have coined the term “glocal” for expressing this tension. While there is some work on this topic<sup>37</sup> it is still unclear how facility of access to knowledge and need to interpret works from other cultures will affect the offerings on the art market.

If we abandon the fiction of a representative consumer, and if consumers have different abilities to interpret works of art or even to identify them, a need for certification arises: either to prevent fraud,<sup>38</sup> or to facilitate the interpretation of the work, e.g., by certifying the origin and therefore the historical context during which the work was created.<sup>39</sup> Who should provide this certification? Certification is often provided after a selection process, a screening process. Traditional gatekeepers (recording studios, galleries) play the role of screening and filter works of art that will be offered on the market; then as in the quote from Duchamp the artist “will have to wait for the verdict of the spectator” – critics, buyers, and historians build his reputation. The Internet by facilitating entry of artists shifts the role of screening to the market; it is not clear at this point whether this shift will improve on the previous system.<sup>40</sup> An indirect consequence of the difficulty to provide certification and screening on the Internet is that there are rents

<sup>37</sup> For instance, Legros and Stahl (2002) provide a theoretical argument showing how the number of varieties offered locally is affected by global competition and how the local market structure may alleviate this variety loss. There is a large “business economics” literature on these glocal strategies, see, for instance, Ghemawat (2001).

<sup>38</sup> For instance, during the inaugural auction of Nart.com, an online auction company, a Picasso drawing had to be withdrawn because of fears that it might be a forgery (*Economist*, January 27, 2000).

<sup>39</sup> Duchamp’s Fountain would have probably not made an impact on art if Duchamp had not been a certified artist.

<sup>40</sup> For instance, as is known from search theory, lowering of search costs may lead to less search by consumers because firms adapt their pricing strategy in equilibrium and may increase the equilibrium opportunity cost

to be gained by offering such services to consumers; this should affect the market for the distribution of works of art, possibly leading to more rather than less concentration.<sup>41</sup>

## References

- Arrow, K.J. (1962). "Economic welfare and the allocation of resources for invention". In: Nelson, R.R. (Ed.), *The Rate and Direction of Inventive Activity*. Princeton University Press, Princeton, NJ, pp. 609–625.
- Bakos, Y., Brynjolfsson, E., Lichtman, D. (1999). "Shared information goods". *Journal of Law and Economics* 42 (1), 117–155.
- Becker, G. (1976). *The Economic Approach to Human Behavior*. University Chicago Press, Chicago, pp. 284–294.
- Besen, S.M., Kirby, S.N. (1989). "Private copying, appropriability, and optimal copying royalties". *Journal of Law and Economics* 32 (2), 255–280.
- Bessen, J., Hunt, R.M. (2004). "An empirical look at software patents". Working Paper 03-17. Federal Reserve Bank of Philadelphia.
- Bessen, J., Maskin, E. (2000). "Sequential innovation, patents, and imitation". Working Paper 00-01. MIT.
- Boldrin, M., Levine, D.K. (2002a). "The case against intellectual property". *American Economic Review* 92 (2), 209–212.
- Boldrin, M., Levine, D.K. (2002b). "Perfectly competitive innovation". Mimeo. UCLA.
- Brynjolfsson, E. (2002). *Understanding the Digital Economy (Data, Tools, And Research)*. MIT Press, Cambridge.
- Conner, K., Rumelt, R. (1991). "Software piracy: An analysis of protection strategies". *Management Science* 37, 125–139.
- Corbett, A.R. (2003). "Software bullet is sought to kill musical piracy". *New York Times*, May 4.
- Danto, A.C. (1986). *The Philosophical Disenfranchisement of Art*. Columbia University Press, New York.
- David, P.A. (1992). "Knowledge, property, and the system dynamics of technological change". *Proceedings of the World Bank Annual Conference on Development Economics*, March, pp. 215–248.
- De Long, B., Froomkin, M. (1999). "Speculative microeconomics for tomorrow's economy". <http://econ161.berkeley.edu/OpEd/virtual/technet/spmicro.doc>.
- Dewatripont, M., Tirole, J. (2005). "Modes of communication". *Journal of Political Economy* 113 (6), 1217–1238.
- DRM Watch Staff (2004). "Apple's iTunes opens wider – by hook or by hack". <http://www.drmmwatch.com/drmtech/article.php/3400151>.
- Duchamp, M. (1959). *The Creative Act*. Paragraphic Books, New York, pp. 77–78. Edited by M. Duchamp and R. Lebel.
- Duchamp, M. (1966). "Apropos of 'Readymades' ". *Art and Artists* 1 (4).
- Ellison, G. (2003). "A model of add-on pricing". NBER Working Paper 9721.
- Gayer, A., Shy, O. (2001a). "Copyright protection and hardware taxation". Mimeo. University of Haifa.
- Gayer, A., Shy, O. (2001b). "Freeware, downloading, and piracy in markets for digital media". Mimeo. University of Haifa.
- Ghemawat, P. (2001). "Global vs. local products: A case study and a model". Mimeo. Harvard Business School.

of search. The documented search behavior of consumers on the Internet is an indication that this is not merely a theoretical argument, e.g., Brynjolfsson (2002). Search may also become less efficient, for instance, if producers adjust their offering and try to manipulate the search of consumers, see Ellison (2003).

<sup>41</sup> The current strategy of Microsoft to purchase the rights of paintings, photographs and other works may be an illustration of this.

- Givon, M., Mahajan, V., Muller, E. (1995). "Software piracy: Estimation of lost sales and the impact on software diffusion". *Journal of Marketing* 59, 29–37.
- Green, J.R., Scotchmer, S. (1995). "On the division of profit in sequential innovation". *RAND Journal of Economics* 26 (1), 20–33.
- Hahn, R.W., Wallsten, S. (2003). "A review of Bessen and Hunt's analysis of software patents". Mimeo. AEI-Brookings Joint Center. Available at [http://www.researchineurope.org/policy/hahn\\_wallsten.pdf](http://www.researchineurope.org/policy/hahn_wallsten.pdf).
- Hellwig, M., Irmen, A. (2001). "Endogenous technical change in a competitive economy". *Journal of Economic Theory* 101 (1), 1–39.
- Hui, K.L., Png, I. (2003). "Piracy and the legitimate demand for recorded music". *Contributions to Economic Analysis and Policy* 2 (1).
- Johnson, W.R. (1985). "The economics of copying". *Journal of Political Economy* 93 (1), 158–174.
- Kahin, B., Varian, H.R. (Eds.) (2000). *Internet Publishing and beyond: The Economics of Digital Information and Intellectual Property*. MIT Press, Cambridge and London.
- Klein, B., Lerner, A.V., Murphy, K.M. (2002). "The economics of copyright "fair use" in a networked world". *American Economic Review* 92 (2), 205–208.
- Lancaster, K. (1966). "A new approach to consumer theory". *Journal of Political Economy* 76 (2), 132–157.
- Landes, W.M., Posner, R.A. (2004). *The Economic Structure of Intellectual Property Law*. Harvard University Press, Cambridge.
- Leggat, R. (1999). "A history of photography – from its beginnings till the 1920s". <http://www.rleggat.com/photohistory/>.
- Legros, P. (2005). "Creativity and copyright". Mimeo. ECARES.
- Legros, P., Newman, A.F. (1999). *Interference: Contracts and Authority with Insecure Communication*. Brussels and London.
- Legros, P., Newman, A.F. (2002). "Courts, contracts, and interference". *European Economic Review* 46 (4–5), 734–744.
- Legros, P., Stahl, K. (2002). "Global vs local competition". CEPR Discussion Paper 3333.
- Lerner, J. (2003). "150 years of patent protection". Mimeo. Harvard University.
- Lerner, J., Tirole, J. (2002a). "The scope of open source licensing". Mimeo. Toulouse and Harvard University.
- Lerner, J., Tirole, J. (2002b). "Some simple economics of open source". *Journal of Industrial Economics* 50 (2), 197–234.
- Lessig, L. (2004). *Free Culture*. The Penguin Press, New York; pdf available under a Creative Commons license on Lessig's site.
- Liebowitz, S.J. (1985). "Copying and indirect appropriability: Photocopying of journals". *Journal of Political Economy* 93 (5), 945–957.
- Liebowitz, S.J. (2002). *Record Sales, MP3 Downloads, and the Annihilation Hypothesis*. University of Texas at Dallas.
- Nordhaus, W.D. (1969). *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change*. MIT Press, Cambridge.
- Novos, I.E., Waldman, M. (1984). "The effects of increased copyright protection: An analytic approach". *Journal of Political Economy* 92 (2), 236–246.
- Ordover, J.A., Willig, R.D. (1978). "On the optimal provision of journals qua sometimes shared goods". *American Economic Review* 68 (3), 324–338.
- Peitz, M., Waelbroeck, P. (2003a). "CD sales and internet piracy: Cross-section evidence". Mimeo. ECARES and Mannheim.
- Peitz, M., Waelbroeck, P. (2003b). "Piracy of digital products: A critical review of the economics literature". Mimeo. ECARES and Mannheim.
- Quah, D. (2002a). "24/7 competitive innovation". LSE.
- Quah, D. (2002b). "Matching demand and supply in a weightless economy: Market-driven creativity with and without IPRs". Mimeo. LSE.
- Quah, D. (2003). "Digital goods and the new economy". Mimeo. LSE.
- Romer, P. (2002). "When should we use intellectual property rights?" *American Economic Review* 92 (2), 213–216.

- Schopenhauer, A. (2004). "The art of controversy". eBook@Adelaide, translated by T. Bailey Saunders. First publication 1896. Available at <http://etext.library.adelaide.edu.au/s/schopenhauer/arthur/controversy/complete.html>.
- Scotchmer, S. (2004). *Innovation and Incentives*. MIT Press, Cambridge.
- Scotchmer, S., Green, J. (1990). "Novelty and disclosure in patent law". *RAND Journal of Economics* 21 (1), 131–146.
- Shavell, S., van Ypersele, T. (2001). "Rewards versus intellectual property rights". *Journal of Law and Economics* 44 (2), 525–547.
- Shy, O., Thisse, J.F. (1999). "A strategic approach to software protection". *Journal of Economics and Management Strategy* 8 (2), 163–190.
- Sutton, J. (1991). *Sunk Costs and Market Structure – Price Competition, Advertising, and the Evolution of Concentration*. MIT Press, Cambridge.
- Takeyama, L.N. (1994). "The welfare implications of unauthorized reproduction of intellectual property in the presence of demand network externalities". *Journal of Industrial Economics* 42 (2), 155–166.
- Takeyama, L.N. (1997). "The intertemporal consequences of unauthorized reproduction of intellectual property". *Journal of Law and Economics* 40 (2), 511–522.
- Varian, H.R. (2000). "Buying, sharing and renting information goods". *Journal of Industrial Economics* 48 (4), 473–488.

## CENSORSHIP VERSUS FREEDOM OF EXPRESSION IN THE ARTS\*

TUN-JEN CHIANG and RICHARD A. POSNER

*University of Chicago Law School, Chicago, IL, USA*

### Contents

Abstract	310
Keywords	310
1. Introduction	311
2. History	311
2.1. Censorship in ancient times	311
2.2. Censorship in Christian Europe	312
2.2.1. The Middle Ages – the iconoclast controversy	312
2.2.2. Reformation – heresy and nudity	314
2.2.3. The censorship of music	315
2.3. Censorship in the East	316
2.3.1. Islam – painting as blasphemy	316
2.3.2. China	316
2.3.3. Japan	317
2.4. Morality and art in the nineteenth century	317
2.4.1. Victorian England	317
2.4.2. France	318
2.4.3. Comstockery in the United States	319
2.5. Censorship in the twentieth century	320
2.5.1. Art in the Soviet Union	320
2.5.2. Public funding of art in the United States	320
2.6. Censorship of film	322
2.6.1. United States	322
2.6.2. Europe	324
2.6.3. Asia	326
3. Analysis: Offensiveness, artistic merit, and the law of freedom of expression	326

\* The Analysis section of this chapter draws in part on Posner's article [Posner (1989), "Art for law's sake", *American Scholar* 58 (4), p. 513]. The comments of Victor Ginsburgh on a previous draft are gratefully acknowledged.

3.1. The problem of objectivity	326
3.2. The law	330
References	333

## Abstract

Whether in ancient Rome or in the modern United States, censorship has existed in every society at every age. Art that challenges the strongly held beliefs of any society – whether those be political, ideological, religious, or otherwise – causes offense and creates pressure for censorship. At the same time, almost every society has found value in the existence of visual art. What limitations on censorship should be made for the sake of artist value, or more broadly freedom of expression? “Artistic merit” and “offensiveness” are nebulous concepts lacking in objectivity, shifting with the tastes of society at any given time. Yet the value of art to society, both positive and negative, cannot be doubted. In modern American society, with its heterogeneous tastes, the tension between the two concepts becomes especially vivid. Given the divergent and unpredictable tastes of society, the fact that destroying a work permanently removes it from future generations, and considering censorship’s dreadful history, the decision to censor is one appropriately made with caution. But neither can it be said that a work should never be censored, for art can and does cause offense, and even a society as diverse as ours will find consensus at the extremes. Rather, striking the appropriate balance calls ultimately for good judgment. In making this judgment, what is the appropriate role of the law and the courts? Those who think of the law as purely objective will desire the courts to either forbid all governmental interference with art, or to themselves abstain from interfering with political decisions on art. But these approaches place legal purity above reality, and make the impossible attempt to divorce law from its social context. The problem of relativism that inheres in the balance between artistic merit and offensiveness in fact exists in every legal controversy. The necessary public respect for our courts is unlikely to be undermined by a cautious display of good judgment, even if the judgment is inherently subjective and involves art causing offense to elements of our society.

## Keywords

censorship, visual art, first amendment, offensiveness, pragmatism

*JEL classification:* K10

## 1. Introduction

Our subject is the limitations that the concept of freedom of expression, including the legal embodiment of that concept in the First Amendment to the Constitution of the United States, places on the censorship of visual art (paintings, statuary, plays, opera librettos, movies, television programs, and so forth), when the censor's emphasis is on the visual rather than the verbal content of the art. We first sketch the history of such censorship and then try to explain the current position of (mainly) U.S. law with respect to it.

## 2. History

### 2.1. *Censorship in ancient times*

Ancient Egyptian civilization seems not to have had any “art” as understood in the modern sense. Workmanship, not creativity, characterized visual representation, and artists followed rigidly a set of prescribed conventions concerning color and proportion. While there are no records of individual artists being censored, the strict adherence to the conventions over a period of three millennia suggests that individual artistic freedom was highly limited to the point of being non-existent [Clapp (1972, p. 15)]. Not that subversive exercises of that freedom could be prevented entirely; satiric paintings, such as one depicting a mouse being fanned and fawned on by a cat, have been discovered [Jackson (2002)].

Plato praised Egyptian art precisely because “modification and innovation outside [the] traditional framework was prohibited” [Plato (1970, II, pp. 656–657)]. Plato was one of the earliest recorded advocates of rigorous censorship. His imagined utopian state imposed strict censorship in order to promote virtue and good morals in the young [Plato (1978, III, p. 401)]. Although he focused his criticisms more on heretical poetry and music than on the visual arts, he also disapproved of painting [Plato (1980, X, p. 603)] and sculpture [Plato (1993, pp. 235–236)] and argued that they should be submitted to state censorship so that their moral content could be monitored and if necessary corrected.

The draconian censorship envisioned by Plato was never fully realized in Greece. Sparta had some censorship relating to sculpture and music [Chambers (1928)], but the general attitude seems to have been liberal. Thus, although much Greek art was religious, the gods were portrayed in human form [Gage (1992)]. Greek vases even depicted the gods engaged in sexual acts – a depiction that would surely engender fierce controversy today if currently esteemed deities were the subject. Depictions of nudity of both men and women were acceptable. But the Greek tolerance for art was limited. Phidias, among the greatest of Greek sculptors, was tried in Athens in 438 B.C. and convicted of sacrilege for placing his own likeness on the shield of Athena. Historical sources disagree on whether he was exiled or imprisoned [Clapp (1972)].



Freedom of speech in ancient Rome depended on one's social rank, and in the republican era many poets were exiled for their writings. The Romans established the office of Censor as early as 443 B.C. [Riley (1998)], and though initially the only function of the office was to make a census of citizens and assess their wealth and taxes, the censors eventually acquired the power of *regimen morum*, or general control over the morals of citizens to determine their fitness for office. Romans occasionally associated art with decadence and corruption of morals. The Roman general Sulla was accused of corrupting his soldiers in foreign lands through art: "for there it was that the army of the Roman people first learned to indulge in women and drink, to admire statues, paintings and chased vases" [Chambers (1928, p. 55)].

We have found no instance in which an individual Roman artist was censored for his work; but then few names of Roman artists have survived. Even if there was no retribution against individual artists, their artwork received official scrutiny. Augustus displayed his displeasure upon seeing a statue of Marcus Brutus in Cisalpine Gaul, accusing the town of harboring a public enemy [Jones (2001, p. 2055)]. And a Roman administrator named Marcus Granius Marcellus was indicted for treason in 15 A.D. after elevating his own statue above those of the Caesars and mutilating the statue of Augustus [Jones (2001, p. 2055)].

Roman attitudes toward sexual representation may have shifted with time. Roman statues include many sculptures of naked male athletes; but in the Republican era naked statues of the gods were rare, and may even have been viewed as a sign of decadence. Naked depictions of the gods became more popular during the late Republic, though not without limits: Arellius was reportedly criticized for painting goddesses in the likeness of his mistresses [Jones (2001)].

Political factors were a frequent cause of censorship in ancient times (as in all subsequent times, for that matter). The destruction of a person's image after his or her downfall was a common act of retribution. When Thutmose III became Pharaoh of Egypt around 1481 B.C., he ordered the destruction of all statues and images of his aunt Hatshepsut erected during the period in which he had been forced to share power with her [Clapp (1972, p. 16)]. Similarly, the statues of Sejanus, chief minister to the Roman emperor Tiberius, were destroyed after Sejanus's downfall in 31 A.D. [Jones (2001)].

## 2.2. *Censorship in Christian Europe*

### 2.2.1. *The Middle Ages – the iconoclast controversy*

The historical relationship between artists and Christianity was complex. The Church (both its eastern and western divisions) was the most important patron of the arts for more than a thousand years, but also the greatest censor. Possessing both the carrot of patronage and the stick of excommunication, as well as other powers of persuasion and intimidation, the Church exercised enormous influence over artistic development.

The early Christians encountered, and were horrified by, the nudity and (as it seemed to them) the idolatry rife in the Hellenistic world. The Council of Elvira in 306 decided

that it should be permissible to exhibit paintings on only the outside walls of a church; and those who worshiped idols could not receive communion [Clapp (1972); Carmilly-Weinberger (1986)]. The Church became more accommodating towards artists after it discovered that art was a convenient way of spreading Christianity among the poor and the illiterate. The Edict of Milan in 367 allowed paintings to be displayed inside a church, but the order of the saints, the use of color and symbols, and other details of the paintings were all specified by church officials [Clapp (1972); Carmilly-Weinberger (1986)].

The most significant debate within Christianity about art during the middle ages was the iconoclast controversy. Nudity and sexual representation also posed problems, but the censorship of nudity was not important until the revival of classicism in the Renaissance revived artistic interest in the nude.

Idolatry is forbidden by the Second Commandment. Saint Clement of Alexandria, an early Christian thinker on art, wrote against art in general and graven images in particular [Carmilly-Weinberger (1986, p. 10)]. But the Church gradually incorporated some pagan traditions, and icons (paintings on wood) came to decorate many churches in both the East and the West [Jones (2001, vol. 1, p. 103, vol. 2, p. 1141)]. When the Bishop of Marseilles removed such images from churches, he received a gentle reprimand from Pope Gregory I, who hoped to enhance the influence of Christianity by means of art [Jones (1977)]. In furtherance of this effort Gregory I had Roman statues destroyed so that the images of saints that remained would be free from heathen competition.

The debate over the propriety of icons exploded into all-out war in the late seventh century. In 692, the Trullan Council prohibited representing Christ symbolically as a lamb and required that He be portrayed in human form [Alexander (1958)]. The Byzantine emperor Leo III went further but in the opposite direction in about 726, forbidding the making of icons and ordering the existing ones to be destroyed [Mango (1977, pp. 1–2)]. Leo's successor, Constantine V, intensified the crackdown on icons. Many churches were redecorated with landscapes depicting trees and animals [Jones (2001)].

Incidents of iconoclasm in the West were sporadic [Gimpel (1969, p. 14)]. Pope Gregory II opposed Leo III's decree, and his successor Gregory III declared that anyone who destroyed pictures of Christ or the Virgin Mary would be excommunicated [Alexander (1958)]. Charlemagne was also more liberal than his counterparts in the East, opposing the worship of images but endorsing their use as decorations in churches [Gimpel (1969)]. Many artists and sculptors fled Constantinople for Italy.

Iconoclasm declined with the accession of Leo IV to the Byzantine throne in 775, and his moderation was continued by Empress Irene after his death [Carmilly-Weinberger (1986, p. 12)]. Decrees of the Second Council of Nicaea ratified by both the western and eastern churches allowed icons to be venerated within broad limits. The Council also prescribed rules on how holy images were to be presented. The order of saints was determined; Jesus had to be on the right side of a painting, St. John on the left; the feet of Jesus, of the apostles, and of the angels could be painted bare but not the feet of the Virgin Mary and the saints. Manuals were distributed in the East informing artists of these rules [Gimpel (1969)].

Iconoclasm returned once more in 813, when Emperor Leo V renewed the prohibition on images. Soldiers destroyed them throughout the Byzantine Empire. This second iconoclast phase lasted until the death of Emperor Theophilus in 842. The Council of Constantinople brought the iconoclast movement to a final end the following year by reverting to the decrees of the Second Council of Nicaea [Clapp (1972)].

### 2.2.2. Reformation – heresy and nudity

The Reformation revived iconoclasm. Martin Luther strongly condemned the idolatry of the Catholic church, and his followers forbade religious paintings [Clapp (1972, p. 55)].<sup>1</sup> Though the reformers were enthusiastic in their destruction of idols, they had little problem with printed images depicting scenes from the Bible, or paintings denouncing the Catholic church. Indeed, they used satirical images to mock the Pope and denounce Catholic practices, and images of Luther were distributed in order to spread the Reformist message [Andersson (1997)]. In response, the Catholic church issued the Edict of Worms in 1521, decreeing that all printed, including pictorial, material contrary to Catholic belief be suppressed.

Censorship of Reformist teachings and propaganda was not the only example of an altered relation between the Catholic church and the art world. The rise of capitalism in Europe created new markets for art and thus eroded the ability of the Church to use its patronage to dictate artistic standards [see Clapp (1972, pp. 63, 65)]. One of the first paintings to cause scandal in this era was the *Sacra Conversazione* by Giovanni Battista Rosso, painted in 1518. The painting was a departure from the serenity of the classical style; and the Santa Maria Novella in Florence, for which the painting was made, refused to accept it [Bazin (1969)]. The sculptor Pietro Torrigiano was brought before the Inquisition in 1522 for sacrilege after smashing a statue of the Virgin Mary in protest against the small payment offered for the work. He starved himself to death before he could be executed [Clapp (1972)].

The depiction of nudity had been a problem for the Church from its earliest days, when Christians had denounced the “decadence” of Greek and Roman art. The Second Council of Nicaea required all adult persons depicted in pictures to be clothed, with limited exceptions for Jesus, the apostles, and angels [Carmilly-Weinberger (1986)]. During the Renaissance, admiration of classical civilization led to increased depiction of nudity in art, placing art on a collision course with Church doctrine. The Council of Trent in 1563 decreed that nudity in religious art was to be avoided in all cases, even where nude depiction was consistent with the Biblical record. Under the direction of Girolamo Savonarola in 1498, the “Bonfire of Vanities” in Florence consumed many paintings and sculptures [Carmilly-Weinberger (1986)]. Other incidents in this

<sup>1</sup> Luther’s own views were more moderate than some of his followers. When supporters were destroying paintings in Wittenberg, Luther stopped them with the statement, “Do not suppose that abuses are eliminated by destroying the object that is abused” [Carmilly-Weinberger (1986, p. 20)].

period included papal condemnation of *Sonetti Lussuriosi*, illustrated copies of which showed various positions for engaging in sexual intercourse [Clapp (1972)]. Controversy erupted even over Michelangelo's *Last Judgment*, the fresco in the Sistine Chapel above the altar, completed in 1541. Michelangelo's previous use of nudity in such works as *David* was accepted as representing heroism, but the nudity in the *Last Judgment* was described by a papal official as "shameless" and "better suited to a bathroom" [Clapp (1972, p. 60)]. In 1558 Pope Paul IV ordered that draperies be painted over portions of the *Last Judgment* [Clapp (1972)]. Even fellow artists joined in the censure; Pietro Aretino, the very author whose *Sonetti Lussuriosi* had been censored by the Church, wrote Michelangelo that the "licentiousness" of the *Last Judgment* made him blush [Clapp (1972, p. 61)]. Several years later El Greco recommended to the Pope that the work be destroyed [Bazin (1969, p. 145)]. Although the Pope did not accept the recommendation, he did order that more draperies be painted over the nude figures in the painting [Clapp (1972); Bazin (1969)]. Many other paintings and sculptures also suffered the indignity of having genitalia concealed by fig leaves [Jones (2001, vol. 3, p. 2059)].

Nudity to one side, the Inquisition kept a watchful eye over signs of heresy in art, though the actual prosecution of painters by the Inquisition was rare [Kaplan (1997)]. Andrea Gilio da Fabriano in his book *Dialogo Degli Errori Dei Pittori* criticized Michelangelo's *Last Judgment* for non-conformity with church regulations because, among other departures from orthodoxy, Jesus was depicted as unbearded and standing, instead of sitting on his throne; angels were represented without wings; and the Apocalypse angels were shown standing next to each other instead of at the four corners of the picture. In 1573 Paolo Veronese was summoned before the Inquisition in Venice for his painting *The Last Supper in the House of Simon* because he had added uncanonical details to a theologically sensitive subject. The inquisitor Aurelio Schellini was especially critical of the figure of a servant with a bloody nose, which he thought made light of the Eucharist [Bazin (1969); Kaplan (1997)]. The inclusion of German soldiers and a dwarf in the painting also drew the inquisitor's ire [Kaplan (1997)]. Yet the Inquisition was satisfied by Veronese's merely changing the title of the painting to *Feast in the House of Levi* [Jones (2001); Kaplan (1997)].

### 2.2.3. The censorship of music

Censorship of instrumental (non-vocal) music is rare, though not, as one might suppose, non-existent. Plato urged that music that appealed to the baser instincts should be censored. Centuries later, this view would be adopted by the Christians. St. Clement of Alexandria wrote against pagan music: "It must be banned, this artificial music which injures souls and draws them into feelings sniveling, impure and sensual . . . One must not expose oneself to the powerful influence of exciting and languorous modes, which, by the curve of their melodies, lead to effeminacy and infirmity of purpose" [Jones (2001, vol. 3, p. 1654)]. The association of pagan music with decadence was reflected

at the Second Council of Nicaea, which warned: “Woe on those who drink their wine to the sound of lyre and harp.”

After Christianity became the official religion of the Roman Empire, musical instruments were banned in worship. Saint Augustine of Hippo explained this ban as distancing Christianity from the “sensual heathen cults . . . and shameless performances of the degenerate theatre and circus” [Jones (2001, vol. 3, p. 1654)]. The ban persisted until 670, when the organ was permitted to be played at the Eucharist. But in the Eastern Orthodox church the ban persists to this day.

### 2.3. *Censorship in the East*

#### 2.3.1. *Islam – painting as blasphemy*

Islam was and still is far more hostile than Judaism and Christianity to artistic freedom, though there is variation within different branches of Islam. In addition to the prohibition of idolatry, Islam prohibits the representation of the human figure. The human form is regarded as the creation of Allah, and the depiction of it is therefore considered a divine prerogative [Arnold (1965)]. *Traditions of the Prophet* quotes Mohammed as saying “those who will be most severely punished by Allah on the Day of Judgment are the painters” [Clapp (1972, p. 33)].

Given such strict prohibitions, portrait and other painting in Islamic countries were and remain rare. Islamic artists focused their creative energies on calligraphy and ornamentation, where they faced few constraints [Papadopoulo (1979)]. Arabesque ornamentation, where human and animal figures are blended into the designs, was generally acceptable to religious authorities.

Art in defiance of religious teachings existed surreptitiously, aided by the lack of a central religious authority to dictate and enforce religious orthodoxy. Emperors and sultans had portraits and murals painted in the privacy of their palaces [Arnold (1965)]. Such art was sometimes destroyed by their successors. In about 869 the Caliph Al-Muhtadi ordered all paintings in his palace to be destroyed [Arnold (1965)]. Later, Sultan Firuz Shah issued an order to destroy not just paintings, but all ornaments, including those on the smallest objects, such as cups and curtains [Carmilly-Weinberger (1986)].

#### 2.3.2. *China*

The Chinese seem not to have been overly concerned with pictures; the written word was considered far more important in a society where literary skills were the key to high office. Art was influenced by Imperial tastes, because various emperors were great patrons of art, especially during the Han, Tang, and Sung dynasties [Carmilly-Weinberger (1986)]. Emperor Hui-Tsung of the Sung dynasty, himself a painter, assembled an art collection of more than 6000 pieces, with Buddhist themes especially prominent. He kept tight control over the painters at Court, dictating the subjects to be painted and

even giving artists examinations. The Emperor especially disliked the landscapes of Mi Fei, a master whose specialty was using blobs of ink in place of drawn lines. This radical technique was banned at the Court [Sullivan (1967)].

By the Ming dynasty, political concerns had found their way into visual art. The use of certain colors was proscribed to ordinary artists, being reserved for the Imperial Court. Painters were executed for insulting the emperor in their art, or dismissed from the Court for lesser offenses [Carmilly-Weinberger (1986)].

Comparatively few pieces of Chinese art have survived. Frequent wars took their toll, such as the Civil War of 190 A.D. that destroyed the collection of the Han emperors, and the Mongol conquest of 1279 that destroyed Hui-Tsung's collection [Carmilly-Weinberger (1986)]. Also, the materials used for Chinese art, such as bamboo, silk, and paper, failed to withstand the ravages of time.

### 2.3.3. Japan

The Tokugawa shogunate, which came to power in 1603, enacted strict censorship laws and exercised considerable control over the production of both woodblock prints and printed books. In Japan during this period the government's emphasis on education led to a dramatic rise in literacy, but woodblock prints remained a powerful instrument of mass communication that was believed to require rigid state control.

The Neo-Confucian ideology of the shogunate insisted upon a plain, strictly moral lifestyle. Though individual members of the ruling warrior class failed to satisfy these lofty aspirations, it was the self-indulgent consumption of the merchant class, whose economic power was steadily growing, that was seen as a threat to the established order. Tokugawa regulation of woodblock prints kept a watchful eye for political subversion and also censured sexual impropriety and excessive luxury.

The sensitivity of political subjects in art was reflected in the prohibition on portraying the Tokugawa family, the hereditary dictatorship that ruled in the name of the emperor, even in the most flattering terms [Thompson (1991, pp. 32–33)]. Sexual and sumptuary restrictions were more loosely enforced. Edicts in the 1720s banning erotic prints, a popular form of woodblock print, were ineffective; erotic prints continued to circulate, although without the signatures of the artist and printer [Thompson (1991, pp. 44–45)]. The banned picture book *Hyakunin Joro Shinasadame* showed streetwalkers, but it was not the streetwalkers *per se* that caused the ban; rather it was depictions of the Empress in the same volume as the prostitutes [Thompson (1991)].

## 2.4. Morality and art in the nineteenth century

### 2.4.1. Victorian England

The Society for the Suppression of Vice, formed in 1802, was expressive of the religious morality of the period. Partly as a result of the Society's efforts, the Obscene Publications Act was passed in 1857 by Parliament, authorizing magistrates to confiscate any articles considered indecent or obscene.

During the same period photography was invented and soon there was pornographic photography. The Obscene Publications Act was used to attempt to suppress this development; a police raid in 1874 reportedly seized over 100,000 obscene photos. Because “obscenity” was not defined, the magistrates were left to exercise their individual judgment. Challenged to produce the confiscated “abominations” for inspection, the Chief Justice of England, the leading proponent of the Act, excused his refusal on the ground that they would shock the modesty of England’s peers [Paul and Schwartz (1961)]. Even those photos with acknowledged artistic merit could cause controversy. One photograph by the well known artist O.G. Rejlander had to be covered by a black drape when it was exhibited in Edinburgh because it depicted gamblers and prostitutes [Jones (2001)].

More traditional media were also subject to scrutiny. In *Du Bost v. Beresford* (1810), the court found the painting *Beauty and the Beast* libelous. This was followed in 1868 by the case of *Regina v. Hickland* (1868), which defined obscenity by reference to its likely effects on the imagination of the intended audience of the work.

The English theater, censored since its earliest days, remained under censorship until well into the twentieth century. Under the Licensing Act of 1937, the Lord Chamberlain was authorized to deny licenses to plays “as often as he shall think fit”, a standard that did not invite consistency.

#### 2.4.2. France

Censorship in art was commonplace in France from before the Revolution. Louis XVI had suppressed various works, such as Beaumarchais’ *Marriage of Figaro* and erotic paintings by Francois Boucher. During the nineteenth century, theater and printed images continued to be heavily regulated; prior authorization was required. The Cour Royale de Paris and the Tribunal d’Assises de la Seine destroyed more than 232 prints, engravings, and drawings deemed immoral or obscene [Carmilly-Weinberger (1986)]. But because the audience for paintings as distinct from prints was overwhelmingly the middle and upper classes, self-regulation of painting was permitted: the artists themselves were the censors. Their instrument was the artists’ jury of the Paris Salon. Though the first jurors were appointed by the government in 1800, after 1830 they were appointed by the Academie des Beaux Arts. Rejection of the work by the Salon jury was not final, however, because after 1863 the work could still be displayed in the Salon des Refusés. It is a testament to the artistic judgment or authority of the Salon juries that the stigma of rejection by the jury would often lead artists to withdraw their works from the market rather than display them in the Salon des Refusés.

One artist who had consistent difficulty with the jury was Auguste Rodin; he struggled with criticism and rejection of his “impressionist” sculpture. Another and greater artist repeatedly rejected by the Salon jury was Edouard Manet. Political considerations were no doubt predominant in the suppression of his *The Execution of the Emperor Maximilian* in 1869, Maximilian having been installed as Emperor of Mexico at the behest of Napoleon III. But other paintings of his, such as *Nana*, were rejected for their supposed sexual content. Manet’s greatest painting, *Déjeuner sur l’herbe*, was not only



rejected by the Salon jury but also caused a furor among viewers and critics when it was displayed in the Salon des Refusés in 1863, and fellow artist Odilon Redon recorded his private disapproval [Clapp (1972)]. Remarkably, Manet's *Olympia* was accepted by the Salon jury, but it had to be guarded by two policemen, and subsequently removed to a remote location, because of popular fury at its style and nudity. The objection, as to *Déjeuner sur l'herbe*, was not to nudity as such, but to the fact that Manet's nudes looked like prostitutes, yet were inserted in classical settings; for example, *Olympia* was a pastiche of Titian's *Venus of Urbino* – but Manet's "Venus" was depicted as a whore [Needham (1972)].

Rodin and Manet were of course not the only victims of Salon rejection. Gustave Courbet's *The Return from the Meeting*, a painting depicting drunken priests, caused such outrage that it was rejected on the order of the government despite Courbet's exemption from jury scrutiny as a past winner [Carmilly-Weinberger (1986)]. A devout Catholic later bought the painting just so that he could destroy it [House (1997)]. Similarly, Courbet's *Venus Pursuing Psyche in Her Jealousy* was rejected by the Paris Salon the next year because of its lesbian theme.

#### 2.4.3. Comstockery in the United States

Puritan morality had a strong impact on life in the United States from its earliest days. One of the first cases to involve obscene art was *Commonwealth v. Sharpless* (1815). The Supreme Court of Pennsylvania found the display of an obscene painting offensive to the morals and dignity of the community, and levied a fine.

Federal legislative efforts to stamp out immorality in art began with the Tariff Act of 1842, in which Congress authorized customs officials to seize and destroy imported "indecent, obscene prints, paintings, engravings, lithographs and transparencies".<sup>2</sup> The first case under this statute was brought the next year and resulted in the seizure of three paintings along with a shipment of snuff boxes [United States v. Three Cases of Toys (1843)].

The prohibition was strengthened in 1865, when mailing obscene books and pictures was made a criminal offense. The passage in 1873 of the "Comstock Act" [An Act for the Suppression of Trade (1873)] increased the penalties and extended the list of prohibited materials. Under this Act, Anthony Comstock became a special agent of the Post Office. A zealot for puritan morality and the suppression of vice, Comstock claimed to have destroyed nearly four million obscene pictures during his career [Cooper, Stark and Zaleski (1994)]. He founded the New York Society for the Suppression of Vice in 1873, which the New York Legislature empowered to search, seize, and arrest for the purpose of eradicating vice.

In one of the many cases in which he was involved, *People v. Muller* (1884), Comstock secured the conviction of August Muller for selling obscene photos; the court

<sup>2</sup> Prior to this, only Vermont had enacted an anti-obscenity statute [Brockwell (1994)].



ruled that a painting is obscene if it “is naturally calculated to excite in a spectator impure imaginations”. He was less successful in an 1888 case against the Knoedler gallery. Four days after Comstock raided the gallery to seize 117 pictures and engravings replicating works of French art, the *New York Telegram* displayed the seized pictures on its front page as a brazen challenge to Comstock. In the eventual court case against 37 of the seized pictures, the judge found that only two were objectionable [Clapp (1972)].

## 2.5. *Censorship in the twentieth century*

### 2.5.1. *Art in the Soviet Union*

At the time of the Communist Revolution in 1917, the exciting development in the Russian art world was futurism, a movement with political undertones and industrial themes. Russian futurists believed that industrialization would be the bedrock of a new social order and vowed to “fight with all our might the fanatical, senseless, and snobbish religion of the past” [Carmilly-Weinberger (1986, p. 121)].

Before the revolution, Communists had supported the futurist and cubist movements on the theory that these movements were a challenge to the bourgeois establishment. But after the Communist Party took power, their attitude changed. Art was reconceived as a method of bringing Communism to the masses. Lenin dismissed modernist art with a simple statement, “I have the courage to declare myself a ‘barbarian’. I am unable to consider the works of Expressionism, Futurism, Cubism and other ‘isms’ as the highest manifestation of artistic genius. I don’t understand them. I experience no joy from them” [Carmilly-Weinberger (1986)]. In order that art should serve Communism, works that did not further that aim had to be destroyed and new ones created. Many monuments honoring Tsar Alexander and others were destroyed during this period.

After Stalin took control of the Communist Party in 1924, the Soviet government tightened its grip on the art world. In 1934 Andrei A. Zhdanov, Secretary of the Central Committee of the All-Union Communist Party, declared that the Central Committee would henceforth promote the “exclusive practice of socialism realism . . . the depiction of reality in its revolutionary development”, i.e. Communist propaganda [Carmilly-Weinberger (1986, pp. 133–134)]. Artists who failed to comply with the dictates of socialist realism were persecuted and their works destroyed.

### 2.5.2. *Public funding of art in the United States*

The Works Progress Administration (WPA) inaugurated significant federal funding for the arts in the United States. The WPA’s Federal Art Project supported many artists during the Great Depression, funding more than 2500 murals, 18,000 sculptures, and 100,000 paintings. But government funding came with strings attached, and works promoting Communist ideology were censored under the program. Even without official censorship, individuals would censor their own art to fit the politics of the time. In

1933, Nelson Rockefeller had Diego Rivera's mural in Rockefeller Center destroyed because Rivera had refused to remove an image of Lenin from it.

As anticommunist emotion gripped the nation in the 1950s, the United States Information Agency canceled two government-sponsored art shows because some of the artists involved were pro-communist [Cooper, Stark and Zaleski (1994, p. 32)].<sup>3</sup> The State Department also canceled a traveling art exhibition organized under a cultural exchange program after conservative organizations denounced the exhibits as Communist propaganda. A leading force behind the State Department's action, Congressman George A. Dondero, urged artists' organizations to expel Communists and reward patriotic painters.

In recent decades, political controversy surrounding the arts have been rare.<sup>4</sup> Notable were the controversies that embroiled the National Endowment for the Arts (NEA) in 1989 over works of two photographers whom the NEA had funded, Robert Mapplethorpe and Andres Serrano.

Robert Mapplethorpe's photographic exhibition *The Perfect Moment* was accused of being obscene after the Corcoran Gallery of Art in Washington canceled the planned exhibition of his pictures shortly after the artist's death. When the collection was eventually moved to the Contemporary Arts Center (CAC) in Cincinnati, the Director of the Center was indicted on charges of promoting obscenity and using a minor in nude materials. The Mapplethorpe photos contained graphic depictions of homosexual and sadomasochistic acts and images of nude children. Among the photographs was one that showed a black man urinating into the mouth of a white man. Nonetheless, the defense of the CAC lined up an impressive array of artists and curators who testified to the artistic merits of Mapplethorpe's photography, and the defendants were acquitted.

Andres Serrano's *Piss Christ*, a photograph of a plastic crucifix immersed in a bottle of the artist's urine, caused even more controversy than the Mapplethorpe exhibition, and was eventually destroyed by a vandal in Melbourne, Australia in 1997. While no legal retribution befell the artist, the political fallout was considerable. Under the leadership of Senator Jesse Helms, Congress passed an Act to bar use of NEA funds to "promote, disseminate or produce materials . . . which may be considered obscene". When a court struck down this law as unconstitutional, Congress passed another law that required the NEA to consider "general standards of decency". Controversy briefly surrounded the NEA once more when artist Karen Finley challenged the constitutionality of the amended decency clause, but subsided after the Supreme Court upheld the statute [National Endowment for the Arts v. Finley (1998)].

<sup>3</sup> The two exhibitions were "Sports in Art" and "100 American Artists of the Twentieth Century".

<sup>4</sup> Though certainly not non-existent. Two well-known incidents in Chicago during the 1980s demonstrate the sensitivity of some political subjects. David Nelson's *Mirth and Girth*, an unflattering painting of former Chicago mayor Harold Washington, was carried off by a hostile alderman [Nelson v. Farrey (1989)]. A display by Scott Tyler at the School of the Art Institute of Chicago inviting visitors to walk on the American flag caused the Illinois legislature to cut the school's funding to \$1.

A recent funding controversy illustrates the difference between attitudes to art in the United States and those of continental Europe. The exhibit *Sensation*, a tour of the collection of Charles Saatchi, traveled from London's Royal Academy, to Berlin, then to New York in 1999. The collection has no shortage of controversial pieces, from a sculpture made from frozen blood to a shark preserved in formaldehyde [Needle (1999)]. But what caused a storm in London was a painting by Marcus Harvey, made up of prints from a plaster cast of a child's hand, depicting the child murderer Myra Hindley. The British tabloids roundly criticized the work, members of the Royal Academy resigned in protest, and it was damaged twice by vandals throwing ink and eggs. After restoration it had to be protected behind a layer of glass.

In Germany, the exhibit attracted some isolated protests [Art and offense (1999)], but there were no violent outbreaks or significant controversy [Fitzpatrick (2002)].<sup>5</sup> Its popularity proved such that the museum extended the exhibition for one month [Steiner (1999)]. When the exhibit came to New York, the Harvey painting attracted little notice. However, the Catholic League condemned a work by Chris Ofili, titled *The Holy Virgin Mary*, that depicted a black Madonna, surrounded by images of buttocks, with the addition of elephant dung. Mayor Rudolph Giuliani threatened to cut funding to the Brooklyn Museum over the matter, but was prevented from doing so by a court decision [Brooklyn Institute of Arts & Sciences v. City of New York (1999)]. A devout 72-year-old Catholic later breached security and smeared white paint over the Ofili work [Richards and Calvert (2000)].

The New York controversy had lasting ripple effects. The National Gallery of Australia canceled its planned exhibition of *Sensation* [National Gallery of Australia (1999)], and plans for a possible tour in Tokyo never materialized.

## 2.6. *Censorship of film*

### 2.6.1. *United States*

The motion picture was invented in the late nineteenth century but not developed until the twentieth. "Nickelodeons", stores where short films could be seen for a nickel, spread across the United States. Like photography before it, pornography quickly followed the development of the new medium and censorship followed closely behind. To forestall official censorship, the industry in 1909 formed the New York Board of Censors to evaluate films before public release. In 1915 this became the National Board of Censorship of Motion Pictures.

Though political censorship in the United States has historically been rare compared to censorship for reasons of morality, the development of film in the first half of the

<sup>5</sup> Not that Germans are unconditionally accepting of all art, a Berlin exhibit showing photographs of "chopped-up tin soldiers in SS uniforms who are being castrated by skinhead mutations" caused some visitors to walk out and others to vomit [Vallely (2000)].

twentieth century happened to coincide with a period of political turmoil. D.W. Griffiths' *Birth of a Nation*, an early film that featured many cinematic innovations, was highly controversial because of its sympathetic portrayal of the Ku Klux Klan. The failure of the National Board of Censorship to ban the film undermined the authority of that body. Some local censors, such as the Kansas State Board of Review, refused to approve the film, and the mayor of Minneapolis threatened to revoke the license of any theater that showed it. The Supreme Court of Minnesota upheld the ban [*Bainbridge v. Minneapolis* (1915)].

War also led to heightened political sensitivity. *The Spirit of '76* was censored in 1917, at the height of World War I, for its negative portrayal of the British during the Revolutionary War. The judge in the case wrote:

History is history, and fact is fact . . . At the present time, however, the United States is confronted with what I conceive to be the greatest emergency we have ever been confronted with at any time in our history. There is now required of us the greatest amount of devotion to a common cause . . . , and as a necessary consequence no man should be permitted . . . to do that which will in any way detract from the efforts which the United States is putting forth. [*United States v. Motion Picture Film "The Spirit of '76'"* (1917)]

After the end of the war, the nation's focus returned once more to morality in film. The Production Code was adopted in 1930. Initially it was administered by the Studio Relations Committee of the Motion Picture Producers and Distributors of America (MPPDA), which allowed filmmakers a fair degree of flexibility. But in 1933 the Catholic Legion of Decency led a nationwide boycott, and the film industry capitulated and appointed a Catholic layperson, Joseph Breen, to head the Production Code Administration (PCA) and enforce the Code strictly. Breen maintained an iron grip over the industry for the next ten years [*Jones* (2001)].

The PCA's authority was undermined in 1946 when after a dispute with the PCA over the amount of cleavage that could be shown *The Outlaw* was released without a PCA seal of approval but nevertheless enjoyed tremendous success at the box office. This story was repeated with *The Moon Is Blue* in 1953, when Breen refused to approve the film's sexual innuendo but audiences apparently didn't mind. The influence of the Catholic Church also waned. Since 1933, a condemnation from the Catholic League of Decency had spelled certain disaster [*Phelps* (1975)] – until 1953, when *From Here to Eternity* received a condemned rating from the Catholic Legion of Decency yet was approved by the PCA, won eight Academy Awards, and was a box-office success.

Competition from television put pressure on the movie industry to show what audiences could not see at home [*Asimow* (2000)], and the MPPDA lost some control over theaters after losing the antitrust case *United States v. Paramount Pictures* (1948). After successive changes to the Code, each more liberal than its predecessor, the PCA was abandoned in 1968 in favor of the new Film Code and Ratings Administration with a suggested rating system, shifting the burden of enforcement from the industry to theaters and individuals.

In addition to the self-censorship of the movie industry, the law of obscenity remains as a mechanism of official censorship.<sup>6</sup> In the early twentieth century movies were a constitutionally unprotected medium [*Mutual Film Corp. v. Industrial Commission of Ohio* (1915)]. Many cities and states maintained local censorship boards [see *Vieira* (1999)]. The Supreme Court reversed course in 1952 by ruling that movies were protected by the First Amendment [*Joseph Burstyn Inc. v. Wilson* (1952)]. Obscene movies, like other categories of obscene speech, remain unprotected. The debate has since been about what constitutes the obscene.

In 1957, the Supreme Court first announced a test for obscene speech; it was whether the dominant theme of the material appeals to the prurient interest of the average person applying community standards [*Roth v. United States* (1957)]. The Court replaced the Roth test in 1966 with a three-part test that involved asking whether the material

- (1) appealed to a prurient interest in sex,
- (2) was patently offensive because it affronted contemporary community standards, and
- (3) was utterly without redeeming social value [*A Book Named 'John Cleland's Memoirs of a Woman of Pleasure' v. Massachusetts* (1966)].

The issue whether a particular work was obscene returned again and again to the Court, which eventually began avoiding the issue. Between 1967 and 1973 the court reversed thirty-one obscenity convictions without opinion [*Curtis* (1986)]. During this period the lower courts became ever more permissive, as few works could be deemed to be “utterly” without redeeming social value.

The Supreme Court revisited the issue in 1973. In *Miller v. California* (1973) it articulated a subtly modified three-part test. The issues are now:

- (1) whether the average person, applying contemporary community standards, would find that the work appeals to the prurient interest,
- (2) whether the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by state law, and
- (3) whether the work, taken as a whole, lacks serious literary, artistic, political or scientific value [*Miller v. California* (1973)].

This test remains the framework for testing obscenity under the First Amendment. As a practical matter, however, only child pornography is suppressed in the United States today, other than on radio and television (other than pay-TV).

### 2.6.2. *Europe*

In Britain, the British Board of Film Censors (BBFC) was formed in 1912 by the Kinetograph Manufacturers Association to preempt official censorship. Local authorities have given the BBFC's classifications legal force by conditioning cinema licenses upon

<sup>6</sup> Prohibitions against sedition and blasphemy are generally acknowledged as unconstitutional, given the Supreme Court decisions in *New York Times Co v. Sullivan* (1964) and *Joseph Burstyn Inc. v. Wilson* (1952).

a theater's agreeing to show only BBFC-certified films. Local authorities retain the power to accept or reject any BBFC classification or establish other conditions [Phelps (1975)].

Layered over this quasi-official regulation of films is the Obscene Publications Act, which in 1977 was amended to clarify that it covered the distribution of films. A jury now decides whether material is obscene under the Act as tending to deprave and corrupt. Thus a film may be banned from Britain by either failing to qualify for a certificate from the BBFC or by offending the Obscene Publications Act.

The German constitution contains a free-speech clause declaring that there shall be no censorship, but there is an exception for the protection of youth. And despite the constitutional right of free expression, the penal code retains the offenses of blasphemy and dissemination of obscenity [The German Penal Code (2002, 106, pp. 118–120)]. In practice, however, the only limits on pornography in Germany are for the protection of children [Jones (2001)]. An industry organization, the Spitzenorganisation der Filmwirtschaft (SPIO), operates like the MPAA in the United States as industry self-censor. Any film not certified by the SPIO is likely to be boycotted. Attitudes, however, vary greatly from the U.S., for the SPIO is tolerant of depictions of sex but prohibits anything that might offend diplomatic or religious sensibilities [Phelps (1975)].

The open display of pornography is technically illegal in the Netherlands, but such restrictions are widely ignored by both police and consumers, resulting in the ready availability of hard-core pornography [see “Controlling pornography” (1998)]. Belgium has never had an official censorship body [Phelps (1975, p. 242)], but unlike other northern European countries has a prudish culture that underwrites a great deal of de facto censorship. Danish film censorship for adults was abolished in 1969 [Media Council for Children and Young People (n.d.)]. Sweden has the oldest film censorship body in existence, the National Board of Film Censors (Statens Biografbyrå), established in 1911, but though the Board has the power to ban films, the power is rarely exercised [see Furhammar (2001, p. 5)]. In contrast to the UK and the U.S., the focus of the Board is on psychological damage to children from violence [see Furhammar (2001)]. Curiously, the film industry in Sweden has been among the defenders of the Board's continued role, citing the uncertainty that would arise from the abolition of censorship [see Furhammar (2001)].

Despite the generally liberal attitudes toward pornography on the continent, there have been movements to curtail pornography in both France and Poland, the latter of course a highly Catholic, conservative nation. In France, where television stations air a large number of hardcore films, the legislature recently raised the tax on pornographic films to 93 percent [Henley (2002)], and there are calls for banning pornography on television completely. In Poland, a sweeping antipornography bill was passed by the Parliament but vetoed by the president [Polish Porn (2000)].

### 2.6.3. Asia

The Japanese Criminal Code prohibits the sale of obscene materials. Primary enforcement has devolved onto the industry's self-governing body, Eirin, as well as customs officials in the case of foreign films [Alexander (2003)]. The censorship system is rigid but seemingly rather futile. The entertainment media in Japan are well known for sexual and violent content, including particularly graphic images of sado-masochism. However, until recently, all displays of the genital area of either sex in film were prohibited, and even today Eirin allows only limited display of genitalia [Pitman (1995)]. For example, Nagisa Oshima's film *In the Realm of the Senses* was explicitly pornographic, and though it was shown at the Cannes Film Festival the director was tried in Japan for obscenity although eventually acquitted. The rereleased version of the film restores scenes cut by the censor but still digitally scrambles images of genitalia [Alexander (2003)].

China has a strict censorship system, the ideological foundation of being the subordination of art to politics, as in Plato's Republic and the Soviet Union. A complex web of government agencies is involved in the regulation of film, including the Ministry of Culture, the Ministry of Propaganda, and the Ministry of Radio, Film, and Television. On its face, Article 35 of the Chinese Constitution confers an extremely broad right of free speech, but in practice this is extremely limited [Calkins (1999)]. Until 1996 individual censors had broad discretion to censor as they wished, resulting in such erratic decisions that film-makers agitated for a law that would clarify standards. The resulting law appears to have achieved little clarification, and films are banned for such vague crimes as depicting "low-class" themes or "bad ideas" [Korski (1996)].

Politics remains the paramount consideration in Chinese film censorship. *Temptress Moon*, a film with no shortage of sexuality and drug use, was banned primarily because the censors associated one character with President Jiang Zemin [Tung (1998)]. Popular Hollywood films such as *Independence Day* are banned because of their "American spirit" [Brent (1999)]. The strict censorship regime also, however, results in the suppression of depictions of nudity and vice in films.

## 3. Analysis: Offensiveness, artistic merit, and the law of freedom of expression

### 3.1. The problem of objectivity

As our historical account, sketchy as it is because of the vastness of the subject, should have made clear, the common element in the censorship of art is perceived offensiveness: when art challenges strongly held beliefs, usually of a political, ideological, moral, or religious character, there is pressure for censorship. Broader generalizations regarding the political economy of censorship of the arts are elusive. Reliable correlates of such censorship are difficult to identify. There is no robust correlation between democracy and censorship, or even between censorship and religion, since not all religions



are as prudish as Judaism, Christianity, and Islam, and there are considerable variations regarding the concept of the obscene in those religions. Moreover, sexual explicitness is only one of the grounds of censorship. About all that can be ventured with confidence concerning the causality of censorship is that the more offensive to the public at large or influential segments of the public an artistic depiction is, the more likely it is to be suppressed either by government or by public opinion inducing self-censorship. To dig deeper into the causality of censorship would require considering the social sources of concepts of offensiveness. Probably those sources can be traced to differences in the status of women, the role of the family, the political structure, and the regnant religion, but we have not tried to conduct such an analysis.

We emphasize “offensiveness” rather than “harm” because censorship of art is rarely based on a plausible causal linkage viewing a work of art, whether highbrow or popular, and engaging in antisocial behavior. The analytical and regulatory problem is that in a democratic, culturally and morally heterogeneous, society such as that of the contemporary United States, there is little agreement on what is offensive, and efforts to suppress offensive work are therefore strongly resisted by those who derive pleasure from it that is not overborne by a sense of its offensiveness and who consider it their right to pursue leisure activities that do not cause demonstrable harm to other people. We need to consider whether, if this is right, it implies that offensive art should get a lot or a little – or even no – protection from governmental interference, however that interference should be defined in this setting.

That the concept of offensiveness is not objective in a strong sense of the word has been illustrated throughout our historical narrative of censorship in Section 2 of this chapter, which demonstrated that concepts of the offensive are culturally specific rather than universal. That doesn’t mean that they are arbitrary, even though they vary not only across countries but also within the same country in different periods. They can probably, as we have said, be traced to social factors. At any given time in any given place, there might be such general agreement on what is offensive that there would be no issue of lack of objectivity. That lack becomes perspicuous only when, a culturally heterogeneous society such as that of the modern United States, one attempts to classify particular works as being “offensive” or when, conceding offensiveness, one claims that it is redeemed by artistic distinction – a move that anyway is convincing only to those who regard offensiveness and artistic distinction as compatible; “moralistic” critics of the arts do not.

Let us try to explore this idea of “objectivity” a little further, beginning with the question of “objective” artistic merit and moving from there to offensiveness. When one says that lead is heavier than aluminum or that an automobile is faster than a rickshaw, one is making a statement that can be verified by methods independent of the tastes or personal values of the people doing the verifying. A Communist, a nudist, a Jehovah’s Witness, and a follower of Ayn Rand will all agree on how to test such propositions and on how to interpret the test results. Such “observer independence” gives the propositions about lead and automobiles truth value. It is quite otherwise if we say that *Piss Christ*



has, or does not have, artistic value in the context of modern American (or, more broadly, Western) values.

The problem is not that artistic value is not a thing which a work either has or has not, for in this respect artistic value is no different from weight or speed, being like them an attribute or property rather than a thing. You don't take apart a Maserati and announce, "This is the carburetor and that is the speed". Similarly, *Piss Christ* is not a composite of urine, a bottle, a crucifix, a photograph, and artistic value. But while it is possible to make objective measurements of physical properties such as weight and speed, it is not possible to make such measurements of artistic value, because people having different values and preferences do not agree and cannot be brought to agree on how to determine the presence of that attribute or even how to define it. A moralistic critic such as Tolstoy might think that the most important question about *Piss Christ* from an artistic standpoint is its likely effect on belief in Christianity. A Marxist critic might agree, and might further agree with Tolstoy that *Piss Christ* would undermine that belief, yet they would disagree about whether this made the work good or bad. Even if everyone to whom judges are willing to listen agrees that a work has no artistic value, we know from historical experience that later generations may find such value in the work even though the artist's contemporaries did not. Conversely, a work highly valued in its time, or for that matter in later times, may eventually come to seem thoroughly meretricious.

Artistic value is something an audience invests a work with, and as the tastes of audiences change, so do judgments of artistic value. About all that can be said in a positive vein is that the longer a work is held in high repute the likelier it is to continue to be held in high repute. This is the "test of time" that Samuel Johnson, David Hume, and George Orwell thought the only objective test of artistic merit.<sup>7</sup> If, to take a concrete example, the Homeric epics are still being read more than twenty-five hundred years after they were composed, then chances are they will continue to be highly regarded for some time; their appeal is robust and resists cultural change.

So far, though, all we have established is an inductive generalization, not an explanation. We could try to figure out what such durable works as the *Iliad* and *Hamlet* and Raphael's Madonnas and *The Marriage of Figaro* and the "Ode on a Grecian Urn" and the Louvre's "Winged Victory of Samothrace" have in common and call that the key to artistic value. But this sort of thing has been attempted for millennia without success, and it now seems clear that the quest is a snipe hunt, so diverse are the durable works of the Western tradition. Conceivably we might identify a *necessary* condition of artistic survival – that a work have a certain "omnificance" or, less portentously, a certain ambiguity or generality that enables it to be taken in different ways in different times and places. But the distinction between a necessary and a sufficient condition is critical here, for we would not concede artistic value to every work that crossed some

<sup>7</sup> With special reference to the visual arts, see Ginsburgh and Weyers (2003). On the failure of other "objective" tests of artistic merit, see Victor Ginsburgh (2003).

threshold of ambiguity or generality. *Piss Christ* deals with a fundamental concern of humankind, religion, and does so in a distinctly ambiguous way, Serrano denied harboring any blasphemous intent and indeed claimed that the work is a Christian commentary on the debasement of religion in modern America. The work may have artistic or even moral value, and then again it may not; it may soon come to be thought a worthless bit of trash, though, since it no longer exists, the issue may have become quite academic. Yet if *Piss Christ*, existent or not, seems altogether too slight and ephemeral a work to have any chance of winning a secure niche in art history, let us remind ourselves that the common urinal that Marcel Duchamp exhibited, one of the *objets trouvés* of the Dadaist movement, has won such a niche along with Aretino's *Sonetti Lussuriosi* and Utamaro's erotic prints.

The conclusion to which we are driven is that ascriptions of artistic value or valuelessness to "works of art" – especially to contemporary works of art – are arbitrary. And likewise with offensiveness, another property of, rather than a thing found in, a work. *Piss Christ* is no more a compound of urine, a bottle, a crucifix, and offensiveness than it is a compound of urine, a bottle, a crucifix, a photograph, and artistic value. Again this property, offensiveness, is largely, perhaps entirely, a matter of public opinion rather than of correspondence to or causation by something that is observer independent, something akin to the forces that determine weight and speed in accordance with the laws of physics.

As we pointed out earlier, this is not a problem when public opinion is united, as perhaps it is over the offensiveness of certain particularly graphic or degrading types of visual obscenity. With specific reference to *Piss Christ* one might be tempted to argue that, while there may be no consensus on what is art, there is a consensus, in Western societies anyway, that the public display of excreta is offensive. Consensus is a highly fallible warrant of truth, yet we might grant it provisionally objective status, even when it is local and temporary – a consensus in our society today, although not in all others and perhaps not in ours tomorrow. But it is a mistake to suppose that there is a consensus concerning the offensiveness of public displays of excreta. If samples of diabetics' or addicts' urine, or the feces of sufferers from Crohn's disease or cancer of the colon were displayed at a medical convention, we would not think the display offensive, it is all a matter of context and purpose. The question of the offensiveness of *Piss Christ* is therefore connected to the question of its artistic value. Those who find the work artistically valuable will not be offended by the (photographic) presence of urine, which they will consider integral to the work's value, just as a display of excreta might be integral to medical training. Those who find the work blasphemous and barren of artistic value will consider the display of urine gratuitous and hence, given our culture's feelings about excreta, offensive. A few people may find the work both offensive and aesthetic, as many probably found Michelangelo's *Last Judgment*. Their judgment on whether *Piss Christ* should be suppressed will depend on how offensive, and on how aesthetic, they find it, and on their personal sense of the proper balance between art and insult.

### 3.2. The law

All our discussion of the difficulty of balancing the offensiveness of a work of (purported) art against its artistic merit would have little or nothing to do with law if law had its own values, if it were morally autonomous. But for the most part it does not and is not. The law that entitles the victim of negligence to collect damages from the injurer is parasitic on – has no life apart from – social norms concerning what is careful and what is careless behavior. The prohibition in the First Amendment against government's abridging freedom of speech and freedom of the press, these freedoms being broadly conceived to include artistic as well as political and scientific expression, is parasitic in the same sense on social norms concerning artistic as well as other "speech" values and offensiveness and other speech harms such as violence. A speaker who urges a mob to lynch a prisoner because his guilt is so plain that a trial would be a waste of time and money will be punishable for incitement to violence because the danger of the speech will be felt to outweigh its value in drawing attention to the problematic character of due process. But if instead he writes a book urging the masses to rise up and liquidate the bosses, he will not be punishable, because such books are thought to have some value and not to be very dangerous, although citizens of Communist states may want to dispute both points. If our society thought such books were dangerous, they would be suppressed. Practical considerations, rather than the text or the eighteenth-century background of the First Amendment, guide the application of the amendment to today's problems.

Certain forms of obscenity are considered by virtually everyone in our society (including many of the consumers of obscene works) to be completely worthless and highly offensive, and they are suppressed without much ado, although, it must be added, also without much success. But the consensus that condemns the extremely obscene does not extend to the class of works illustrated by *Piss Christ*, which are thought valuable and non-offensive by some, worthless and offensive by others, and worthwhile but offensive by a handful. If there is no objective way to arbitrate such a disagreement, what should the courts do? More broadly, what are the implications for law of the kind of cultural relativism that we are describing?

There are three possibilities, of which the first two reflect a desire to secure the certainty of the law at any price. The first is to forbid *any* governmental interference with "art", no matter how offensive the "art" is. This approach does not escape subjectivity entirely; rather, it pushes inquiry back a stage, to the question of whether the work in question *is* art (and also to what counts as governmental "interference"). If a work is sufficiently offensive, it is classified as obscene, and therefore as non-art. At the other extreme is the judicial-hands-off approach: Courts are the forum of principle, there are no principles to apply to questions of aesthetic merit and offensiveness, so let the political branches do what they want with these questions. Such an approach is likely to appeal to those who are especially protective of courts – who want the judges to shine and believe that the judicial escutcheon is tarnished when the judges mess with indeterminate questions such as artistic value and offensiveness.

The third possible approach, the intermediate or pragmatic, is to acknowledge that the problem of relativism, moral as well as aesthetic, is a general feature of American, and perhaps of any, legal controversy. Judges need not feel that they must shy off from judgment merely because the issues raised by offensive art are spongy. That's just the way things are in law; the nature of the legal enterprise ensures that judges will frequently find themselves wrestling with indeterminate questions because those are the questions least likely to be settled without recourse to lawsuits that have to be pressed all the way to the Supreme Court or to another high appellate court before the question can be answered. Judges struggle with such questions all the time yet somehow manage to retain that minimum of public respect which is indispensable to the effectiveness of a court system. They are unlikely to forfeit it if they venture – with appropriate caution – into the controversy that eddies around issues of value and taste in purported works of art.

The first thing to note about this venture is that although artistic value is largely, perhaps entirely, unknowable, there is little doubt that art is valuable. If this seems a paradox, consider: the lesson of history is that many of the scientific theories in which we firmly believe today are almost certainly false, just as Euclidean geometry as a theory of spatial relations, the geocentric theory of the solar system, the luminiferous ether, the spontaneous generation of bacteria, and Newton's laws of motion are now known to be false after having been believed by the scientific community for centuries to be true. Yet the fallibility of scientific theory does not lead a sensible person to doubt the existence, growth, or value of scientific knowledge. Even if every current scientific theory is someday falsified, we will still be able to make atomic bombs, fly airplanes, and immunize people against polio. Likewise it is a fact that art museums are thronged, that works of art command huge prices, that some people devote a lifetime to the study of art, and – more to the point – that many people would feel a profound sense of deprivation had the nineteenth-century French artistic establishment succeeded in suppressing Impressionist art, just as they feel that the world is a poorer place because so little classical Greek sculpture has survived.

If we grant that art has value and add that the censorship of art has a dreadful historical record, we can derive, in order to guide judicial review of controversies over offensive art, a presumption in favor of letting the stuff be produced and exhibited to whoever is willing to pay the price of admission. The Supreme Court's decisions affirming the constitutional right to burn the American flag as a form of political expression is illustrative here. Burning a nation's flag is an offensive, inarticulate, and immature mode of political communication, but as long as one is burning a flag that one bought and paid for, and burning it before a willing audience, the burning contributes, however feebly, to the marketplace of ideas without impairing anybody's property rights – though many people will be offended.

We can bolster the presumption in favor of a permissive judicial attitude toward offensive art by noting that the “test of time”, which is the closest we seem able to get to an “objective” measure of artistic merit, presupposes, like natural selection in the theory of biological evolution (which the test of time resembles), the existence of variety,

from which history makes its selections. The thrust of censorship is to reduce variety, to suppress outliers, and by doing so to disrupt the test of time and impoverish art's legacy to the future.

Consider now a case in which the presumption in favor of freedom of artistic expression was successfully rebutted. *Piarowski v. Illinois Community College* (1985) involved a small junior college near Chicago that, being public, was subject to constitutional limitations on restricting free expression. The artist in the case, who was the chairman of the college's art department, made an improbable effort to fuse his two loves – the making of stained-glass windows and the art of Aubrey Beardsley – by creating stained-glass versions of some of Beardsley's illustrations from Aristophanes' comedy *Lysistrata*. The illustrations, like the comedy, are bawdy even by today's standards (how fitting that they should be on public display in the Victoria and Albert Museum in London). They are, of course, line drawings of white men and women – Greeks. But to transpose the drawings to the stained-glass medium the artist used pieces of colored glass for each of the figures, and the colors had to contrast, and he made the innocent but, as it turned out, unfortunate choice of amber glass for the women and white glass for the men. As a result, one of the stained-glass windows depicts a brown woman, naked except for stockings, on her knees, embracing in an attitude of veneration the huge white phallus of a robed man. The other two stained-glass windows depict brown women passing wind and masturbating, respectively.

The artist displayed the stained-glass windows in the art department's annual exhibition of faculty work, held in an alcove (the "gallery") off the main corridor of the college (the "mall", as it is called), on the ground floor. As the college had only one building, the exhibition was visible to all students, faculty, and visitors, whether or not they wanted to see it – there was no wall between the gallery and the mall. The first group to complain was the cleaning staff, which was black. Most of the students in the college are black, and they, too, were offended by the stained-glass windows and complained to the president of the college, who ordered the artist to shift the display from the first floor corridor to a smaller exhibition room on the fourth floor, a room normally used for exhibiting photographs but suitable for exhibiting other works of art as well. When the artist refused, the president took down the stained-glass windows and placed them in his office. The artist sued the college, charging a violation of the First Amendment.

The stained-glass pastiches seem largely free of both artistic value and offensiveness. On the one hand, Beardsley's charm is in the line, and it is lost when lines give way to chunks of colored glass. On the other hand, there was no contention that the artist was attempting a commentary on race or sex; he was merely trying to use different colors, vaguely human, to distinguish the figures in the windows from one another. The very crudeness of the windows, moreover, neutralizes any obscene impact. A "hands-off relativist" might take the position that since issues of artistic and moral taste are not objective, the artist should have lost his suit even if the college had refused to allow him to exhibit his stained-glass windows anywhere on (or for that matter off) the campus. This was not the court's view, but neither did the court think the Constitution *entitled*

Piarowski to exhibit his stained-glass windows in the most public place in the college. The college's president had offered an alternative place of exhibition that while indeed less conspicuous was by the same token less offensive. Consider how many great works of art, such as Manet's *Olympia* and *Déjeuner sur l'herbe*, survive today probably only because they were denied the most conspicuous places of exhibition. Racial sensitivities are a fact in our society, and if offensiveness ultimately is no more objective than artistic value, neither is it less so.

The college president's action seemed a reasonable compromise, and the court gave judgment for the college. In so doing, the court affirmed that "academic freedom" is a two-way street. It is the freedom of a college to manage its affairs without undue judicial interference no less than it is the freedom of the teacher or scholar to teach or write or, in this case, create works of art without undue interference by the state (for remember that this was a public college, and hence an arm of the state). A further point is that the power of a single junior college to affect the art scene by shunting offensive works to less conspicuous places of exhibition is distinctly limited. But the example might prove catching.

So particularistic and fact-specific – so pragmatic – a mode of adjudication that led to the judgment for the college, and one that implies that the scope of First Amendment protection may be different for works of art than for political or scientific works, is not to everyone's taste. Lawyers have a predilection for rules, and there are many situations in which hard-edged rules are indeed preferable to fuzzy standards. But controversies over offensive art may not be one of them. It is not even clear that art would be helped rather than hindered by a rule that forbade any and all public regulation of offensive art. Such a rule – a rule that gave privileged status to the flaunting of offensive art – might engender public hostility to art that would be out of all proportion to the benefits in artistic freedom gained. At the same time, a rule that gave government carte blanche to suppress art deemed offensive by any vocal, assertive, politically influential group in a diverse, teeming, and rather philistine modern society, such as that of the United States, could impair the future of art, a costly consequence. Perhaps, therefore, the watchword in the legal regulation of offensive art should be caution.

## References

- A Book Named 'John Cleland's Memoirs of a Woman of Pleasure' v. Massachusetts (1966). United States Reports 383, 413–462.
- An Act for the Suppression of Trade in, and Circulation of, Obscene Literature and Articles of Immoral Use (1873). United States Statutes at Large 17, 598–600.
- Alexander, J.R. (2003). "Obscenity, pornography, and the law in Japan: Reconsidering Oshima's 'In the Realm of the Senses' ". *Asian-Pacific Law & Policy Journal* 4, 148–168.
- Alexander, P.J. (1958). *The Patriarch Nicephorus of Constantinople; Ecclesiastical Policy and Image Worship in the Byzantine Empire*. Clarendon Press, Oxford.
- Andersson, C. (1997). "The censorship of images in Reformation Germany, 1520–1560". In: Childs, E.C. (Ed.), *Suspended License: Censorship and the Visual Arts*. University of Washington Press, Seattle, pp. 32–58.

- Arnold, T.W. (1965). *Painting in Islam: A Study of the Place of Pictorial Art in Muslim Culture*. Dover, New York.
- Art and offense in Brooklyn (1999). *Washington Post*, September 29, A28.
- Asimow, M. (2000). "Divorce in the movies: From the Hays Code to *Kramer vs. Kramer*". *Legal Studies Forum* 24, 221–267.
- Bainbridge v. Minneapolis (1915). *Northwestern Reporter* 154, 964–966.
- Bazin, G. (1969). *The Avant-Garde in Painting*. Simon & Schuster, New York.
- Brent, W. (1999). "China censors cut deep swatch in film". *Variety*, September 27, 18.
- Brockwell, P.H. (1994). "Grappling with *Miller v. California*: The search for an alternative approach to regulating obscenity". *Cumberland Law Review* 24, 131–145.
- Brooklyn Institute of Arts & Sciences v. City of New York (1999). *Federal Supplement Second Series* 64, 184–206.
- Calkins, M.L. (1999). "Censorship in Chinese cinema". *Hastings Communications and Entertainment Law Journal* 21 (2), 239–338.
- Carmilly-Weinberger, M. (1986). *Fear of Art: Censorship and Freedom of Expression in Art*. Bowker, New York.
- Chambers, F.P. (1928). *Cycles of Taste*. Harvard, Cambridge.
- Clapp, J. (1972). *Art Censorship: A Chronology of Proscribed and Prescribed Art*. Scarecrow Press, Metuchen, NJ.
- Commonwealth v. Sharpless (1815). *Sergeant and Rawle* 2, 91–106.
- Controlling pornography: How Britain compares with other countries in dealing with the problem of obscenity (1998). *Guardian*, August 13, 4.
- Cooper, S., Stark, C., Zaleski, D. (1994). *A Survey of Art Censorship: Painting Sculpture and Photography: An Annotated Bibliography*. Queens College CUNY, New York.
- Curtis, M.K. (1986). "Obscenity: The justices' (not so) new robes". *Campbell Law Review* 8 (3), 387–419.
- Du Bost v. Beresford (1810). *English Reports* 170, 1235.
- Fitzpatrick, J. (2002). *Art Censorship: A Panel Discussion*. Online at [http://www.thersa.org/acrobat/james\\_fitpatrick170402.pdf](http://www.thersa.org/acrobat/james_fitpatrick170402.pdf).
- Furhammar, L. (2001). *Swedish Culture: Swedish Film*. Swedish Institute, Stockholm. Online at [http://www.sweden.se/upload/sweden\\_se/english/factsheets/si/si\\_fs112b\\_swedish\\_film/fs112b.pdf](http://www.sweden.se/upload/sweden_se/english/factsheets/si/si_fs112b_swedish_film/fs112b.pdf).
- Gage, N. (1992). "Introduction". In: Buitron-Oliver, D. (Ed.), *The Greek Miracle: Classical Sculpture from the Dawn of Democracy*. National Gallery of Art, Washington.
- Gimpel, J. (1969). *The Cult of Art: Against Art and Artists*. Stein & Day, New York.
- Ginsburgh, V. (2003). "Awards, success and aesthetic quality in the arts". *Journal of Economic Perspectives* 17 (2), 99–111.
- Ginsburgh, V., Weyers, S. (2003). *Persistence and Fashion in Art: Italian Renaissance from Vasari to Berenson and beyond* (unpublished). ECARES, Université Libre de Bruxelles.
- Henley, J. (2002). "French to clamp down on porn with 'full-frontal' tax". *Guardian*, October 12, 20.
- House, J. (1997). "Manet's Maximilian: Censorship and the salon". In: Childs, E.C. (Ed.), *Suspended License: Censorship and the Visual Arts*. University of Washington Press, Seattle, pp. 185–209.
- Jackson, K. (2002). "Visual arts: Mummy's boys". *Independent*, April 23, 19.
- Jones, D. (Ed.) (2001). *Censorship: A World Encyclopedia*, vols. 1–4. Fitzroy Dearborn, London.
- Jones, W.R. (1977). "Art and Christian piety: Iconoclasm in Medieval Europe". In: Gutmann, J. (Ed.), *The Image and the Word: Confrontations in Judaism, Christianity and Islam*. Scholars Press for the American Academy of Religion, Missoula, pp. 75–105.
- Joseph Burstyn Inc. v. Wilson (1952). *United States Reports* 343, 495–540.
- Kaplan, P.H.D. (1997). "Veronese and the Inquisition: The geopolitical context". In: Childs, E.C. (Ed.), *Suspended License: Censorship and the Visual Arts*. University of Washington Press, Seattle, pp. 85–124.
- Korski, T. (1996). "Warning focuses on violence in films". *South China Morning Post*, September 24, 8.
- Mango, C. (1977). "Historical introduction". In: Bryer, A., Herrin, J. (Eds.), *Iconoclasm*. Center for Byzantine Studies, Birmingham, pp. 1–6.

- Media Council for Children and Young People (n.d.). History of Danish Film Censorship. Online at [http://www.medieraadet.dk/html/gb/history\\_gb.htm](http://www.medieraadet.dk/html/gb/history_gb.htm).
- Miller v. California (1973). United States Reports 413, 15–48.
- Mutual Film Corp. v. Industrial Commission of Ohio (1915). United States Reports 236, 230–247.
- National Endowment for the Arts v. Finley (1998). United States Reports 524, 569–623.
- National Gallery of Australia (1999). Sensation. Online at <http://www.nationalgallery.com.au/press/prarchive2.htm>
- Needham, G. (1972). “Manet, ‘Olympia’ and pornographic photography”. *Art News Annual* 38, 80–89.
- Needle, J.B. (1999). “Brooklyn Institute of Arts and Sciences v. City of New York: Mayor of New York violates First Amendment right to experience sensation”. *DePaul-LCA Journal of Art and Entertainment Law* 10, 77–112.
- Nelson v. Farrey (1989). *Federal Reporter Second Series* 874, 1222–1234.
- New York Times Co. v. Sullivan (1964). United States Reports 376, 254–305.
- Papadopoulo, A. (1979). *Islam and Muslim Art*. H.N. Abrams, New York.
- Paul, J.C.N., Schwartz, M.L. (1961). *Federal Censorship: Obscenity in the Mail*. Free Press of Glencoe, New York.
- People v. Muller (1884). *New York Reports* 96, 408–414.
- Phelps, G. (1975). *Film Censorship*. Victor Gollancz, London.
- Piarowski v. Illinois Community College (1985). *Federal Reporter Second Series* 759, 625–633.
- Pitman, J. (1995). “What the censor did not see”. *Times*, June 26, 17.
- Plato (1970). *The Laws*. Penguin, London.
- Plato (1978). *Republic*, vol. 1. Harvard, Cambridge.
- Plato (1980). *Republic*, vol. 2. Harvard, Cambridge.
- Plato (1993). *Sophist*. Hackett, Indianapolis.
- Polish porn bill overturned (2000, March 27). BBC News Online. Online at <http://news.bbc.co.uk/2/hi/europe/692586.stm>.
- Regina v. Hickland (1868). *Queens Bench* 3, 360–379.
- Richards, R.D., Calvert, C. (2000). “Counterspeech 2000: A new look at the old remedy for ‘bad’ speech”. *Brigham Young University Law Review* 2000 (2), 553–586.
- Riley, G.B. (1998). *Censorship. Facts on File*, New York.
- Roth v. United States (1957). United States Reports 354, 476–514.
- Steiner, R. (1999). “Charles Saatchi to open chain of restaurants”. *Sunday Times*, July 18, 3.
- Sullivan, M. (1967). *A Short History of Chinese Art*. University of California Press, Berkeley.
- The German Penal Code* (2002). William S. Hein & Co., Buffalo.
- Thompson, S.E. (1991). “The politics of Japanese prints”. In: Thompson, S.E., Harootunian, H.D. (Eds.), *Undercurrents in the Floating World: Censorship and Japanese Prints*. Asia Society Galleries, New York, pp. 29–91.
- Tung, L. (1998). “Waiting for the ice to melt”. *Asiaweek*, September 4, 40–42.
- United States v. Motion Picture Film ‘The Spirit of ’76.’ (1917). *Federal Reporter* 252, 946–949.
- United States v. Paramount Pictures, Inc. (1948). United States Reports 334, 131–181.
- United States v. Three Cases of Toys (1843). *Federal Cases* 28 F Cases, 112–113.
- Vallely, P. (2000). “Profile: Jake and Dinos Chapman”. *Independent*, September 16, 5.
- Vieira, M.A. (1999). *Sin in Soft Focus: Pre-Code Hollywood*. Harry N. Abrams, New York.



## THE ARTS IN THE “NEW ECONOMY”\*

WILLIAM J. BAUMOL

*New York University and Princeton University, USA*

### Contents

Abstract	340
Keywords	340
1. Introduction	341
2. Technical innovation and its effects on the arts	342
2.1. Artistic creation and “modern art”	342
2.2. Dissemination	343
2.2.1. Availability of the arts to consumers	343
2.2.2. Artistic products in international trade	345
2.3. Preservation	346
2.4. Funding problems and the cost disease	346
3. Pricing issues for the arts in the new economy	348
3.1. Repeatedly sunk costs and pricing problems	348
3.2. Pricing under copyright and compensation of creative activity	351
3.3. Efficient component-pricing	353
4. Other problems contributed by the mass media	355
4.1. Superstars	355
4.2. Contraction of secondary performance venues	356
5. Concluding comment	357
References	357

\* With deep thanks to my colleague, Sue Anne Batey Blackman, for her contributions to this chapter. When the editors of this volume invited me to prepare this piece, they suggested a grand evaluation of what the field of arts economics had accomplished, what work was currently under way and what most urgently still needed to be done. Though my longevity in the field appears to make this an appropriate assignment for me, the work has already been carried out with exemplary competence, notably in a well-known piece by David Throsby (1994), in a more recent article by Mark Blaug (2001), and, by implication, in the superb compendia edited by Ruth Towse (notably 1997), as well as the writings of Alan Peacock (e.g., 1992), and Bruno Frey and Werner Pommerehne (e.g., 1989). Evidently, the ground is well covered and there is no urgent need for yet another entry.

*Handbook of the Economics of Art and Culture, Volume 1*

*Edited by Victor A. Ginsburgh and David Throsby*

*Copyright © 2006 Elsevier B.V. All rights reserved*

*DOI: 10.1016/S1574-0676(06)01011-8*

**Abstract**

The revolutionary growth in economic prosperity and technological change that underlie the “new economy” have profoundly affected the arts. They have evidently contributed new and previously unimaginable methods of dissemination and preservation. But they have even had revolutionary effects on goals and standards. They have substantially affected training. These developments also raise profound problems for financial support and pricing. The paper characterizes these developments and suggests the nature of the relationships. It also provides a brief discussion of relevant pricing principles dealing with the trade-off between encouragement of creativity and facilitation of dissemination.

**Keywords**

arts financing, art dissemination, art preservation, new media, artistic standards and technological progress, pricing of reproduction rights

*JEL classification:* Z11

## 1. Introduction

The “new economy” – whose hallmark is extraordinary technological change – is a very real phenomenon, one whose manifestations and consequences are so mind-boggling that we have had to readjust ourselves to it by a complete revision of expectations. As a consequence, we have simply grown to take it for granted. But in my view it is not something that began only in the second half of the twentieth century. Rather, the unprecedented world of the new economy emerged perhaps some two centuries earlier and has since been accelerating at a dizzying pace. The available historical statistics, such as they are, suggest that by the eighteenth century real per-capita incomes barely exceeded those of Imperial Rome, which translates on average into more than 1500 years of near-zero growth. Then in the 18th century per-capita incomes in England rose some twenty or thirty percent. Nineteenth century incomes increased 100 to 200 percent in much of Western Europe and the US. As for the 20th century, it has been argued by knowledgeable observers that the oft-suggested 600 or 700 percent rise is a gross *underestimate*.

What has this extraordinary growth in economic wealth to do with the arts? The answer is that it has brought with it opportunities, capacities and *problems* that our ancestors could never have imagined. One writer on the accomplishments of the “new economy” (in my broad, longer-period connotation) has cited as a characteristic example the fact that before the 18th century no one had ever dreamed that a person could travel faster than on horseback, much less communicate across oceans instantaneously.<sup>1</sup> For the arts, it is easy to provide examples of comparable wonders: Today we can readily experience (somewhat imperfectly) a performance by Caruso, and I have heard the voice of Johannes Brahms.

Thus, economic circumstances powerfully influence both innovation and the arts. The state of the economy cannot produce creativity, but it can stimulate the exercise of creativity and facilitate dissemination and utilization of its products. On the other side, economic conditions can starve and otherwise handicap creative activity and condemn its products either to very limited use or even to oblivion. Clearly in the case of innovation – interpreted in Joseph Schumpeter’s sense to encompass the entire innovative process from the birth of new ideas to the full utilization of their products – the market economy has produced an outpouring of novel products and processes not remotely paralleled in any other form of economic organization.<sup>2</sup> And the new economy has undoubtedly affected the arts by stimulating technological developments that have even served to create new art forms. But it has also revolutionized the means of dissemination

<sup>1</sup> This is not quite true. There were dreamers who went beyond what was known. As early as the 13th century Roger Bacon, in a remarkable and much quoted passage, had foreseen a future that provided ships and wagons driven by machines that would move “. . . with incredible speed and without aid of beast, flying machines . . . [and machines that] make it possible to go to the bottom of seas and rivers” [quoted in White (1962, p. 130)].

<sup>2</sup> See further in Baumol (2002).

and the options for preservation and has influenced the amounts and sources of funding of the arts. Most obviously, the new economy has surely freed artists from dependence on royal and noble patronage and led them to turn for funding to perhaps equally risky new sources, including the marketplace.

I will focus here on two general subjects: the influence of technical innovation on the arts, and the consequences – actual and prospective – for the arts' ability to obtain funding in the marketplace. In particular, I will deal with several themes, including changes in the general orientation of the arts, their dissemination and preservation, and the compensation of artists. My discussion of the last of these topics will perhaps be the most substantive. That is because funding, pricing and other associated matters are most squarely in the terrain of the dismal science, and so lend themselves most easily to its analytic ministrations.

## 2. Technical innovation and its effects on the arts

### 2.1. Artistic creation and "modern art"

Patently, the attributes that most fundamentally characterize developments in the arts that have accompanied the new economy are exploration of the abstract and pursuit of the unconventional. And though some may consider the hypothesis outrageously philistine, it is at least plausible that these reorientations were initially stimulated to a considerable degree by technological developments. Is it so plainly untrue that the introduction of photography helped to undermine the pursuit of accurate or at least persuasive representation of reality in the visual arts? Even if the availability of photographic images did not lead artists directly into seeking other directions, surely the market must have driven them to new orientations by depriving them of purchasers – those clients who were no longer so anxious to pay for handmade and reasonably accurate depictions of reality, notably in portraiture.

Developments in other art forms such as discordant and atonal music, unconventional writings like those of James Joyce or Gertrude Stein, and modern dance forms introduced by Isadora Duncan and Martha Graham surely only followed in the wake of the first radical departures in the visual arts – the Impressionists, *les fauves*, the secessionists and their ilk – whose success in *épatant le bourgeoisie* must certainly have served to invite the practitioners of the other arts to follow. Evidently these remarks are a vast oversimplification; yet it surely cannot be a pure coincidence that the determined effort of creative artists to break so radically with their past seems totally unprecedented. There is apparently no parallel before the industrial age, which heralded the beginnings of the new economy. Of course, much of what I have said so far rests purely on conjecture and casual observation. This means that it merits no more than acceptance as a hypothesis that can serve as a subject for research by cultural economists. Technology's stimulus of artistic reorientation can probably never be determined by formal research methods and objective evidence, though there is much here that can be studied and

tested. And the subject is surely significant, for what is at issue is what may well prove to be the most dramatic example of the influence of economics on the orientation of the artist.

But there are other forms of reorientation in the artistic media themselves. The new economy has provided artistic creation with new tools, some of them already profoundly affecting what artists can do, some offering promise for the future. The introduction of motion pictures, for example, has played a significant role in the creative process, in addition to its enormous contribution to dissemination and preservation. As is widely recognized, a screen performance has fundamental differences from a performance on the stage. A straightforward film of a staged play usually constitutes uninspired cinematography. Moreover, film provides a rich menu of possible new approaches, and the history of the cinema provides a multitude of well-recognized examples. The cinema really is a new art form that had to be thought through and mastered. Arguably, many decades of cinematic creation were required before those engaged in the activity began to see its possibilities clearly and learned how to act on them. Television was a clear runner-up.

And there are more-radical breaks. Composers such as Milton Babbitt now regularly create works to be performed by electronic instruments such as the synthesizer guided by the computer. Indeed, there have even been experiments in composition using computers as the composers. Painting in which the computer serves as the medium is now widely available and promises to open roads to new directions. Computer painting and graphics open up vast possibilities, though it may be argued that up to now these media have exercised rather more control of the artist than the artist has been able to exercise over the media. Computers can now also be used to compose and guide the performance of intricate lighting design for the dance. Choreographer Mimi Garrard and visual artist James Seawright, for example, have created a powerful program for this purpose. Undoubtedly, all these are only beginnings, and there are in prospect artistic instruments and art forms that we have not yet imagined.

## 2.2. *Dissemination*

### 2.2.1. *Availability of the arts to consumers*

Printing with movable type (as invented belatedly in Europe) is generally recognized as a revolution in the dissemination of ideas and ultimately of literature. But while it has been credited with having become in relatively short order a stimulus for upheaval in religion and politics, its speed in making reading matter widely available was considerably less than is generally realized. The early books produced with movable type were hardly inexpensive, and since the vast bulk of the population was illiterate, printing's initial role in the spreading of writings was severely circumscribed. In any event, the printed page was not inherently different from the page produced by a medieval scribe. The difference was only (!) that once one produced a single printed page, it became

relatively easy to provide many others (and to do so without a successive accumulation of copying errors).

But from the 19th century on, there have been revolutions in artistic dissemination that faced far fewer constraints and that, in terms of their break with the past, were even more dramatic. Recordings, film, radio, television, and now the Internet, have produced magical change – total breaks with anything known in the past. Yet the speed with which we have come to take almost for granted these mass media and the new means of communications is itself impressive. Of course, many people do not yet have access to the Internet, but even in remote villages in inaccessible geographic locations movies are shown and people gather round the community's television set. Telephone service, too, has penetrated widely, and machines for the playing of recorded music are available. This evidently means that music and drama are obtainable as never before, and in ways that constitute a complete change from past experience.

When a Renaissance monarch was negotiating for a bride, he could demand a likeness prepared by a painter, and would obtain one that had taken weeks to produce, perhaps weeks to transport, and was probably an idealized version of the subject. Today, the ways around these difficulties are many and obvious. As one economist has remarked as an illustration of the point, "Could Thomas Aquinas have . . . dispatched [a letter] to 1000 recipients with the touch of a key, and begun to receive replies within the hour?"<sup>3</sup> These new means of communication are not only able to transmit the written word. They bring music, paintings and photographs to particular intended recipients or they can be broadcast to the world. Not only are announcements of art exhibits in distant places readily available, the Internet user may be able to download reproductions of some of the exhibited items. And, from anywhere in the world, the armchair art enthusiast may now view art collections like the Tate Gallery (<http://www.tate.org.uk>), and even take a virtual tour of the Louvre Museum (<http://www.louvre.fr>). A personal example is constituted by my own recent paintings, executed entirely on the computer and posted on my web page, which regularly elicit comments from China and elsewhere from people I do not know.<sup>4</sup>

Why spell out these observations, when I am well aware that any reader of this piece will already know all this? There are two reasons. First is the fact that these incredible developments have become so commonplace, if not quite totally banal. Second is the more obvious observation that the new economy has, indeed, made all forms of art accessible to a degree beyond anything previously experienced. This is clearly a benefit to the creative artists, who surely do not want have their works ". . . born to blush unseen and waste [their] sweetness on the desert air".<sup>5</sup> But at the same time the explosion in the dissemination of the arts is the source of serious difficulties that threaten to exacerbate significantly the unceasing economic problems of the arts, as I will discuss later.

<sup>3</sup> J. Bradford DeLong, *The Economic History of the Twentieth Century: Slouching toward Utopia?*, draft copy, <http://www.j-bradford-delong.net>, accessed September 2001.

<sup>4</sup> [www.econ.nyu.edu/user/baumol/w/](http://www.econ.nyu.edu/user/baumol/w/) (advertisement).

<sup>5</sup> From Thomas Gray, "Elegy Written in a Country Church-yard"; see Tovey (1904, p. 33).

### 2.2.2. *Artistic products in international trade*

Not only do economic and technical influences affect the arts, but the arts have achieved a new and expanded role in the economy – the relationship has become a two-way street. Rather than being the almost exclusive preserve of a narrow privileged group, the arts have expanded in the consumption patterns of society as a whole. Their contribution to GDP and to employment has evolved, and their role in other arenas has evidently changed as well, sometimes markedly. The dissemination revolution has been one of the foundations of these developments, most strikingly shown by the emergence of artistic products as significant elements in international trade. It used to be understood with good reason that only agricultural and manufactured products lent themselves readily to international exchange. An artistic activity such as the performance of drama or orchestral music is neither a manufactured nor an agricultural product. Neither medical examinations nor haircuts supplied in the UK are readily consumed by the inhabitants of Los Angeles, and so for a similar reason it used to be true in the nineteenth century that the New York Philharmonic Orchestra felt itself under little competitive threat by an orchestra whose home base was Berlin.<sup>6</sup>

Matters have now changed for many of the services. For example, engineers in Seattle may now email the specifications for a new product part to New Delhi, where the blueprints will be drawn up within 24 hours and transmitted instantly, again via e-mail, to a manufacturing plant in Ireland. But it is in the popular art forms that the most dramatic manifestation of this change is to be found. Though these activities could surely have been no more than a flyspeck in the national economic accounts of the US in the 19th century, nowadays via the vast “entertainment” universe – including movies, videos, television programs, sound recordings, computer software, and the publishing industry – the popular art forms constitute an important economic activity. They are now a principal US export – with revenues from foreign sales reported to be somewhere between about \$25 billion [US Census Bureau (2003)] and \$85 billion [International Intellectual Property Alliance (2002)], depending on the definition of the scope of this sector. Indeed, the very substantial competitive incursions made possible by easy dissemination of these art forms have become a contentious and much publicized political issue in a number of countries. It is felt that cultural exports, particularly from the US, threaten native artistic activities and undermine distinctive local culture and its traditions. The response has sometimes been the adoption of quotas that affect both performance and imports.<sup>7</sup> But for the issue under discussion here, these competitive incursions constitute clear testimony to the degree to which the new economy has been able to modify the role of the arts in the economy.

<sup>6</sup> Still, one must make such statements cautiously. It is noteworthy that what has been described as the worst riot in the history of New York City occurred in 1849 in Astor Place near Broadway and caused widespread injury and property damage. It involved a battle between the partisans of two popular actors, one American, the other British, performing in theaters located near one another.

<sup>7</sup> See further in Chapter 33 by Acheson and Maule in this volume.

### 2.3. *Preservation*

The broadly interpreted “new economy” has evidently also changed preservation of the arts beyond anything previously experienced and probably beyond anything previously anticipated. This was in fact brought out by the example mentioned earlier: the revolutionary role of sound recording. We presumably never will have any clear idea of what medieval song really sounded like, and can only make surmises about some of the improvisation and ornamentation in 18th century musical performance. We do not know how Elizabethans pronounced the English language, despite clever conjectures from what must ultimately be shaky evidence such as rhyme in 16th and 17th century English poetry and current speech patterns in relatively isolated communities. Certainly, we have little idea about how a contemporary performance of *Macbeth* would have sounded.

Worse yet, there were creative products that were rendered completely ephemeral by the absence of means of preservation. The works of 19th century choreographers clearly do not survive except via tradition whose modifications and evolution one cannot readily evaluate. Film and the invention of dance notation have since changed all that. There are, of course, limits to what film can describe of an activity in which three dimensions are so fundamental. But if the new forms of preservation cannot tell us everything, they can at least give the viewer a good idea of earlier practices, and can offer enough information to later choreographers to permit them to do a reasonably good job in reconstructing the work of their predecessors.

The computer and the accompanying developments of storage capacity have done far more. There is now virtually no effective limit to what can be taken into the world’s collective cultural memory. The contents of entire libraries can be transferred to minuscule recording media. More and more works of art are preserved in some image form, so that nowadays when a painting turns out to be missing, its likeness is nevertheless apt to be readily available. The Mellon Foundation, under the leadership of William Bowen, is engaged in a major digitization project in remote areas of China, partly to preserve their inaccessible art works, partly as a demonstration project for others to undertake.<sup>8</sup> In sum, future “Seven Wonders of the World” will not be so readily lost, their configurations available only to frustrated conjecture as has been the case in the past.

### 2.4. *Funding problems and the cost disease*

The last of the influences of the new economy on the arts that will be discussed here is financial. Growing wealth in the industrialized nations may have facilitated the financing of the arts, but other influences have made it more difficult. For example, ease of

<sup>8</sup> However, the declaration of victory may be premature. Programmers have indeed invented fantastic methods that can store works enduringly in binary code. But computer programs are replaced and superseded with remarkable rapidity. It is not clear that in 50 years anyone will be able to read what we painstakingly store today. Specialists are deeply concerned about this problem, and have not yet provided a solution.



copying has made it harder in many artistic fields to recover costs and earn a living in the marketplace. Furthermore, given what I and my associates have perpetrated in the past in our writings on cultural economics, I will no doubt be expected to argue that it is primarily through the “cost disease of personal services” that the new economy creates financial problems for the arts. Taken at face value, this is indeed true; the cost disease asserts that, at least for live performance,<sup>9</sup> costs are expected to rise cumulatively and almost constantly, and surely that is the *bête noir* of the impresario and the producer. Yet, as will be argued below, the new economy brings both the disease and the means that enable society to deal with it. In other words, the cost disease is, and yet need not be, a primary concern for funding of the arts.

First, a few words on the relation of the new economy to the cost disease are called for. It is easily argued that without the advent of the new economy, in the broader sense in which I am using the term here, there would be no cost disease. It may be recalled that the explanation of this phenomenon lies in the pattern of productivity growth; the reason live artistic performance – like health care, education and other activities with handicraft attributes – suffers from the cost disease is simply that the growth in their productivities, i.e., their rate of labor saving, inherently tends to be markedly less rapid than that of other products in the economy. With their productivity falling behind, it is mere tautology that their relative real cost – that is, the relative input quantities they require – must be rising. If input prices throughout the economy are determined in markets that are at least moderately competitive, the relative monetary costs of those outputs must rise as well. Before the advent of the new economy, when productivity growth was truly negligible in virtually every economic activity, the differences in these growth rates hardly mattered. Thus, if there were any cost disease, it would have been a very minor disturbance that hardly merited attention. It is in the new economy with its unprecedented and accelerating growth that the issue becomes significant.

Happily, the nature of the source of the problem brings with it the means by which the effects of the disease can be dealt with. Even the activities that are beset by the disease generally benefit from *some* productivity growth, however modest. But if even these, along with most of society’s other economic activities, are experiencing productivity growth, it follows that the community *can* afford more of each and every one of its outputs, if it chooses to allocate its purchasing power appropriately. With more output per hour of labor in every field of endeavor, there is no necessity to cut back on the consumption of *any* of the products. It is only a sort of money illusion that tells us we must cut back on the arts (or health care or education). Thus, it is the new economy and its sensational productivity growth that presents us with the cost disease problem, and it is that very productivity growth that provides us with the means to cope with its consequences.

<sup>9</sup> It will also be recalled that the mass media are not generally immune from the disease, because the handicraft components of the activities entailed in their operation tend asymptotically to dominate their budgets. On this see, e.g., Baumol, Blackman and Wolff (1989, Chapter 6).

### 3. Pricing issues for the arts in the new economy

The profound upheavals in the circumstances of the arts under the influence of the new economy have to a substantial degree been favorable, making the arts easier to create, to disseminate and to preserve. But some of these developments also have another side that threatens to handicap rather than help the arts. These issues have been recognized and analyzed by economists, but there is an underlying structure to these problems that seems not to be universally recognized. It is best brought out by analogy with another side of the new economy: the important role of innovation and the problems of its financing.

It is not without reason that the term “innovative” is often used to describe a new piece of music, a recently published novel, or some other product of artistic endeavor. In music or dance or literature, a creative contribution has one inviolate attribute – it cannot replicate some previous work. Unlike the production of automobiles or shoes, in which identical products can be turned out in apparently endless succession, the choreographer’s efforts (like the research and development division of a firm) must provide creations, each of which differs significantly from each and every one of the other products in the arena. “Innovation is a heterogeneous product”, runs Baumol’s third tautology, and the tautology is evidently applicable to artistic production. Now, this requirement of heterogeneity is the source of a number of problems both for practice and for analysis. We are all aware of the complications it poses for value theory particularly in general equilibrium form, problems that have only begun to be dealt with in recent years. And it may be that such problems help to explain the paucity of formal theoretical analysis of pricing in the arts. Perhaps cultural economists working on these problems may conceivably find it helpful to turn for guidance to the growing literature on general equilibrium analysis of monopolistically competitive markets and the equally valuable literature in the theory of technical innovation.

#### *3.1. Repeatedly sunk costs and pricing problems*

But that is not the issue here, since my concern is more with practice than with theory. And from this point of view there are at least two critical issues. One is the public-good property of information, a property shared with many outputs in the arts, and the second is the role of sunk investment of a special sort – not the kind of once-and-for-all sunk outlay that economists quite rightly tend to ignore as irrelevant ancient history once it has been incurred. For in technical innovation as in artistic creation, the need never to replicate, or at least not to do so precisely, means that investment is required each time an output emerges. That investment is sunk in the production of that novel item. Thus in both fields the need to sink costs is inescapably a frequently repeated if not continual phenomenon. Moreover, it is a phenomenon that appears to have grown in magnitude in the new economy. The investments that are routinely incurred in the production of a new film or in preparation to launch a new television series dwarf anything previously experienced in live theater, at least in absolute magnitude. In the launching of a popular

song via CD, the size of the advertising outlay can also be staggering; the CD manufacturers and distributors complain that the production of a brief video publicizing the piece becomes a major cost component.

For the economics of these activities, repeatedly sunk costs are fundamentally different from the sunk costs more usually mentioned in the literature – those that are incurred only once, as, for instance, in the purchase of a license required to open a business enterprise. The economic literature stresses that such costs once incurred are irrelevant for current pricing decisions because by their nature it is now too late to do anything about them, even if management feels it has made a mistake in undertaking the outlays in the first place. If the firm does not subsequently earn the amounts required for recoupment, that is the investors' hard luck. But when sunk costs must constantly be repeated, they must be covered by current prices or by some other means – otherwise a profit-seeking firm will refuse to repeat the outlay when the time for it comes, and a non-profit enterprise will not be able to do so.

This normally means that prices in these fields can be expected to exceed marginal costs because, as will be noted again in the next section, neither increased use by current users nor an increase in the number of users requires any increase in the sunk outlays; that is, these outlays constitute no part of the marginal cost figures.<sup>10</sup> Even in the non-profit arts organizations we find that prices do often exceed marginal costs; for example, the price of a ticket to concerts by a performing group whose auditoriums are not filled to capacity will usually substantially exceed the virtually-zero cost of admission of another attendee. Indeed, in such circumstances discriminatory prices are frequently required to make ends meet. In the arts, such pricing typically takes the form of discounts to students, the elderly and other special groups, last-minute price reductions, and other familiar variants.

Of critical importance here is the relative magnitude of the sunk cost in comparison with the other costs. This is so because of the public-good property, which means that the sunk costs are not increased by additional use or additional users; in dance, for example, the effort that went into choreographing and rehearsing are not increased either by a rise in the number of attendees at a particular performance or by an increase in the number of performances. Hence, if the repeatedly sunk costs are very substantial relative to other costs, their recoupment will require prices that are well above the marginal costs.

<sup>10</sup> Clearly, repeatedly sunk costs are generally not considered by firms to be the same as costs sunk once and for all, but rather as costs of operation of the firm. That is, they cannot simply be disregarded by a profit-maximizing firm on the usual argument that they are no more than a piece of unchangeable and therefore irrelevant ancient history. Nevertheless, repeatedly sunk costs do not enter marginal costs, because they are not a function either of scale of output nor of number of customers. Thus, the cost of producing the sets of a stage play do not rise with an increase in attendance, at least up to the capacity of the theater, nor with the number of performances per week nor even in the number of performances in total (at least until replacement is needed).

This, in turn, gives rise to two other difficulties. The first is the classical problem of public goods: Even if such higher prices were feasible in the market, their welfare consequences are questionable at best. Higher prices discourage use – that is the economic role they serve in the presence of scarcity. But in goods with public-goods properties, the standard form of scarcity vanishes. A public good has the property of a Hanukkah oil lamp, kept continuously alight without depleting the fuel supply. Additional attendees in a half-filled auditorium do not increase production costs, so why ration attendance with prices that exceed the marginal costs that an added attendee imposes? That is one part of the normative side of covering the sunk costs, even if they are regularly replicated. But of course if there is inadequate recovery of costs that must be repeatedly sunk, will not the supply process be undermined, with all potential attendees the losers?

In practice, there is another problem that often can prove even more urgent. The need to sink costs repeatedly threatens survival of firms if competition drives prices towards marginal costs. Fortunately, the market mechanism will generally permit recoupment even where entry costs are the same for all entrants, because no prospective entrant will be willing to begin operations in an arena where the prospects for recovery of sunk outlays are dim. Hence, contrary to the standard conclusion, here the threat of entry will not preclude either prices that exceed marginal costs or even prices that are markedly discriminatory.

But where all firms are not created equal, low marginal costs when combined with substantial sunk costs are an irresistible temptation to entry. However, the entrants do not seek to operate on the same terms as those of the already incumbent suppliers. The entrants brought in by the cost disparities are those who seek to profit by covering the marginal costs and escaping the sunk investments. In short, they are the “pirates”, those who hope to prosper parasitically from the sunk investments repeatedly incurred by the providers of innovations and artistic creations.

In technical innovation, the problems just described are often rendered acute by great disparity in the magnitudes of the two pertinent types of costs. In computer software, for example, estimates are that even the most successful firms’ annual expenditures on R&D can amount to something like 20 percent of revenues. But the marginal cost of supplying a copy of a new computer program to an additional user is negligible – it can amount to a fraction of a dollar. In some of the arts, such disparities can also be large. For instance, in the theater the cost of mounting a new production can equal a substantial proportion of the running costs for a year of performances. But in other art forms, partly because the creative artists traditionally often bear much of the sunk cost themselves, the monetary portion of the costs that need to be repeatedly sunk may not be so substantial.

Here, yet another feature of the supply process becomes critical. In a number of art-forms the middleman plays an important, often an indispensable, role, and it is from the sunk costs incurred by the middleman that the financial issue just discussed takes its most acute form. This is most obviously so in the mass media. The firms that make and distribute musical recordings are a clear example. In popular music, just the investment in the video to publicize a new recording, as already suggested, can be very large – it

can run upwards of a million dollars. And the bulk of the real remaining costs, including the efforts of the performing artist, are incurred in the production of the recording, all adding up to a substantial sunk cost that needs to be repeated each time an additional recording is turned out. Indeed, since both musical performances and computer programs are distributed on CDs, the reasons for the comparably negligible marginal cost should be equally obvious here. Similar relationships now hold in movies, where expenditure of many tens of millions of dollars per production is common and well recognized. But the creation of a tape of a movie that can be played on the home VCR is, once again, negligible.

Even if the financial problems stemming from the difference between sunk outlays and marginal costs are not of very direct importance for the creative artists themselves, they are generally urgent for the middleman enterprises. Furthermore, the significant role of the intermediary firms in many of the arts means that when the problems affect the one type of enterprise, they must ultimately have a reflection upon all. Financial pressures that fall initially upon the middlemen will not remain confined to them alone; ultimately, they will be shared by others – composers, choreographers and most severely by those in the weakest position to defend their interests. I will return to this part of the story presently.

### *3.2. Pricing under copyright and compensation of creative activity*

Copyright is, of course, the instrument that has been invented to protect the rights of creative artists and to stimulate the creative process. In practice, it is directly helpful primarily to those who work in the “popular arts”, since the market for the work of artists in less popular venues is, as we know, generally quite limited. But even the latter are affected indirectly by the support they receive from the mass media. Script writing, composition of background music for film and television, and classical music recordings do at least help to finance activity in what may dispassionately be referred to as the “elite art forms”. But the amounts that the mass media are willing and able to pay out in these ways is dependent on what they are able to earn from their activities in the more popular art forms and the degree of protection of those earnings that is contributed by the copyright laws.

Recent developments in the new economy have put enormous strains on these protective powers. Among the accomplishments of the new economy that have already been stressed is the remarkable progress in ease of reproduction, including the reduction of its costs to negligible levels, and the ease and speed of transmission of the reproduced material. The photocopier has already caused great difficulties for publishers, as has the Internet for the producers of recorded music. There is very likely to be a race between the design of means to prevent illicit reproduction and the instruments the reproducers have at their disposal. The outcome is unpredictable, but the relative position of the two groups may well oscillate with the advantage going from one to the other and then back as time passes.

One extreme outcome may be partial collapse of current commercial institutions and invasion of the fields by amateurs who undertake the activities out of dedication and personal interest rather than pursuit of gain. The free software (“shareware”) movement, most notably the development of the Linux operating system, may be a foretaste of such a future. So far the prosperity of firms such as Microsoft, some of the film studios and the recording industry indicates that the prospect is not yet imminent, and shows that some such enterprises have managed via technical devices and with the support of the courts to stay one step ahead of the pirates. But where such protection of the financial interests of the major suppliers is effective, there is the danger that matters may swing too far in the other direction. A copyright, after all, is nothing else than a grant of monopoly power, however virtuous its purpose. The question as usual comes down to a matter of pricing. What price ideally should the copyright holder be permitted to charge?

If the issue is a mere matter of recoument of sunk costs and efficiency of resource allocation in the standard sense, economic analysis provides a solution which is now well known: the Ramsey price, which is the second-best price consistent with a given revenue requirement constraint. In the simplest case in which the cross elasticities of demands are all zero, the supplier’s socially-ideal product prices will exactly exceed marginal costs in inverse proportion to the demand elasticities of the products, up to the point where the resulting rise in revenues is just sufficient to recoup the sunk costs. For middleman firms this may be the right solution, though as we will soon see, here ideal pricing policy runs into another fundamental complication.

But before getting to this, a bit more can usefully be said about the Ramsey price of an artwork with public-good properties. In the extreme case of the ideally-pure public good, the marginal cost of increased usage will be zero, as in the case of an added audience member in an uncrowded theater or an additional reader of a poem or viewer of a painting. Then, the second-best pricing solution to the simplest Ramsey model clearly becomes

$$\frac{P - MC}{P} = \frac{P}{P} = 1 = \frac{1}{E} = E = -\frac{\frac{dQ}{dP}}{\frac{P}{Q}}, \quad \text{or} \quad \frac{Q}{P} = -\frac{dQ}{dP}. \quad (1)$$

That is, in this case the second-best welfare solution is to set price so that the quantity of sales elicited is that at which demand is unit-elastic. This means that the more responsive the quantity demanded of the product in question is in absolute terms to changes in price, i.e., the greater the value of  $-dQ/dP$ , the lower the price should be relative to the quantity sold. The intuitive explanation is that the objective here is to raise prices above their first-best levels sufficiently to increase revenues so as to cover the unavoidable costs, but to do so with minimum distortion of quantities demanded from their first-best magnitudes, the magnitudes that would be elicited by zero prices. This means that prices should be positive<sup>11</sup> but lowest for those items whose demand quantities are most substantially modified from their first-best magnitudes.

<sup>11</sup> Price evidently should be zero for any good for which the demand elasticity is infinite; for more on Ramsey pricing of pure public goods, see Baumol and Ordover (1977).

On the other hand, where the objective of price setting is not recoupment of some determinate sum but stimulation of creativity, we do not know the answer. Partisans of the arts will be tempted to say "the more the better", and in practice one may well argue that the typical compensation of the creative artist is sufficiently niggardly as to leave us in the range where this answer is valid. But it is conceivable that the allocation of talent to creative activity will be excessive. This is an obvious possibility in the case of technical innovation where the greater part of the labor force still must be engaged in production for the present rather than for the future, for the future is what invention primarily serves. But there are economists who argue that even now the pressures of the new economy are apt to result in excessive allocation of resources to innovation, because the loss of value of still-usable but obsolescent assets is what creative destruction entails. *A*'s innovation is likely to render obsolete an asset that is not her own, but rather belongs to other individuals, *B*, *C*, and *D*, who have no role in *A*'s activities. Thus, creative destruction is a detrimental externality and like other such externalities in a free market it is likely to lead to excessive output of the externality-creating product.<sup>12</sup>

In the arts, it can also be argued that there is no shortage of creative activity. There is a profuse stream of plays written without let-up, and paintings (albeit of uneven quality) are produced in vast numbers. The problem, apparently, is not further stimulation of the supply but creation of venues in which the products of creative activity in the arts will be seen and appreciated. In short, not only are we unsure of the effectiveness of copyright in benefiting the creative artists but, particularly in the new economy where creative artists are so dependent on market forces, it is not clear what prices it is desirable for copyright to yield.

### 3.3. *Efficient component-pricing*

It should be clear from the above that the determination of appropriate prices for output supplied under copyright is not really the critical issue in relation to compensation of the individual creative artist. Rather it is the giant middlemen – the film studios, the television broadcasters and the producers and distributors of musical recordings – where the matter is of substantial interest. But even here it may well be argued that the issue is not one of exploitation of monopoly power, because competition can still be a force sufficiently powerful to drive profits down toward the competitive level in the economy, even though the products are markedly heterogeneous. However, in the new economy, given the critical role of the mass media, there is an additional complication that besets the pricing issue. Others engaged in cultural activities will want permission to reuse material covered by copyright for employment in their own pursuit of profit. For example, films and recorded television programs are often replayed and rebroadcast by others than their proprietors, and recordings of music are played by broadcasters all the time. In some cases, the users are direct competitors of the owners of the copyright.

<sup>12</sup> On this see, notably, [Aghion and Howitt \(1998\)](#).

The question is whether there exists an economically efficient price for such access to copyright material.

The answer is that there does indeed exist such an efficient price. Though it will not be derived here,<sup>13</sup> it seems worthwhile to describe the efficiency consequences that are at issue. From the point of view of abstract economics, a piece of intellectual property such as a technical invention or a musical composition is interpretable as just another input into some production process where the final output is a product that incorporates the invention, or a concert or a film in which the music is played. Now, it is clear that the firm that invested in the R&D that produced the invention is not necessarily the invention's most efficient user in the production of the final product, just as the composer of the music is not necessarily the music's most gifted performer. The price that the holder of the copyright charges others for use of a musical recording or rebroadcast of a TV performance determines whether or not that activity will devolve upon the most efficient user. For example, the composer who demands a prohibitive price may end up as the sole performer of his piece, even if he is far from the ideal person for the task.

The theory of economic regulation provides a formula for the efficient price for access to such proprietary intellectual property. The formula is called the "efficient component-pricing rule" (ECPR), and yields what has come to be called the "parity price". This rule has elicited considerable discussion in the literature on economic regulation, and has been adopted by a number of courts and regulatory agencies in several countries, though other legal opinions have rejected it. But it is an issue of some importance for the economics of the arts, and it is one that yet again arises primarily in the new economy. Specifically, ECPR requires that the license price must equal the copyright owner's final-product price minus the copyright owner's incremental cost of remaining inputs, i.e., that the licensing price satisfy the following rule:

$$P_i = P_{f,i} - IC_{r,i}, \quad (2)$$

where:

$P_{f,i}$  = the copyright owner,  $I$ 's, given price per unit of final product;

$\min P_{f,c}$  = the competitor,  $C$ 's, minimum viable price of final product;

$P_i$  = price charged for a license to use the copyright, per unit of final product;

$IC_{r,i}$  = the incremental cost to the copyright owner of the remaining final-product inputs, per unit of final product;

$IC_{r,c}$  = the corresponding figure for the competitor; and

$IC_i$  = the incremental cost to the copyright owner of use of the copyright by itself or by others.

Equation (2) establishes a tight link between the price  $P_{f,i}$  that the copyright owner charges for its final product and the ECPR price  $P_i$  that it charges its rivals for the license to use the copyright. If incremental production costs do not change, efficiency requires that a rise in one of these prices must be matched dollar for dollar by a rise in the other.

<sup>13</sup> For a full exposition, see Baumol (2003).



It is trivial to prove the efficiency property of the rule via the Level Playing Field Theorem, which is stated as follows: The parity price as given by Equation (2) for use of material covered by copyright is both necessary and sufficient in order for the “playing field” to be level as between the copyright owner and any licensee. This means that the maximum difference between the remunerative prices of the perfect-substitute final-products of the two firms, the copyright owner ( $I$ ) and its final-product competitor ( $C$ ), is exactly equal to the difference in the firms’ remaining incremental costs (other than the license fees). That is,

$$\min P_{f,c} - P_{f,i} = IC_{r,c} - IC_{r,i}. \quad (3)$$

In other words, an ECPR price is necessary and sufficient for the lowest compensatory price the competitor can afford to charge in order to differ from the copyright owner’s exactly by the amount (positive or negative) that the former’s remaining costs are below the latter’s. This means that the competitor will (will not) be able to take over the use of the licensed product and the market for the resulting output if and only if it is the more (less) efficient user of the licensed item for the purpose. This clearly illustrates how the standard approaches of economic analysis can be used to solve one of the efficiency problems in the arts that the new economy has brought to the fore.

#### 4. Other problems contributed by the mass media

Through the creation of the mass media, the new economy has contributed other problems for the arts that have already received considerable discussion in the literature. Because they have been analyzed by others, I will limit my discussion to two issues here. The first is the exacerbated “star” system and its influence on the distribution of income in the arts,<sup>14</sup> and the second is the contraction and partial disappearance of secondary performance venues that served (like the minor leagues in sports activity) as training grounds for creative artists and performers.

##### 4.1. Superstars

Of course, there is nothing new about the existence of stars in the arts, star performers, star composers, and so forth. What the new economy has created is the prospectively vast audience that these stars can now draw, an audience that would not exist without the mass media. It used to be that an audience numbering in the lower thousands was considered sensational. With the mass media, millions are now the relevant order of magnitude. This means that the marginal revenue product of the stars may perhaps have been multiplied a thousand-fold, a fact that is often reflected in the remuneration they command. The lesser mortals among creative artists and performers can be expected

<sup>14</sup> See further [Chapter 25](#) by Adler in this volume.

to receive a much more modest multiplication of their fees, since viable substitutes for their services are far more readily available. The consequence has been an enormous increase in disparity of income between stars and lower-ranking performers and, often along with it, some increase in the power of the stars to command their conditions of employment, as well as the artistic uses to which their work will be put. Whether the effect of this development on artistic quality has been advantageous or disadvantageous is difficult to say, however.

#### 4.2. *Contraction of secondary performance venues*

Somewhat more easily assessed, though probably no more easily documented, is the implication for artistic quality of the second consequence of the advent of the mass media mentioned above. Because mass media enable star performances to reach such vast audiences, in many venues the demand for second-tier performances has dried up. For example, vaudeville stage entertainment used to be a pervasive activity, with live performance interspersed with film performances in theaters throughout the land and with performers constantly traveling in great numbers from one city or smaller community to another. Today the vaudeville theater is dead. Even more striking is the disappearance of live amateur concerts in private homes. A century ago in many a middle-class household several members of the family could play musical instruments with tolerable skill. Today children still take music lessons, but often their performances are listened to primarily to shore up egos, while it is the CD player that provides the musical performances that are really wanted by the listeners.<sup>15</sup> One consequence is the shrinking of a critical training ground. Another example is the effect on comic performances. The fabled comedians of the mid 20th century – the Marx Brothers, W.C. Fields, Jimmy Durante, Burns and Allen and many others – had their training on the road, and they refined their acts through much repetition, and through constant interaction with their audiences. The result was a product whose polish is considered by some observers not to be approximated by more recent performers. There are now other venues for budding comics: local comedy clubs, the *Comedy Central* cable channel, even *Saturday Night Live*, a television show from which quite a few unknowns have gone on to stardom. While one may well question whether this is enough to sustain the old level of professional polish, it may persuasively be argued that quality here is strictly a matter of taste.

Yet a suggestive analogy comes to mind. Elsewhere, my wife and I [Baumol and Baumol (1994)] have surmised that the extraordinary riches of the composers' activities in 18th century Vienna and much of Germany and the Austro-Hungarian empire was at least in part attributable to political fragmentation, with the multitude of courts of minor royalties providing jobs that attracted people into careers (sometimes menial jobs)

<sup>15</sup> Schools have made up for some of this loss, often providing significant alternative venues for young amateur musicians. It is not unusual for school-based choral and orchestral groups to perform in local communities, go on international tours, and produce CDs.

as composers, served as training grounds and as sorting instruments that could identify those who had special ability or promise.<sup>16</sup> This suggests that with the unification of Germany, the disappearance of this multiplicity of employment opportunities tended to circumscribe the country’s advantages as a location for the work of composers. The hypothesis I am proposing here is that the new dominance of the mass media in dissemination of the arts may have similar consequences.

## 5. Concluding comment

The central point of this chapter is that the new economy – the one that had its origins in the Industrial Revolution at the end of the 18th century – has indeed affected the arts and has not done so only marginally. In a variety of ways it has changed their circumstances beyond anything that humans living earlier could have recognized or imagined. The new economy has led to total upheaval in the technology of distribution, communication and preservation in the arts. It has given rise to profound and novel financial pressures. Arguably it has even had a major influence on the standards and goals of those engaged in creative activity in the arts. And, indirectly, in the new economy, the arts have had new effects on the structure of the economy. But what has been pointed out here has largely been impressionistic and discursive, entailing little data or formal analysis. I suggest that there are many elements in the relation between the arts and the new economy that merit considerably more serious exploration than is offered here. I hope that others will be tempted into following these leads.

## References

- Aghion, P., Howitt, P. (1998). *Endogenous Growth Theory*. MIT Press, Cambridge, MA.
- Baumol, W.J. (2002). *The Free-Market Innovation Machine: Analyzing the Growth Miracle of Capitalism*. Princeton University Press, Princeton.
- Baumol, W.J. (2003). “Applied welfare economics”. In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham, pp. 20–31.
- Baumol, W.J., Baumol, H. (1994). “On the economics of musical composition in Mozart’s Vienna”. In: Morris, J.M. (Ed.), *On Mozart*. Woodrow Wilson Center Press, Washington, DC, pp. 72–101.
- Baumol, W.J., Ordover, J. (1977). “On the optimality of public-goods pricing with exclusion devices”. *Kyklos* 30 (1), 5–21.
- Baumol, W.J., Blackman, S.A.B., Wolff, E.N. (1989). *Productivity and American Leadership: The Long View*. MIT Press, Cambridge, MA.
- Blaug, M. (2001). “Where are we now in cultural economics?” *Journal of Economic Surveys* 15, 123–143.
- Frey, B.S., Pommerehne, W.W. (1989). *Muses and Markets: Explorations in the Economics of the Arts*. Basil Blackwell, Oxford.

<sup>16</sup> But see Scherer (2001, 2004), who uses historical statistical evidence to argue that the activities of composers who relied primarily on the market rather than primarily on the patronage of the courts were even more fruitful.

- International Intellectual Property Alliance (2002). "Copyright industries in the US economy". The 2002 report, prepared by Siwek, S.E. Available at <http://www.iipa.com>, accessed May 2003.
- Peacock, A. (1992). *Paying the Piper: Culture, Music and Money*. Edinburgh University Press, Edinburgh.
- Scherer, F.M. (2001). "The evolution of free-lance music composition, 1650–1900". *Journal of Cultural Economics* 25 (4), 307–319.
- Scherer, F.M. (2004). *Quarter Notes and Bank Notes: The Economics of Music Composition in the 18th and 19th Centuries*. Princeton University Press, Princeton.
- Throsby, D. (1994). "The production and consumption of the arts: A view of cultural economics". *Journal of Economic Literature* 32, 1–29.
- Tovey, D.C. (Ed.) (1904). *Gray's English Poems*. Cambridge University Press, Cambridge.
- Towse, R. (1997). *Cultural Economics: The Arts, the Heritage and the Media Industries*, 2 vols. Edward Elgar International Library of Critical Writings in Economics, Edward Elgar, Aldershot, UK.
- US Census Bureau (2003). *Current Business Reports, SAS/01, Service Annual Survey: 2001*. US Government Printing Office, Washington, DC. Available at <http://www.census.gov>
- White, L. Jr. (1962). *Medieval Technology and Social Change*. Clarendon Press, Oxford.

## CULTURE AND ECONOMIC PERFORMANCE\*

MARK CASSON

*University of Reading, UK*

### Contents

Abstract	360
Keywords	360
1. Introduction	361
2. Basic concepts and definitions	363
2.1. The definition of culture: Culture as a public good	363
2.2. Cultural diversity	364
2.3. Stereotypes	366
2.4. Culture as an asset	366
2.5. A typology of social groups	367
3. Towards an economic theory of culture	368
3.1. Leadership	369
3.2. Competition between groups	369
3.3. The changing nature of competition between leaders	372
4. Key dimensions of culture	373
4.1. Four main dimensions of cultural variation	373
4.2. Trade-offs involved in a high-performance culture	377
4.3. Refining the dimensions of culture	378
5. Method and history	389
5.1. Methodological issues in modeling culture	389
5.2. Historical perspectives	391
6. Conclusion	394
References	395

\* I am grateful to Janet Casson, Marina Della Giusta and the editors for comments on an earlier version of this paper.

**Abstract**

Culture, defined as shared values and beliefs, can influence the performance of an economy in many ways. The culture of a group, whether national, regional or ethnic, may be regarded as a particular type of intangible public good. The chapter summarizes and critiques a positive theory of inter-cultural competition. According to this theory, culture is created by leaders, who specialize in the production of culture, and is shared by their followers. Leaders compete for followers in order to increase the rents that they can extract from their groups. Whilst some of these rents may be pecuniary, most are non-pecuniary, such as the enjoyment of pursuing a public project which glorifies the leader and their group. There are four main dimensions of culture which influence performance, and there are trade-offs between them which are governed by the environment of the social group. The positive theory is useful in interpreting historical evidence on the rise and decline of societies, institutions, and organizations of various kinds.

**Keywords**

culture, institution, leader, transaction cost

*JEL classification:* H8, D7, P5

## 1. Introduction

It is popularly believed that culture has a significant effect on economic performance [Buruna (1999)]. Whilst some economic historians are sympathetic to this hypothesis [Landes (1998)], most economists are skeptical. They question the intellectual rigor of the underlying theory, and the objectivity of the evidence. In *The Wealth of Nations*, Adam Smith downgraded cultural factors from the prominent position they had occupied in his previous work, and subsequent economists have largely followed his lead [Macfie (1967)]. Recently, however, theoretical interest in the economics of culture has revived [Olson (2000)]. This chapter reviews attempts to bring greater rigor to the subject. It is argued that models of rational action, on which conventional neoclassical economics is based, can be extended to allow for cultural influences. Such models suggest that certain cultures promote economic performance better than others.

Culture may be regarded as an economic asset – a form of cultural capital. It is an intangible public good, shared by the members of a social group. The analysis below identifies four major dimensions of culture which influence the performance of a group:

- individualism versus collectivism,
- pragmatism versus proceduralism,
- the degree of trust, and
- the level of tension.

Individualism emphasizes personal autonomy, and echoes the former UK Prime Minister Mrs. Thatcher's dictum that 'there is no such thing as society', whilst collectivism asserts that it is natural for people to be socially embedded in a larger group. Pragmatism favors improvisation and flair in taking decisions, whilst proceduralism emphasizes reliance on rules. High trust reflects a belief that other people are honest and hard-working, whether they are supervised or not, whilst low-trust reflects a belief that people will take every profitable opportunity to shirk and cheat. The level of tension reflects the level of achievement to which people aspire, and their determination to succeed.

The analysis distinguishes between economic performance in a material sense, and overall quality of life. Quality of life depends on emotional as well as material rewards. Culture is not merely instrumental in the pursuit of material rewards, but is a direct source of emotional rewards as well. Boosting emotional rewards can also boost material rewards – as in highly-motivated teams – but there are trade-offs too: for example, a religion that encourages prayer and fasting may reduce material performance even though it improves quality of life. Bias in the measurement of the material living standards adds a further complication. A market economy may appear to out-perform a non-market economy in material terms simply because a higher proportion of its output is recorded in the national income statistics.

It is relatively easy to show that culture can have a positive effect on quality of life. Quality of life depends heavily on the provision of intangible public goods such as visual amenity, safety on the streets, and so on. Culture is not only a public good itself, but is instrumental in creating popular support for investment in other public goods. It is more challenging, however, to show that culture can improve the material output

of private goods, and it is this challenge that is therefore the focus of attention in this chapter. Furthermore, since material performance is easier to measure than quality of life, hypotheses linking culture to material performance are in principle easier to test.

Modern neoclassical economics implicitly endorses a Western culture of 'competitive individualism', which is individualistic and low trust. The collapse of Soviet communism and the 'triumph of the market' was widely interpreted as demonstrating the advantages of an individualistic culture over a collectivist culture. However it said nothing about the advantages or disadvantages of high trust. Until the 1970s, the justification for markets was seen mainly in their ability to adjust to incremental change. Globalization, however, precipitated major changes, and led to the growth of 'enterprise culture', which emphasized the value of pragmatic improvisation over routine procedure when taking key decisions. At the same time, Soviet communism remained wedded to procedural decision-making. Thus Western capitalism was not only individualistic but pragmatic, whilst Soviet communism was both collective and procedural. It is therefore unclear whether the superiority of individualism over collectivism, or pragmatism over proceduralism, was mainly responsible for the revealed superiority of the West.

The success of many newly industrializing countries in pursuing state-led export programs suggests that where government has been pragmatic rather than procedural it has sometimes been able to achieve remarkable results. It may therefore be that excessive reliance on procedure, rather than collectivism *per se*, caused the collapse of communism.

Western capitalism and Soviet communism were both high-tension cultures, whilst developing countries, on the whole, exhibit low-tension cultures. In the third world, high-trust culture seems to perform better than low-trust culture [Sherman (1997)]. Combining the lessons from these various comparisons therefore suggests that the most promising culture is individualistic, pragmatic, high-trust and high-tension. This is entrepreneurial associationism – a culture which encourages people to freely commit themselves to ambitious pragmatic team-based projects. It differs from competitive individualism in having a high level of trust. No country has been able to sustain associationism for very long, however, and so competitive individualism has emerged as a 'second best' solution.

High tension stimulates competition, which tends to undermine trust. It is sometimes suggested that trust arises naturally, through repeated interaction, but it remains the case that selfish individuals have a strong incentive to cheat in the final play of any 'repeated game'. If trust is to prevail generally, it cannot be regarded as natural, but must be engineered [Casson (1991)]. This is achieved by moral leadership, as explained below. From this perspective, lack of trust reflects a scarcity of leadership – indeed, there are grounds for believing that there is a systematic shortage of suitable leaders in most countries. An unfortunate legacy of inter-war Fascism is that the very concept of moral leadership has fallen into disrepute. This has discouraged the systematic production of moral leaders through education. Families and local communities have under-invested in the supply of leaders for future generations. Furthermore, it is argued below that the growth of mass media has distorted competition between potential leaders to favor those who appeal to



narrow self-interest. It is suggested that ineffective moral leadership has impaired the performance of Western economies over the last twenty years. Individualism and high-tension have been pursued to the point where they undermine trust, creating a consumer society marred by crime and anti-social behavior. Undermining trust has raised the costs of coordination, eroded material performance, and caused serious detriment to quality of life.

If this economic theory of culture is correct, and its diagnosis of events is sound, then the policy implication is that nations must improve the supply of moral leadership. Intellectual leaders such as priests, politicians, philosophers and artists all have an important role to play in stimulating the imagination of political and business leaders; in a successful society such intellectual leaders will tend to embrace a high-trust high-tension culture.

The chapter is organized in four sections. Section 2 introduces basic concepts and definitions; Section 3 outlines an economic theory of culture, concerned with competition between groups; Section 4 discusses the key dimensions of culture, whilst the final section examines broader methodological and historical issues.

## 2. Basic concepts and definitions

### 2.1. The definition of culture: Culture as a public good

There are many important contemporary economic issues in which culture is a significant factor, such as

- Is a common European currency a symbol of political unification?
- Will contracting out public services such as health to private firms undermine the public service ethic?
- What exactly is 'consumerism'? Do heavily advertised 'lifestyle' consumer brands delude consumers with false hopes, and does it matter if they do?

It is necessary to define culture in a way that captures the common elements in these questions. For the purposes of this chapter, therefore, culture is defined as *shared values and beliefs relating to fundamental issues, together with the forms in which they are expressed*. This suggests that there are three main aspects to culture:

- values, which represent the moral aspect of culture,
- beliefs, which represent the technical aspects, and
- forms of expression, which represent the symbolic and artistic aspects.

These values, beliefs and forms of expression are shared within a social group.

It can be seen that this approach to culture is more general than that employed in the economics of the arts. Arts tend to be identified with 'high culture', involving the expression of emotion through artifacts (e.g., paintings, books) and performances (e.g., drama, ritual). Culture, as defined above, relates not only to emotional responses, but to quite detached views connected, for example, with scientific topics. Furthermore, it

encompasses more than just expression – it includes the formation and dissemination of views as well.

Culture is an intangible good. Cultural values and beliefs can be shared, which indicates that culture, like knowledge, has the property of a public good [Reisman (1990)]. The fact that one person holds certain beliefs, for example, does not preclude another person from holding these same beliefs too. Thus there is no rivalry in the consumption of culture. Culture may be a good because it has intrinsic value, or because it is instrumental towards some other purpose. People may value certain beliefs because holding these beliefs makes them happy [Layard (1980); Easterlin (1998, Chapter 10, 2001)]. They may value other beliefs because they are purely instrumental – for example, holding correct beliefs eliminates mistakes and so reduces waste, thereby improving the material standard of living. It follows that culture can also be a ‘bad’. Some beliefs make people unhappy – for example, the belief that nobody likes them. Other beliefs may be damaging because they are wrong – mistakes are made when acting on these beliefs, and resources are wasted as a result. From an economic perspective, therefore, the elimination of cultural bads is just as important as investment in cultural goods.

## 2.2. *Cultural diversity*

Cultural diversity is a topic which generates considerable controversy. Conventional economic theory suggests that culture is simply a set of beliefs which will ultimately converge on correct beliefs as a result of learning. According to this theory there is a unique set of correct beliefs on which everyone will eventually agree; groups that refuse to learn will fail to survive. The only cultural guarantor of economic success is a correct economic theory and the implementation of policies derived from it. Some economists seem to believe that convergence on the correct theory is almost instantaneous. Adherents of rational expectations theory, for example, maintain that everyone holds correct beliefs because they already know the true model of the economy [Lucas (1981)]. Others allow the process of adjustment to take a little longer; they concede, for example, that the final collapse of authoritarian socialism in the 1990s occurred only after a century of institutional experimentation.

Simple economic models such as rational expectations, assume that information is costless to collect and communicate, and easy to verify. These assumptions about costless information are critical to the prediction that incorrect beliefs will be eliminated, and only correct beliefs survive.<sup>1</sup> However, any plausible economic theory of culture must recognize the significance of information costs. Whilst knowledge is a public good, it is costly to share. No one has complete access to all available knowledge. Costs of collecting information mean that everyone bases their beliefs on only a limited amount of information. Optimal search theory shows that once a certain amount of

<sup>1</sup> The rational expectations approach to economics is a recent innovation which is very much at odds with traditional mainstream writing, even in the Chicago School; see, e.g., Leacock (1998) and Viner (1972, 1978).

information has been collected, it is no longer cost-effective for an individual to refine their beliefs by collecting more. Beliefs are therefore based on a very limited amount of information.<sup>2</sup> Access to information can be improved by pooling information, but this requires communication between people, which is costly too. It is often more efficient to leave someone to discover something for themselves rather than incur the costs of telling them about it. Information sources are typically localized, which means that when people rely upon their own resources, different groups of people in different localities have different sets of information. Each group generates beliefs on fundamental issues by generalizing from its own experience. This leads to different sets of beliefs, and so to cultural diversity.

Cultural diversity is likely to diminish over time. Much information is a by-product of action – it is acquired through ‘learning by doing’ – and so accumulates over time. Additional information can be captured through scientific experiment. As a result, the information available to each group is likely to become more and more the same. Groups can also compare beliefs, and refine them through a process of criticism. In this way the accumulation of knowledge, combined with critical debate, encourages the emergence of consensus. Diversity cannot be eliminated, however, because there is a lack of decisive information on certain crucial issues. Evidence is decisive when it convinces not only believers but also skeptics. Much of the evidence used in social science is difficult to replicate because it cannot be collected under fully controlled conditions. It therefore lacks the ‘objectivity’ that would convince a skeptic. Lack of objectivity is particularly problematic in the investigation of fundamental issues such as the origin of consciousness, inequality of intelligence, and the relative importance of ‘nature’ and ‘nurture’. Lack of objectivity allows people to remain attached to beliefs which explain their own experience but not the experiences of others.

Disagreements are even more difficult to resolve in the field of values. Some value systems can be criticized for lack of consistency, although not everyone would accept that logical consistency is a requirement of a value system. Religious value systems often appeal to revelation and sacred texts as a source of authority, but secular critics deny their validity. Diversity in values therefore tends to be not only greater, but also more enduring, than diversity in beliefs.<sup>3</sup>

Overall therefore, fundamental problems in assuring the quality of information mean that despite the increased quantity of information that flows within the world economy, cultural convergence on a true model is unlikely to be attained. The spread of the internet, for example, may well promote convergence on relatively superficial issues such as the consumption of heavily advertised brands, but it is unlikely to promote convergence on more fundamental issues. Indeed, the proliferation of special issue lobbies such as

<sup>2</sup> Indeed, it is interesting to note that recent research has introduced costs of rationality into rational expectations modeling, which has aligned the approach more closely with that set out in this chapter; see, e.g., Ginsburgh and Michel (1997).

<sup>3</sup> For further discussion of the influence of diversity in values see Baxter (1988), Hahnel and Albert (1990) and O’Brien (1988).

anti-globalization protest groups coordinated through the internet suggests that increasing skepticism about the quality and integrity of 'official' information is generating new sources of cultural diversity. Thus while cultural diversity in international consumption patterns may be reduced through greater quantities of information flow, the limitations on information quality mean that intra-national diversity in political and religious beliefs may well increase.

### 2.3. *Stereotypes*

There is considerable popular awareness of differences between the cultures of particular groups of the same type. These differences are usually expressed in terms of stereotypes. A stereotype is an oversimplified characterization of a social group that ignores diversity within the group. It is a form of group reputation. The members of a group generally view their own group more favorably than do outsiders (which partly explains why they are happy to remain within the group). Indeed, competing groups often adopt negative stereotypes of each other in order to justify their antagonism. For this reason stereotypes are often condemned for promoting distrust between groups. However, different outside groups often hold rather similar views of any given group, which lends support to the idea that there is an objective kernel to the outsider's view. Thus although stereotypes ignore internal diversity and are often hostile, they are still useful because they usually contain significant insights too.<sup>4</sup>

### 2.4. *Culture as an asset*

Culture is a durable asset: values and beliefs are memorized by individuals, and are transmitted to the next generation through parenting and education. Education is strengthened when culture is recorded in books, embodied in art and artifacts, and embedded in rituals and routines. The durability of culture has encouraged some writers to see it as the 'dead hand of the past'. Culture is acquired from early childhood when critical faculties are undeveloped. People become very attached to their early beliefs for emotional reasons – loyalty to parents, a concern for their 'roots', or fear of change. Beliefs are not revised in the light of new circumstances and hence there develops a disjunction between culture and the real world. This view ignores the fact that people often review their beliefs in adolescence or when they come of age. It also has the misleading implication that a very old culture is likely to be less appropriate than a newer one.

An alternative view is that culture adapts to changing circumstances, but with a lag. It is sometimes suggested that a traumatic set-back such as a military defeat is necessary to undermine confidence in a culture. Defeated groups may sometime adopt their conqueror's culture (or selected aspects of it). On this view cultures which survive do so not because of mere inertia but because the beliefs they embody are more correct or more successful than those they replace.

<sup>4</sup> For the use of national stereotypes to analyze economic performance see Casson (1990, Chapter 4).

The most efficient way for a culture to cope with change is to adapt its beliefs in an incremental fashion, up-dating them in response to significant events and new discoveries. Monitoring the environment and up-dating beliefs is a complex task, however, and benefits from specialization. It is impossible for everyone within a group to find the time to continually re-examine their beliefs for themselves. To understand how culture changes, therefore, it is necessary to understand the division of labor within social groups.

### 2.5. A typology of social groups

The basic unit of cultural analysis is the social group: it is the unit within which culture is shared [Newman (1983); Pryor (1977)]. The most significant types of group from a cultural perspective are listed in Table 1. People are born into families and the local community where they live. They also acquire nationality at birth. When they come of age they can take decisions for themselves. They can choose the firm for which they work, the profession (if any) they wish to follow, and the clubs and societies they wish to join. They can also decide whether they wish to be active members of a church or a political party. In taking these decisions they affirm certain values and beliefs they have acquired from family and friends and reject others.

In a high-tension society, belonging to a group involves significant commitments; furthermore, in a high-trust society there are significant emotional penalties for breaking such commitments – disloyalty and lack of perseverance bring guilt and shame. Within a group there are distinctive roles. Roles with greater responsibility generally carry higher

Table 1  
Typology of social groups

Type of group	Membership system
Nation state	Citizen by birth or naturalization. Tax-payer by residence.
Market	All buyers and sellers of a product are members of the relevant market – especially competing sellers who locate close to each other.
Network	Member by regular contact with other members – often met through introductions arranged by existing members.
For-profit associations: firm	Member by negotiation. Core members supply services on a regular basis: e.g., shareholders and employees. Customers may be regular, casual, or one-off purchasers.
Non-profit associations: profession, club, church, charity, political party, etc.	Member by application, invitation, qualification or election.
Local community: friends, school, etc.	Member by residential location.
Family	Member by birth or adoption.

status. High-status people can demand deference from other members of the group. In addition, there are differences in status between different groups. Some groups are task-oriented (like the firm) whilst others are support-oriented (like the family), although most types of group combine elements of the two. In a task-oriented group the clients or customers who consume the output are usually different from the workers who produce the output, whereas in a support group the consumers and producers are often the same. In a charity, for example, the donors who supply the funds are quite distinct from the beneficiaries or clients who receive them, whereas in a support group like Alcoholics Anonymous the members support each other [Bolnick (1975)].

Clients usually have low attachment to a task-oriented group. Customers may have only casual contact with a firm, for example, whereas workers are heavily involved on a daily basis. Those who provide finance usually have less attachment than those who provide labor. Shareholders in a large firm can easily sell out for speculative gain whereas employees may serve for life; similarly, donors to a charity are usually less involved than the volunteers. There are also differences amongst workers; whilst some may be permanent full-time staff, others may be casual part-time staff. In a high-trust society commitment from workers and volunteers may be readily forthcoming, but in a low-trust society people will prefer low-commitment involvement instead. People may prefer to give money rather than time to a charity and to take only casual work, while shareholders may be very concerned that their holdings are liquid.

Some groups have formal structures: these are typically large and long-lived groups. Formal structures institutionalize the division of labor, creating posts or offices to which people are appointed; some posts may be filled on a rotating basis, often by election. Other groups are informal. For example, a market consists of all the people who turn up in the market place to trade – whether the market is a physical location, a commercial publication, or a web-site. Although access to the market may be free, traders must abide by the rules for enforcing contracts. A network is even more informal – it is simply a group of people who are in regular contact with each other [Putnam (1993)]. Networks are typically governed by customs which are enforced through reputation effects. A low-trust culture requires formal rules and procedures, whereas a high-trust culture is more versatile: both formal and informal systems can be used. Networks are useful for sharing information, particularly between entrepreneurs. In a high-tension culture networks can foster innovation, but in a low-tension culture they may simply foster collusion instead.

### **3. Towards an economic theory of culture**

Up to this point, the discussion has simply taken existing insights from sociology and social anthropology and reformulated them in economic terms. Further development of an economic approach to culture requires specific analysis of competition between cultures, leading to an explanation of the competitive strategies employed by social groups. This section outlines a set of assumptions on which a formal model of cultural competition can be developed.

### 3.1. Leadership

Leadership is the most important role within a group. The leader typically manages the external relations of a group. ‘Take me to your leader’ say outsiders who need to negotiate a commitment from a group. The leader demands loyalty from the members in order to guarantee the delivery of commitments and to maintain the reputation of the group. The leader has the power to discipline or expel disloyal people.

The logic of leadership is very simple. In a highly complex and uncertain world, people cannot resolve every issue for themselves. In particular, fundamental questions about the future of the world and the destiny of the individual cannot be easily answered; the costs of collecting and processing all the relevant information would be prohibitively high. Specialist leaders such as priests and politicians are required. Even then their answers cannot be definitive. Different leaders give different answers to the same question, based on different information, and so different cultures prevail. Leaders also provide answers to more specific questions; thus the leader of a firm decides what type of product is most in demand, and the leader of a charity decides what kind of people are most in need of help. The leader is the person deemed to have a comparative advantage in processing the relevant information. He may also claim to have privileged access to information, perhaps through external contacts. Alternatively, he may claim to be able to interpret information in a better way [Casson (2000)].

Leadership styles vary. Some charismatic leaders seek publicity, whereas others are self-effacing. Some leaders even seek to disguise their identity – such as an agitator leading a demonstration or the ‘brain’ at the center of a spy-ring. The common notion that groups can achieve ‘spontaneous order’ without a leader is a myth. It is simply a consequence of failing to identify where leadership really lies.

Leadership requires very scarce talents and as a result many leaders lack appropriate qualifications for the job. Successful leaders must justify the trust that their followers place in them. A leader who has lost the trust of his or her followers is of little value to the group, since members no longer feel secure in following their orders or advice. An alternative leader may emerge ‘from the ranks’ of ordinary members and constitute a rival source of authority – the militant British shop-steward, for example. The rival leader may organize a revolution to depose the incumbent if the incumbent cannot appoint a successor first.

### 3.2. Competition between groups

In a free society people can choose which leaders they follow. At any given time rival leaders will disagree about fundamental issues and people will have to decide with whom they agree. In particular, different political parties promote different ideologies, based on different theories of the economy and different views of human nature. In principle, only one of the rival leaders can be right. Indeed, the most likely scenario is that none of the leaders is right, since each is promoting an over-simplified and somewhat distorted view of the situation. Disagreements may persist because it is impossible to

find any decisive evidence for or against a particular view. In practice most leaders do not debate upon an abstract level, but rather in terms of strategy and policy. They promote specific projects which embody the values they promote and which, it is claimed, will work because the theory on which they are based is sound. For example, a political leader may promote a project to create a Welfare State, based on the optimistic view that new technology makes 'welfare for all' an affordable proposition. A business leader may motivate their workforce by claiming that their product is the best in the world, and a great benefit to all who consume it.

An articulate leader offers their followers a vision of what the project can achieve. The leader's rhetorical skill in creating 'sound bites' and 'buzz words' may be supported symbolically – perhaps by a launch at a prestigious location. The vision typically ignores the short run constraints under which the project operates, and emphasizes its long run potential instead. A vision will often be deliberately vague. It may be expressed in an artistic form which conveys an overall impression without revealing much key detail. The rationale for this ambiguity lies in the fact that much can change before the project achieves its goal, so that it would be misleading to be too specific about the final outcome. Indeed, the more ambitious the project, the longer it is likely to take to complete and so the vaguer the final outcome will be at the initial stage.

Competition may also induce leaders to scorn their rival's visions – arguing that they represent unworkable delusions. In Western democracies, debate between party leaders sometimes degenerates into mutual scorn. The emergence of negative stereotypes, promoted by leaders who wish to discourage their members from defecting to rival groups, can be explained in similar terms. This negative strategy has its limitations, however – too much emphasis on another leader's faults may suggest to honest followers that a leader is simply distracting attention from their own defects instead.

A key feature of a vision is that it arouses an emotional response in the follower. Such emotions are often described as 'beauty' (in the discovery of a simple theory, for example), 'glory' (as in winning a great team victory) or 'awe' (as in creating a monumental piece of architecture or engineering). The follower is enthused by contemplating the vision; by assessing their own emotional response to the vision, the follower can assess the magnitude of the emotional rewards that they will obtain through participation in the project.

Participation in each project involves a contract – usually an implicit contract assured through trust, but sometimes also a formal contract too, which is backed by law. There is an important psychological dimension to this contract. The leader emphasizes that the reward obtained by contemplating the vision will be strongest for those who make the greatest effort. Each follower will know how much effort they have committed to the project; the greater the sacrifices they have made, the greater the rewards they will obtain. These rewards come from two main sources. The first is the satisfaction from being absorbed in a worthwhile project to the point where the worker is unaware of their surroundings or of the passage of time. The second is a sense of pride and contentment when they rest from their work and reflect not only on what they have already achieved but also on what will be achieved when the project is complete. Followers who



know they have made little effort will experience little reward, whilst those who have deliberately shirked will experience guilt and wish that they had never joined.

An effective leader will show appreciation of followers' efforts. But the leader cannot always monitor individual effort with great accuracy. In certain types of work this agency problem may be overcome by basing rewards on measured output. But output may be only weakly correlated with individual effort, particularly in large teams. The 'psychological contract' is particularly valuable, therefore, in motivating effort in teams. However, team-work is not just a matter of effort. Loyalty is important in any project, and particularly so in teams, where the loss of a member can be very disruptive. Every new member has to learn their role, and the cost of training usually falls on the leader. Loyalty is thus an important element in the psychological contract. The stronger a person's emotional attachment to the project at the outset, the greater their sense of guilt when quitting.

When an individual is deciding whether to follow a particular leader, therefore, they will need to know both how they are likely to respond to the vision, and how they will actually perform. They therefore need to know their own competencies and their own emotional characteristics. If these characteristics are incorrectly assessed then a mis-match will occur between the individual and the project, and thus between the individual and the group. This will in turn lead to a waste of resources, in both material and emotional terms. It is typically assumed in economics that individuals possess full information on their own personal characteristics. In practice, however, it can be argued that they do not. In neoclassical economic theory, asymmetric information is usually construed as meaning that an individual knows their own characteristics, but others do not. It is possible, however, to construe the concept differently, and to suppose that other people know a person's characteristics better than they do themselves. Focusing on emotional characteristics highlights this point. Most parents have a better understanding of their children's emotions than the children do themselves. Many people remain 'child-like' (or even 'childish') in their emotions when grown up, and so not only family but also friends may be better aware of a person's emotional characteristics than the person themselves. Indeed, using biological evidence Frank (1985) has argued that people signal their own emotions to others unselfconsciously through facial expression and posture, and that their inability to control these emotional signals gives a credibility to their statements that they would otherwise lack. In a similar vein Freudian psychoanalysts have argued that people sublimate their emotions in order to disguise their feelings from themselves. People not merely lack self-knowledge and self-awareness – they are also systematically deny the existence of certain emotions too.

It is unnecessary to accept all of these claims in order to agree that many followers may be unaware of their emotional characteristics at the time they take a decision to join a group. Joining a group is therefore not only risky because of uncertainty about the leader and about the behavior of other members of the group, but because of uncertainty about one's own characteristics too. People's uncertainties about their own characteristics provide a significant opportunity for plausible leaders who are a good judge of character. The leader can invite people who in their judgment have the cor-

rect characteristics to join their group. People who feel very uncertain about their own characteristics are likely to respond in a positive fashion to such an invitation. Trusting people are also likely to respond, as they are more likely to accept the leader's judgment. An honest leader pursuing a socially worthwhile project can turn such mechanisms to good advantage, but it is equally obvious that an unscrupulous leader can take advantage of vulnerable followers. The most vulnerable people are those who are unaware that their own uncertainties and trusting nature are very obvious to others. Those whose competencies are obviously limited are particularly vulnerable, because it is obvious that they will receive few offers from other leaders. They may however receive some offers from honest but highly altruistic leaders, who wish to save them from falling under the influence of unscrupulous leaders instead.

### 3.3. *The changing nature of competition between leaders*

The nature of competition between leaders has been changed fundamentally by the growth of the communications and media industries – from the growth of print journalism in the eighteenth century to the spread of cinema, radio and television in the twentieth. The lower cost of mass communication has intensified competition between the leaders of high-level groups, especially political parties. Most significantly, the technologies of photography, film and video have reduced the cost of pictures relative to words, giving pictorial images an increasing role in propaganda and persuasion. Images liberate arguments from the requirement of a literate readership. They make use of a natural visual language which transcends any specific written language, and therefore reaches a mass multi-lingual audience.<sup>5</sup>

Certain images elicit strong emotional reactions. These reactions are almost instantaneous and are therefore invaluable to leaders in gaining attention for their messages. Indeed these reactions are so strong that the image itself may become the argument. Pictures of starving children or police brutality, for example, make their own political points without any need for verbal interpretation. Competition between leaders for visual attention encourages the pursuit of the outrageous. In any collection of competing images the most outrageous is likely to win. People may be attracted by beauty but surprise and horror have an even greater fascination.

The abstract nature of competition between ideologies does not lend itself readily to visual expression. The loss of media space to more visual subjects may be one reason why vigorous political debate appears to have declined as consumption of media services has increased. Social projects are easier to promote, as visions of better houses, schools and hospitals are easy to project. This encourages politicians to argue less about ideology and more about specific projects – a strategy recently adopted by New Labour in the UK [Protherough and Pick (2002)].

Consumer products are remarkably easy to promote by picturing the consumer as relaxed and self-assured; this works particularly well for simple products which provide

<sup>5</sup> The links between culture and language are explored further from an economic perspective in Jones (2000).

emotional benefits of a social nature – cosmetics and alcoholic beverages, for example. The multi-lingual nature of a visual proposition benefits multinational consumer brands. Faces attract attention – particularly faces that are instantly recognized. This favors the promotion of ideas through celebrity endorsement. Since sportsmen and entertainers are not generally noted for their political wisdom, celebrity endorsement works best in product promotion, although it has been used with some success in politics too.

Commercial advertisers are unlikely to increase their sales if consumers give money to good causes instead of spending it on themselves. The implicit message of a typical product advertisement is therefore that low-trust is the norm. Similarly, many products are advertised as impulse purchases, which allow the consumer to show off in a social setting. This promotes a low-tension spontaneous lifestyle as the norm, rather than a single-minded high-tension lifestyle which would produce better long-term results.

The optimization of visual image for persuasive purposes requires very scarce skills. Creative workers in advertising and public relations can command substantial economic rents. The financial requirements of major promotional campaigns constitute a significant barrier to entry for many types of leader. A highly visual political campaign may require powerful industry backers who expect rewards if their candidate is elected to office. Thus leadership becomes more like commercial entrepreneurship as the economic requirements converge on the funding of media campaigns.

In most modern societies newspapers, magazines, radio and television rely heavily on advertising revenues rather than sales and subscriptions. They have a strong financial incentive to attract an audience that is susceptible to advertisers' messages. This can induce the 'dumbing down' of content in order to attract the people most likely to be influenced by the visual message that the advertiser plans to use. Some messages are easier to dumb down than others – for example, a blatant appeal to short-term self-interest is easier to communicate than a sophisticated appeal to long-term social concerns.

To summarize, there are many reasons why in a modern society characterized by competitive individualism the role of moral leadership is difficult to carry out. Whilst the power of visual imagery favors the promotion of certain types of charitable project such as child poverty or animal welfare, it discriminates against the promotion of high-trust high-tension political values. Competition for attention in the visual media is on average biased against the promotion of high-trust cultural values.

## **4. Key dimensions of culture**

### *4.1. Four main dimensions of cultural variation*

There are many fundamental issues which cultures must address. Some are very general, such as 'What are people really like?', whilst others are more specific such as 'Whom can you trust?' and 'How do you motivate people?' Other issues include 'What forms of organization are natural?' and 'How far can technological progress advance?' Describing a culture in full can therefore be a very complex task.

A parsimonious theory of culture must identify just a small number of dimensions along which cultures vary. By focusing on those aspects of culture which are likely to influence economic performance, four main dimensions of culture can be derived. These dimensions were introduced at the outset, and are summarized in the first two columns of Table 2. The first column of the table identifies the end of the dimension which is found in a typical Western competitive individualistic society, whilst the second column indicates the dimension which corresponds to ‘Utopian solidarity’ – the kind of culture that would be found in an idyllic closed society of the kind visualized by Rousseau. This four-way classification is a refinement of a classification proposed in Casson (1993).

- (i) *Individualism versus collectivism*: An *individualist* believes that people are autonomous. Everyone is different and each person values personal ‘lifestyle’ projects above others [Earl (1986)]. The information required for coordination is widely distributed – shocks are individual-specific. Ownership and control of resources should be vested in individuals, since only individuals have the information required to take decisions that affect themselves. A *collectivist* believes that we are all part of the community into which we were born. Even as adults we remain dependent on others for our survival. A collectivist also believes in uniformity – everyone is the same, and everyone values large awesome projects. Information required for coordination is centralized – shocks have collective impact. Collectivists believe that ownership and control of resources should be vested in the group [Ekelund and Tollison (1997)].
- (ii) *Pragmatism versus proceduralism*: *Pragmatists* believe that intuitive judgments based on wide personal experience hold the key to successful decisions. Hunches can also be tested through informal conversation with other people. The best de-

Table 2  
Four dimensions of culture

Limit of dimension corresponding to competitive individualism	Limit of dimension corresponding to Utopian solidarity	Corresponding dimension in Hofstede	Optimal combination
Individualism	Collectivism	Individualism–collectivism	Voluntarism
Pragmatism	Proceduralism	Low–high uncertainty avoidance	Good judgment
Low-trust	High-trust		Warranted trust
High-tension	Low-tension	Femininity–masculinity	Warranted self-confidence

Note: Only three of the four dimensions identified by Hofstede appear in the table. The missing power–distance dimension in the Hofstede classification may be loosely construed as a hybrid which combines elements of individualism–collectivism with elements of low-trust–high-trust.

cisions are made promptly. A single individual should be ultimately responsible for each decision. *Proceduralists* believe that good decisions are generated by closely following formal procedures, whose design is underpinned by theory, and which involve the systematic collection of objective information. The use of committees may delay decisions, but it is better to ‘get it right’ than to do it quickly.

- (iii) *Low-trust versus high-trust: High-trust* individuals believe that others will be honest, work hard, be loyal, and generally keep their promises even when they have little material incentive to do so. *Low-trust* individuals believe that others are guided by material incentives, and will therefore often lie, cheat or shirk. High-trust is particularly important in an individualistic society, because individuals do not have the same power of enforcement as a collective body [Holmes and Sunstein (1999)].
- (iv) *High-tension versus low-tension: A high-tension* person is attracted to ambitious projects, while low-tension person prefers easy projects. The high-tension person is stressed because they are aiming high, and will be ashamed of failure.<sup>6</sup> Conversely, a *low-tension* person is relaxed, because they are aiming low, and they will blame any failure on factors outside their control. Low-tension people like to behave in a spontaneous manner, which often has anti-social consequences [Casson (2002)], although it is a manner of which some economists approve [Scitovsky (1976)].

There are many other classifications of culture which have been devised for a variety of purposes, but there is one particular classification, due to Hofstede, which has been particularly influential in management and organizational studies and is particularly relevant to performance issues [Hofstede (1980); Graham (2001)]. Hofstede’s classification was arrived at empirically by applying factor analysis to a large-scale cross-national study of the employees of a multinational firm. Unlike the classification used here, Hofstede did not deduce his classification from first principles, but nevertheless a comparison is useful. It is interesting that he also focused on four dimensions, some (though not all) of which correspond to the theoretical classification as noted in the third column of Table 2.

Taking the two limits of each of the four key dimensions described above identifies 16 ideal types of culture which are presented in Table 3. Some of these are particularly interesting, especially the high-trust analogues of competitive individualism. These embody the principle of voluntary association for the purpose of pursuing ambitious projects but add the notion that the aims of the project may be altruistic, that competition between the projects is orderly rather than aggressive, and that coordination of projects relies heavy on trust between members of a team. It is known as associationism.

To keep the theory really simple it would be nice to identify just one of these 16 cultures as the best from a performance point of view. It would then be possible to compare

<sup>6</sup> For an excellent discussion of high-tension in the context of fundamentalist religious sects see Stark and Bainbridge (1987).

Table 3  
Typology of cultures

	HG High-tension pragmatic (judgmental)	HD High-tension procedural (administrative)	LG Low-tension pragmatic (spontaneous)	LD Low-tension procedural (bureaucratic)
IS Individualistic low-trust (competitive individualism)	<i>Enterprise culture:</i> Aggressive competition between highly entrepreneurial selfish people.	<i>Big business culture:</i> Aggressive competition between selfish, ambitious but unimaginative people controlling formal organizations.	<i>Libertarianism:</i> Social anarchy constrained only by legal enforcement of market contracts.	<i>Play-the-system culture:</i> Unprincipled competition between formal organizations regulated unsuccessfully by weak and corrupt bureaucracy.
IH Individualistic high-trust (associationism)	<i>Entrepreneurial associationism:</i> Orderly markets allocate resources between ambitious altruistic projects.	<i>Administrative associationism:</i> Orderly competition between ambitious altruistic people running professional organizations.	<i>Good neighbor culture:</i> Social ambitions are limited to relief of current problems such as poverty. Individuals act on impulse to help the needy who are known to them.	<i>Charity culture:</i> Compassionate leaders set up formal organizations to help the needy, and recruit volunteers.
CS Collectivistic low-trust (coercive collectivism)	<i>Revolutionary state:</i> Totalitarian dictator personally promotes prestige projects in which people are forced to participate.	<i>Soviet-style planning:</i> Professional government planners implement ambitious projects using conscripted workers.	<i>Arbitrary dictatorship:</i> Dictator with ambition simply to survive in power improvises strategies to defeat rival bids for power.	<i>Conformist culture:</i> Coercive bureaucracy resists change and demands conformity from apathetic people.
CT Collectivistic high-trust (paternalism)	<i>Charismatic leadership:</i> Paternalistic leader with Utopian vision enthruses population.	<i>Welfare state:</i> Ambitious altruistic programs are devised by a paternalistic leader and administered using public service ethic.	<i>Familism:</i> Paternalistic leader presides over low-productivity economy where socialization is more important than work.	<i>Utopian solitariness:</i> Low-productivity economy is coordinated through compulsory participation in traditional rituals presided over by leader.

the actual culture of any social group with the ideal culture, and measure how many dimensions were in agreement: the closer the actual culture to the ideal culture, the better the economy would perform. Given the advantages of a high-trust culture in reducing agency costs and transaction costs, some form of associationism would be a natural

choice. The form that is closest to classic Western individualism is entrepreneurial associationism, and so this appears to be the natural choice as the ideal.

#### 4.2. Trade-offs involved in a high-performance culture

There are three difficulties associated with identifying entrepreneurial associationism as the unique high-performance culture, however. The first is that a combination of four extreme values is rarely an optimal choice in any problem. There are strong grounds for believing that along each of the four dimensions there is scope for a trade-off.<sup>7</sup> Typical results of the trade-offs are listed in the right-hand column of Table 2. They may be summarized as follows, taking each dimension in turn:

- (i) *Voluntarism*: Individuals are encouraged to transfer their resources to institutions on a voluntary basis. They are encouraged to identify opportunities for projects which these institutions can carry out. Individuals like group projects, but prefer to choose the type of project with which they are involved
- (ii) *Good judgment*: Procedures work well in dealing with frequent minor shocks of a transitory nature. Improvisation is required in dealing with intermittent major shocks of a persistent nature. Successful improvisation requires good judgment, which is based on wide experience.
- (iii) *Selective warranted trust*: Whilst trust reduces coordination costs, naïve trust is of little value, since naïve people provide easy pickings for cheats. A high-trust equilibrium is what counts, in which the majority of people (who are trustworthy) can identify each other and transact with each other, whilst the minority of people (who are untrustworthy) cannot transact at all. Trust is engineered through moral leadership. Leaders demand loyalty and hard work from those who join their teams.
- (iv) *Warranted self-confidence*: High tension delivers results in task-oriented projects. But high-tension cannot be sustained indefinitely. A high-tension person relaxes in a secure environment where they reflect on their performance and learn from their mistakes. The low-tension person likes to mess around at work, and have lots of fun when relaxing.

A combination of voluntarism, good judgment, selective warranted trust and warranted self-confidence may be termed refined associationism, and may be taken as the most accurate characterization of optimal culture from a performance point of view.

The second difficulty with this choice is that none of the forms of associationism discussed above correspond to the cultures of the most successful Western economies. These tend to be much lower-trust than associationism would imply. It could therefore be argued that the entire theory is a predictive failure. This leads on to the third point, however, which is that the exact position of the trade-off will reflect the local

<sup>7</sup> The importance of trade-offs in culture is recognized by many writers on culture; see, e.g., Hampden-Turner and Trompenaars (1997).

circumstances with which a culture has to contend. Thus a very large, transient and widely-dispersed group may have to reconcile itself to lower levels of trust than a small, stable and compact group. It is therefore unrealistic to expect every group to conform to the same ideal. In another case, one group may have an outstanding moral leader – a ‘man of the moment’, say – who intervenes at a critical moment when change is required, whereas another group may have to cope without such a leader. Drawing upon a larger number of less able and less trustworthy individuals to do the same job, they may institute a division of powers between the leaders and even endeavor to promote a degree of competition between them.

It is in fact possible to explain the current predominance of competitive individualism in successful Western countries such as the US in terms of adaptation to changing global conditions in the period since World War II. In the post-war period, volatility has increased as a result of accelerated technological change and the globalization of trade, driven by lower transport costs and tariffs. An increase in the volatility favors a switch from collectivism to individualism, and from proceduralism to pragmatism, because of the need for greater flexibility.<sup>8</sup> Globalization has also reduced trust between trading partners, as local networks of trade have been disrupted by the emergence of foreign competition; social trust has been eroded too, as migration has disrupted the customs of local communities. The globalization of communications has encouraged a switch from low-tension to high-tension culture as people in low-productivity economies have become aware of the opportunities presented by innovation and export-led growth. Countries across the world have therefore switched towards a specific type of competitive individualism, namely an individualistic, pragmatic, low-trust high-tension ‘enterprise culture’, as indicated in the top left-hand cell of [Table 3](#).

The economic theory of culture therefore predicts that culture will adapt to the environment, both across space and over time. This accords with basic economic intuition that despite all the qualifications noted above, a successful culture must correspond closely to the realities of a situation facing a group. As circumstances change, so the optimal culture changes too and forces of adaptation, driven by competition between rival leaders, come into play.

#### *4.3. Refining the dimensions of culture*

Sociological writers on culture have between them identified over a hundred different dimensions of culture. Furthermore, cultural analysis of cross-country differences in industrial policy has identified other dimensions besides those mentioned above [[Foreman-Peck and Federico \(1999\)](#)]. Almost all of these additional dimensions can however be subsumed under the four key dimensions; indeed, these key dimensions were developed in part as composite dimensions under which various other dimensions could be subsumed. [Table 4](#) lists 22 dimensions of culture, including many of the most

<sup>8</sup> For earlier examples of such switching see [Hirschman \(1982\)](#).



Table 4  
Sub-dimensions of culture

Characteristic favoring competitive individualism	Characteristic favoring Utopian solidarity
<i>Individualistic (I)</i>	<i>Collectivistic (C)</i>
Atomistic	Organic
Dynamic	Static
Incremental	Radical
Democratic	Elitist
Market-based	Planning-based
Efficiency-oriented	Equity-oriented
Consumer-oriented	Producer-oriented
<i>Pragmatic (G)</i>	<i>Procedural (D)</i>
Empirical	Theoretical
Outcome-based	Process-based
Risk-taking	Risk-averse
Artistic	Scientific
Personal	Impersonal
<i>Low-trust (S)</i>	<i>High-trust (T)</i>
Unprincipled (moral skepticism)	Principled (morally committed)
Secular	Religious
Selfish	Altruistic
Autocratic	Consultative
Aggressive	Orderly
<i>High-tension (H)</i>	<i>Low-tension (L)</i>
Aspirational	Complacent
Deliberative	Spontaneous
Optimistic	Pessimistic
Confident	Unsure
Progressive	Conservative

frequently cited dimensions, and attributes each of them to one of the four key categories.

Where issues relating to political constitutions and national economic policy are concerned, the sub-dimensions associated with the first dimension – individualism versus collectivism – are most important. Where issues of organisational structure and management style are concerned, the sub-dimensions associated with pragmatism versus proceduralism are most important. The quality of personal relationships within organizations, the intensity of competition between organizations, and the general quality of social life are governed by the sub-divisions of the third dimension – the degree of trust. The extent to which people are energized and inspired by visions of better life, either for themselves or others, is governed by the sub-divisions of the fourth dimen-

sion – the degree of tension. Since there is insufficient space to examine each of these sub-dimensions in detail, their principal features are summarized in Tables 5–8. The middle columns of these tables explain the nature of the variation along the dimension concerned, whilst the right-hand column considers the point along the spectrum on which the best results are likely to be obtained. This is described as the ‘high performance mix’. It is important when studying culture to remember that the dimensions of cultural variation do not normally run from ‘good’ to ‘bad’ or *vice versa* but, like ordinary economic variables, express a trade-off in which the optimum is usually at an interior point. Just as in conventional economics, extremes are rarely efficient in cultural life.

An optimum in this context represents a cultural mix which is likely to prove efficient in the long run. In modern parlance, it is ‘sustainable’. However, the optimum along any one dimension cannot be determined without reference to the other dimensions of culture. Thus a culture that promotes a distrustful attitude to other people may have an optimal degree of competitiveness which is quite high, whilst a culture which encourages people to trust each other may have an optimal degree of competitiveness which is much lower. In the long run there will be a tendency for competition between cultures to select the culture that is most efficient in overall terms. The characteristics of an optimal culture are summarized in general terms in the next section. However, as noted above, the optimal culture varies according to environmental constraints, and this means that the optimum is difficult to specify in terms of all the 22 dimensions discussed in Tables 5–8. Furthermore, the process of competition between cultures is so slow and disjointed that for the foreseeable future the detailed predictions of the theory merely identify a sub-set of viable cultures which are likely to remain in competition with each other for a considerable time to come.

Some of the dimensions described in Tables 5–8 are much more relevant at one level of leadership than another. Individualism versus collectivism and the sub-dimensions associated with it are particularly important for high-level leaders of large groups such as the nation state. They influence their attitude to the decentralization of power. A high-level leader must decide how far his or her followers should be allowed to form lower-level groups on their own initiative. Should the emergence of lower leaders be encouraged, as a welcome display of initiative, or discouraged as a potential threat to the leader’s power? Other dimensions apply at every level. The issue of trust, for example, is fundamental at every level. A high-level leader who does not trust lower-level leaders will either discourage the formation of low-level groups or will promote aggressive competition between them, whereas a trusting leader may encourage low-level groups and promote co-operation and orderly competition between them [Knight (1935)]. At the same time leaders of lower level groups must decide whether to monitor their members and offer material rewards for good behavior, or whether to trust the members to monitor themselves and to reward themselves emotionally for good behavior.

Table 5  
Detailed analysis of individualism *versus* collectivism

Characteristic:	Commentary	High-performance mix
Individualist/ collectivist		
Atomistic/ organic	An atomist believes that individuals are autonomous and independent of society. Their personal rewards derive from their own activities and their attitude to others is purely instrumental. Atomists play down emotions as a source of utility and emphasize pleasure from material consumption instead. Organists believe that the most important rewards are emotional and derive from participation in social activity. Activities devoted to improving and strengthening society generate especially large rewards. The more sacrificial effort people put in, the greater the emotional rewards they get out.	Atomism is bad psychology since it underestimates the importance of emotional rewards, particularly those derive from harmonious social interaction. The atomist is correct, however, that ultimately it is individuals that take decisions. A high-performance culture recognizes that economic performance depends on the interaction of numerous individual decisions – decisions taken by people with real concerns about the society in which they live.
Dynamic/ static	A dynamic culture regards the environment as highly volatile. Change is endemic and it is necessary to adapt and evolve in order to survive. Change is exciting and people can thrive on it. A static culture believes that the environment is stable. Change can be neutralized in order to preserve the status quo. Homeostasis provides much-needed security.	The environment is volatile. Major changes usually require adaptation but minor changes can sometimes be neutralized by an appropriate respond. People can only stand so much excitement from change.
Incremental/ radical	An incrementalist believes that changes are typically small and localized. They relate to particular products or places. The people close to the changes are in the best position to respond. A decentralized system that empowers individual decision-making produces the most effective responses. A radical believes that changes affect the entire economy. Radical actions are required to take advantage of new opportunities or respond to emergent threats. This requires a centralization of power.	Volatility in the environment takes different forms. Minor changes occur all the time, whilst major changes occur only intermittently. Minor changes can easily be delegated to individuals to handle; indeed, standard procedures can be developed to deal with the most common types of change. Major changes can take many different forms and require a more consultative and collective response. Leaders have an important role in building consensus where radical change is required.
Democratic/ elitist	A democrat believes that everyone has unique life experiences which make them worth consulting on how to respond to ma-	Leaders are specialists in taking complex decisions. Leaders constitute an elite – but they should be an ‘open elite’ which anyone can

(continued on next page)

Table 5  
(continued)

Characteristic: Individualist/ collectivist	Commentary	High-performance mix
	<p>major changes. So far as minor changes are concerned, they can be left to handle them themselves.</p> <p>An elitist believes that only a select group of people of high intelligence or 'good breeding', etc. have the ability to form correct opinions and to carry out the appropriate calculations.</p>	<p>attempt to join. Leaders should consult their followers, but ultimately they must act on their own judgment. Ineffective leaders should be replaced – followers should be able to replace a bad leader, or quit a badly-performing group. Leadership roles require people of exceptional ability, but this ability is difficult to identify in advance.</p>
Market-based/ planning-based	<p>The atomist recognizes that markets provide the flexibility that allows different people to respond in different ways to similar events. Market-making middlemen adjust prices to match long-run supply and demand; they also hold inventories to buffer short-run fluctuations.</p> <p>From an organic perspective, planning is the most direct means of achieving consistency between individual responses since it uses a single directing mind. A planner may administer prices or ration quantities.</p>	<p>Planning and markets need to be combined. Firms are planning units which coordinate tightly-coupled systems. Households also plan, but on a smaller scale. Markets link these different planning systems in a loosely-coupled way. Factor markets price the labor and capital employed by firms. Firms which attempt to plan activities which are better coordinated by a market will fail to break even. By allocating scarce factor supplies to the most viable firms, the factor markets determine which activities are planned and which are not.</p>
Efficiency-based/ status-based	<p>The atomist exploits market competition to eliminate waste. An inefficient producer cannot match the price of an efficient producer, and so customer switching eliminates wasteful production methods. Consumers who value products most out-bid those who value them least, so outputs are not wasted by consumers who do not value them. The organicist notes that a consumer's ability to pay depends on income. Consumption should reflect basic needs and social status. Since basic needs are similar, necessities should be allocated fairly. Luxuries should reward service to society as a whole and not just wealth derived from scarce factors of production.</p>	<p>People care both about their own consumption and about the kind of society in which they live. Market-based incentives to eliminate wealth can make everyone better off, but only if those who make the savings are prepared to share them with others. If they are forced to share them, then the incentive to make the effort to drive out waste is reduced. An ethic of community solidarity, which provides emotional rewards to those who reduce waste for the benefit of others is the best solution. Thus a market system can usefully be supplemented by a 'honors system', provided that honors are awarded for sacrificial effort and not simply sold to the highest bidder.</p>

(continued on next page)

Table 5  
(continued)

Characteristic: Individualist/collectivist	Commentary	High-performance mix
Consumer-oriented/producer-oriented	<p>The atomist believes that people derive rewards mainly from material consumption. Novelty and fashion, packaging and presentation, are not trivial matters, but sources of serious satisfaction. The proliferation of different product varieties made possible by technology and trade is to be welcomed. So too are the efficiency gains generated by specialization, even though work becomes monotonous. Services are also valuable, even though no tangible artifact is produced.</p> <p>Organicists believe that people derive rewards mainly from producing goods. They value product variation only when it arises from the use of local materials, and from the personal style of the worker. They value tangible product over intangible services, and craft work over mass production. Producer motivation is strengthened by a long-term relationship with the customer which allows the producer to witness the product in use.</p>	<p>Consumer culture promotes the development of new technology. It exploits advances in technology and communication to significantly improve the material living standards of the poor.</p> <p>However, workers ‘alienated’ by mass production will produce poor quality, so ‘job enrichment’, which limits specialization, may actually improve overall efficiency. They may also seek enrichment through trade union activism.</p> <p>Not all workers may require job satisfaction, however. Satisfactions can also be obtained from hobbies and recreations. Boring jobs may indirectly enrich cultural life by encouraging people to seek satisfaction in community activity instead.</p>

Table 6  
Detailed analysis of pragmatism *versus* proceduralism

Characteristic: Pragmatic/proceduralist	Commentary	High-performance mix
Empirical/theoretical	<p>A pragmatist believes that the response to change should be based on evidence rather than theory – it should be improvised on the basis of previous experience. Everyone has unique life experiences which help to prepare them for taking decisions. Belief in the uniqueness of personal experience links pragmatism to atomism. A proceduralist believes that decisions should be explicitly ra-</p>	<p>Theory and experience need to be combined. Neither evidence without theory, nor theory without evidence, will produce good decisions on how to respond to change. In some situations there is no relevant theory, whereas in other cases there are multiple theories, and hence confusion. Theories invariably abstract from certain factors, and may therefore distort a decision if the omitted factor is im-</p>

(continued on next page)

Table 6  
(continued)

Characteristic: Pragmatic/ proceduralist	Commentary	High-performance mix
Outcome-based/ process-based	<p>tional, in the sense of being grounded in some theory. Without the correct theory, evidence cannot be properly interpreted. Decisions should be based on calculation rather than improvisation. Since the mastery of theory often requires intellectual ability, theoretical orientation is often linked to elitism.</p> <p>Proceduralists believe that a correct theory can suggest a rational procedure which will guarantee a correct decision. A group of people (e.g., a committee) may be involved in taking the decision. Pragmatists believe that procedures normally delay a decision, and make the outcome worse. Disagreements in committees can add to delays; it is better to make one person clearly responsible for a decision, and let them 'get on with it' right away.</p>	<p>portant. On the other hand, ignoring relevant theory can mean that the significance of key evidence is not appreciated.</p> <p>Rational procedures may be useful in dealing with transitory volatility – e.g., in recording reservations or managing inventory. But there are few cases where theory is good enough to identify an optimal procedure. Procedures can also be useful in encouraging autocratic individuals to consult with knowledgeable people. Otherwise it is individual experience that is crucial. Selecting the right individual is more important than optimizing the procedure they employ.</p>
Risk-taking/ risk-averse	<p>A proceduralist believes that risk can be reduced through rational decision-making processes, whereas a pragmatist denies this. The proceduralist worries that correct procedures have not been properly followed, whereas the pragmatist, having improvised their decision, simply sits back and waits for events to unfold.</p>	<p>Large intermittent shocks cannot easily be addressed by routine procedures, and so risk is inescapable. Frequent minor shocks can often be addressed by rational procedures which involve collecting and processing information before a decision is made. The collection of information allows risk to be managed, although it cannot be eliminated altogether. People who are responsible for dealing with large intermittent shocks must be willing to take substantial risks.</p>
Artistic/ scientific	<p>Science analyzes local situations in terms of timeless universal laws, whereas the artist often expresses surprise and wonder at a situation. The scientist typically values uniformity whereas the artist values diversity. A scientific approach supports the development of a theory and the collection of evidence in a systematic way. It therefore underpins a procedural approach. Art tends to emphasize an emotional or even mystical response to a situation which is not fully un-</p>	<p>Economic theory has employed social scientific principles, such as the division of labor, specialization according to comparative advantage and global competition, with considerable success. Decision-makers who do not understand these principles are at a major disadvantage in business life.</p> <p>Economics has proved much less successful, however, in analyzing the emotional rewards that people derive from work and social activity. A combination of scientific un-</p>

(continued on next page)

Table 6  
(continued)

Characteristic: Commentary	High-performance mix
<p>Pragmatic/proceduralist</p> <p>derstood. It focuses on situations which are difficult, or impossible, to understand in purely scientific terms. It therefore supports a pragmatic approach to decision-making.</p>	<p>derstanding of the laws of markets on the one hand, and an artistic appreciation of emotional factors on the other, is therefore the appropriate combination for successful decision-making.</p>
<p>Personal/impersonal</p> <p>Pragmatists believe that people know a great deal more than they realize, and so it pays to converse with them rather than wait for them to tell what they know. People can also say more than they can write, because tone and gesture can aid expression. Pragmatists try out their ideas in conversation with other people, provoking others into revealing what they think. This helps them to arrive at a decision quickly. Proceduralists believe that written communication is superior to the spoken word because it is more precise. There is less scope for ambiguity and reason is unlikely to be clouded by emotion. Proceduralists prefer to consult through memoranda, which they study carefully before arriving at their decision.</p>	<p>Complex arguments benefit from being set out formally, but simple powerful ideas can often be expressed most vividly in conversation. Highly original ideas are difficult to articulate in a formal way. Original solutions to problems are therefore more likely to be generated through personal interaction.</p>
<p>Unprincipled/principled</p> <p>Principled persons believe that they are under moral obligation to a higher authority. They are called to play a particular role in society. They can only achieve peace of mind by doing their duty. Their higher nature (conscience, or spirit) recognizes that they need to control their lower nature (body, or passions). Self-control can be exercised through positive emotions, e.g., enthusiasm for a cause, or negative emotions, such as guilt and shame. Principles need to be based on functionally useful moral values: honesty, loyalty, hard work, and so on. These support teamwork on projects and facilitate coordination between different teams. An unprincipled person believes in satiating their biological needs. The only source of authority is their body; their objective is pleasure rather than peace of mind.</p>	<p>People need to respect their bodily requirements for physical survival, but over-indulgence can damage health. People have emotional as well as material needs, and those who realize this will be happier than those who do not. A moral framework enhances emotional rewards derived from participation in socially beneficial projects. Traditional moral principles such as honesty, loyalty and hard work facilitate coordination in complex economies by reducing transaction costs, encouraging investment, and promoting hard work. An effective leader will therefore promote traditional moral principles, even if their ambitions are purely materialistic.</p>

(continued on next page)

Table 6  
(continued)

Characteristic: Commentary	High-performance mix
Pragmatic/ proceduralist	
Secular/ religious	<p>The secular moralist expects to derive emotional benefits as part of an enhanced quality of life, whereas the religious person expects a dividend in the after-life. Religious people are therefore motivated by deferred rather than immediate emotional rewards. Their moral conduct is therefore more robust to disappointments. On the other hand, their beliefs in the after-life can prove vulnerable to attack from skeptics.</p> <p>Rivalry between religious groups can promote distrust as well as trust. Religious commitment can make religious conflict very intense. On the other hand, religious commitment can also promote extreme forms of self-sacrifice and heroism, such as those involved in fighting in defense of a country. While both secular morality and religious belief can generate emotional satisfactions (for people of good conduct), religion adds a further dimension to motivation which secularism lacks.</p>

Table 7  
Detailed analysis of degree of trust

Characteristic: Commentary	High-performance mix
Low-trust/ high-trust	
Selfish/ altruistic	<p>Selfish people cannot empathize with others. Their concerns are focused on their own consumption, work and leisure. They may be concerned with status, but only in an instrumental way – as a means of gaining privileged access to resources. They are concerned with the state of society only in so far as it impacts on their own material interests. Altruistic people empathize with others – either personally, e.g., friends – or impersonally, e.g., concern for the poor. They can derive vicarious pleasure from other people's happiness, and share their suffering too. Degrees of altruism differ depending on the weight that people place on other people's interests.</p> <p>Altruism is important in channeling high-tension people into providing support for others. Self-interested ambition can stimulate high-tension but generates external diseconomies, and leads to under-provision of emotional support. It does nothing to address the income inequality generated by competition between self-interested people, or to support the losers from the competitive process and their dependents.</p>
Autocratic/ consultative	<p>When other people are selfish and cannot be trusted, their opinions will reflect where their own interests lie. Consultation creates a risk of distorting decisions through lobbying from vested interests. If you cannot believe</p> <p>Consultation is useful not only in improving a decision but in motivating people to implement a decision through participation in the decision process. Opinions received need to be critically examined, however. Where</p>

(continued on next page)



Table 7  
(continued)

Characteristic: Commentary	High-performance mix
Low-trust/ high-trust	
	<p>what other people say, there is no point in asking their opinion. If other people are honest and their preferences are aligned with those of the decision-maker, their opinions may be valuable since they are likely to have been thinking about similar issues for themselves. Hence consultation is worthwhile.</p>
Aggressive/ orderly	<p>vested interests are important, conflicting opinions from the different interests will reveal that a problem exists.</p> <p>The high-trust view is correct. An advanced society is highly complex and the 'law of the jungle', which usually rewards aggression, does not work well. Reprisals can lead to feuds which originate with a simple misunderstanding.</p> <p>Competition is not just about challenging monopoly but about stimulating and diffusing socially useful innovations. Competitors who sabotage each other's activities do not benefit society and so 'rules of the game' are required. Competition works best when rivals can be trusted to abide by the rules.</p> <p>While aggression may sometimes motivate innovation, other motivators such as public recognition are available too. Channeling aggression into competition may be a useful way of controlling a potentially disruptive biological urge, but it still needs to be moderated through self-control.</p>

Table 8  
Detailed analysis of the degree of tension

Characteristic: Commentary	High-performance mix
High-tension/ low-tension	
Aspirational/ complacent	<p>Aspirational people have high norms. These norms may correspond to ideals deduced from moral or theoretical principles. Alter-</p> <p>The high norms of the aspirational person are indispensable to a high-performance culture.</p>

(continued on next page)

Table 8  
(continued)

Characteristic: High-tension/ low-tension	Commentary	High-performance mix
Deliberative/ spontaneous	<p>natively, people with wide horizons may know that higher standards are being achieved elsewhere. They are dissatisfied with the status quo. They believe that it can and must be changed. Complacent people have low norms. They have narrow horizons due to a parochial outlook. They are satisfied with the status quo, and their chief ambition is to maintain it.</p>	<p>Deliberation prevents people with high norms from giving up too easily. Spontaneity undermines the value of aspirations, since the aspirations are merely fantasies.</p>
Optimistic/ pessimistic	<p>A deliberative person concentrates single-mindedly on achieving his objective. He remains focused on it until he has either achieved it or has irretrievably failed. Success is quietly satisfying but failure is mortifying. Spontaneous persons focus on whatever has caught their attention most recently. It is not necessary to finish one task before starting another. Success is a cause for celebration, however minor it may be. Failure is attributed to bad luck or blamed on others.</p> <p>An optimist believes that the environment is favorable for the successful completion of a project whereas a pessimist believes that it is unfavorable. An optimistic culture promotes general optimism through notions such as 'the time is right' and 'it's all up for grabs'. A pessimistic culture promotes the idea that if something was really a good idea then someone else would already have done it.</p>	<p>Deliberation prevents people with high norms from giving up too easily. Spontaneity undermines the value of aspirations, since the aspirations are merely fantasies.</p> <p>Optimism reduces perceived risks and thereby encourages investment and innovation. However, unwarranted optimism can lead to wasteful projects being undertaken. Where the private benefits of investment are less than its public benefits, optimism may induce investors to risk losses for the public good. If private and social benefits are aligned, realism is better than either optimism or pessimism, as it leads to better investment decisions.</p>
Confident/ unsure	<p>When an optimist is confronted by a group of pessimists they may decide that they must be wrong. They need self-confidence to believe that they can be right when everyone else is wrong. A confident culture sustains the idea that people in the group are always right, at least compared with people in other</p>	<p>Most leaders require self-confidence to take the initiative in setting up groups, and take the responsibility if things should go wrong. A combination of optimism and self-confidence is a hall-mark of an entrepreneurial culture.</p>

(continued on next page)

Table 8  
(continued)

Characteristic: High-tension/ low-tension	Commentary	High-performance mix
Progressive/ conservative	<p data-bbox="280 371 654 475">groups. It may be based on a notion of innate superiority. People who are unsure usually adapt their opinions to conform with the majority view.</p> <p data-bbox="280 498 654 811">Progressives regard change as largely benign. They believe it provides opportunities rather than threats, whereas conservatives take the opposite view. Progressives are continually raising their norms in line with new possibilities whereas conservatives are more concerning with ensuring that existing norms are maintained. Being progressive involves innovation rather than conservation. Both are demanding, but innovation tends to be more demanding because the element of novelty increases the risks.</p>	<p data-bbox="678 498 1048 651">A high-performance culture requires a combination of science-driven innovation with the maintenance of functionally useful traditional morals. It therefore requires both a progressive technical agenda and a conservative moral agenda.</p>

## 5. Method and history

### 5.1. Methodological issues in modeling culture

This third and final part of the chapter attempts to draw together the threads of the preceding discussion. It begins by summarizing the principal differences between conventional neoclassical economics and the economic theory of culture outlined above. Five main differences have been identified. Contrary to conventional neoclassical economics, the economic theory of culture put forward above asserts that:

- Information is costly, both to collect and communicate. Where fundamental issues are concerned it is often impossible to collect objective evidence that will discriminate between alternative theories: thus different systems of beliefs can co-exist almost indefinitely. Conflicts between rival value systems are even more difficult to resolve; their authority often derives from tradition or from spiritual experiences whose authenticity it is impossible to assess. Information costs help to explain uncertainty – uncertainty exists because it is prohibitively costly to collect all the relevant information before taking a decision. Many uncertainties are radical and existential, because fundamental issues are peculiarly difficult to resolve. It is not just ‘facts’ that are uncertain – theories are uncertain too.

- The economic environment is volatile. Factual information is therefore continually obsolescing. A steady flow of new information is required to permit the economy to adapt appropriately to changing circumstances. Information sources are localized, so different people have access to different information. Furthermore, since different people use different theories to interpret this information, different people will react to similar events in very different ways. An important advantage of decentralization is that it empowers people to act immediately on their judgment of a situation. Where opinions differ about the advisability of change, competition permits the optimists to bid resources away from the pessimists, and so the weight of opinion as expressed in the market determines whether how much change takes place.
- Because information is a public good, it is inefficient to replicate its collection unless communication costs are high. Furthermore, it is better to concentrate information processing on people with a comparative advantage in interpretation – i.e. those whose beliefs are closest to the truth. These will tend to be the people with a track record of successful decisions. Intermediaries therefore emerge who specialize in processing information of particular kinds. Entrepreneurs intermediate by setting up new firms to sell new products, whilst social leaders intermediate by setting up new clubs and charities.
- Each person's utility depends upon emotional as well as material rewards. Change often elicits a powerful emotional response; some people thrive on the excitement of change, while others fear its consequences. Leaders need to be calm when taking decisions – they have to be confident in their judgments. They also need to understand the anxieties of their followers and provide them with reassurance if they can.
- Emotions are morally framed. Pride and self-esteem on the one-hand, and guilt and shame on the other, are powerful emotions. Leaders can associate positive emotions with actions that promote coordination and negative emotions with actions that undermine coordination. This engenders trust and so reduces agency costs and transactions costs. Improved coordination enhances the performance of the group. The leader can recover costs from this enhanced performance by various means – taxes, membership fees, voluntary donations – depending upon the type of group involved.

These assumptions are perfectly compatible with a rational action approach to modeling. However, the detailed specification of a model is rendered difficult by the fact that both theories and facts are uncertain. Nevertheless, the basic structure of the model can be set out using three propositions:

- Leadership operates at different levels. High-level leaders control nation states, organized religions and international pressure groups. Middle-level leaders manage firms, clubs and charities, whilst low-level leaders manage families and local communities. High-level leaders set a high-level culture within which the other leaders must operate. Lower-level leaders can 'free-ride' on useful values and beliefs inculcated by the high-level leader, but if they disagree with the values promoted at

the higher level they must invest in counteracting them. This issue separates people into those who prefer to assimilate and conform, and those who oppose or resist instead [Jones (1984)].

- Followers have a choice of leader. In a democracy they are free to vote for a political party and to practice their preferred religion; they can also decide which firm to work for, which clubs to join, and which charities to support. People recognize that when they decide to follow the leader of a particular group they must adopt the leader's values and beliefs. Many key decisions regarding choice of leader are made around the time a person comes of age. Using the prior beliefs inculcated in their childhood by their family and community, people decide which leaders they will follow in their adult life. They evaluate the risk that given leader's values and beliefs will turn out to be wrong. They take account of their own personal characteristics, as they perceive them, because these will determine their emotional responses later on. The final choice that an individual makes will reflect not only their beliefs but also their preferences – whether he is selfish or altruistic, material or emotional, and so on.
- Leaders seek to optimize the values and beliefs they promote in order to fulfill their own objectives. Honest leaders will promote their true beliefs – acting on conviction – but dishonest leaders may adapt their values in order to maximize their following. Culture change will occur both through leaders modifying their values to maintain market share, and by followers switching between committed leaders who are unwilling on principle to adjust their values for the sake of expediency.

These propositions show how the basic economic principles of choice and competition can be applied to culture. The economic theory of culture subsumes standard neoclassical economics as a special case. In a simple neoclassical economic model, there is just a single culture which corresponds to the 'true' model of the economy. This 'true' model assumes that people are selfish and materialistic. It is therefore a model of a low-trust society. The high-trust alternative is excluded by assumption. It is also a model of an individualistic society, since people care nothing about the welfare of others and take a purely instrumental view of the kind of society in which they live.<sup>9</sup>

## 5.2. Historical perspectives

The empirical and historical literature linking culture to economic performance is extremely diffuse. It is possible, however, to identify three specific issues which have had a significant impact on the economic analysis of culture: the Weber thesis, obstacles to development, and the role of freedom.

Economic historians have long debated the Weber thesis that the Protestant Ethic promoted the growth of capitalism [Weber (1930)]. There is broad agreement that the spread of international commerce in Europe coincided with the Reformation (although

<sup>9</sup> For a comprehensive critique along these lines see Roberts and Holden (1972) and Schoeffer (1955).

pre-reformation origins in Italian city-states must not be overlooked). However causality has been questioned. The Protestant Ethic can also be understood as accommodating Christian beliefs to the requirements of an emerging mercantile middle-class [Schlicht (1995)]. Behind the theological revolution, therefore, a vested business interest may be detected. Protestantism 'dis-intermediated' the Papacy and gave people a direct relationship with God through prayer. It undermined the case for paying the church for indulgences and the upkeep of chantries, and for obeying prohibitions on usury – and thereby reduced the economic burdens on the middle-class.

The theological content had real effects, however. The Protestant convert accepted grace through personal salvation. The sign of grace was not monastic seclusion, as before, but spreading the Gospel through engagement with the world. Business was a 'calling' which could promote missionary work. It supported the expansion of commercial empires into the 'darker corners' of the world. Whilst the origins of Protestantism may be questioned, therefore, its effects appear to be those which Weber predicted. Protestantism replaced the collectivist and procedural culture of the Roman Catholic church with a more individualistic and pragmatic culture, which formed the foundations of the competitive individualism that characterizes the West today.

Jones (1981, 1988) examines the 'take off' of commercialism in Western Europe from a different perspective and arrives at rather similar conclusions. Jones regards entrepreneurship as a natural human behavior which supports survival by encouraging people to show initiative in meeting their material needs. However entrepreneurship can be stifled by political tyrannies, in which collectivism and proceduralism are imposed [Rosenberg and Birdzell (1986)]. The motive is to monopolize the tax-base and use its revenues to support a leisured lifestyle for the elite. From this perspective the Reformation is a protest movement which, by overthrowing a parasitic religious elite, liberates people to follow their natural entrepreneurial inclinations. China and other Asian powers have never liberated themselves in this way; when one elite is deposed, another simply takes its place. Once again, however, the explanation may be cultural – perhaps Western society is intolerant of political oppression in the way that some Asian societies are not.

Development economists have addressed similar issues but from a more secular perspective [Bardhan (2000)]. A drive to 'modernize' post-colonial societies is typically advocated [McClelland and Winter (1969)]. In the 1960s modernization became the secular equivalent of the Protestant ethic. The object was to engineer a high-tension society driven by a desire to catch up with the West, in place of a low-tension society where people are content with low living standards and high mortality. Individualism was a secondary consideration; in the 1960s planned industrialization behind protective tariffs was the recommended strategy, and it was only in the 1990s that privatization and liberalization took over.

A major obstacle to economic development in the poorest countries is weak internal communications which perpetuate a cellular social structure based on local family and tribal loyalties. High levels of local trust are combined with low levels of trust at the national level. National government is too corrupt to intermediate the flow of funds between international agencies and local people. The engineering of trust at the na-

tional level has been accomplished in a number of Asian economies but with one or two notable exceptions there has been little success in Africa. As noted earlier, creating a high-tension high-trust society has proved difficult even in prosperous Western countries.

The disintegration of Soviet communism has led to a resurgence in research dedicated to showing that 'freedom' holds the key to economic performance [Gwartney and Lawson (2003)]. The guarantor of freedom is usually said to be a US-style constitution [Scully (1992)]. A range of freedom indicators has been developed, and cross-country statistical regressions have been reported which confirm the impacts of freedom on living standards and economic growth. On the whole these regressions simply confirm that, other things being equal, Western-style competitive individualism promotes economic growth. The point is not difficult to make if a sufficient number of poor African dictatorships is included in the sample of countries. As in any cross-section regression, there are omitted variables, and much of the sample variation remains unexplained. The apparent significance of some of the variables may be due to the presence of omitted cultural variables, including the legacy of traditional religion [Kohut et al. (2000)]. Whilst these regressions are a significant advance on anecdotal evidence, the range of explanatory variables is too narrow to offer a full account of cultural factors in economic performance.

Advocates of freedom as the critical factor are usually unsympathetic to a cultural interpretation of their findings and this biases the way in which they interpret their results. They typically believe that laws, not morals, reduce agency costs and transaction costs. They believe that a written constitution enforced through impartial courts is better than an unwritten constitution enforced through social sanctions. They believe that the biological drives such as greed and aggression are better guarantors of competition than a genuine desire to benefit the customer. They therefore ignore crucial issues such as why greedy judges do not accept bribes, and how the basic needs of people with low incomes are met.

The historical significance of culture is related to the historical significance of other intangible public goods, such as technological know-how. It is therefore not surprising that modern writers on convergence of national economic growth rates have begun to develop an interest in cultural issues. The traditional way of analyzing the convergence of growth rates focuses on technological diffusion, but there is no reason why the analysis should not include cultural diffusion too. The rapid spread of free-market ideology in the 1990s, with many governments reducing tariffs and privatizing and deregulating their utilities, is a clear example of cultural diffusion. Such cultural diffusion can lead to convergence in institutions as well as in rates of growth. A particularly interesting development has been the incorporation of religion in the convergence model.<sup>10</sup> Whilst the European empires of the nineteenth century are often credited with the spread of Christianity, the US-led Western 'empire' of the late twentieth century is noted chiefly

<sup>10</sup> For a significant step in this direction see Barro and McCleary (2003).

for its spread of secularism. This raises the issue of whether religion or secularism is best for economic growth. If religion is best then the spread of secularism could lead to convergence on a sub-optimal level of growth. The analysis in this chapter suggests that it is the specific content of religious belief that is crucial in this respect, because it is the specific beliefs that determine the emotional incentive structure which motivates people. A simple distinction between religion and secularism is therefore too crude to properly identify the link between religious belief and economic performance. The impact of the spread of religion and culture on the convergence of growth rates is clearly an important topic which warrants further research.

## 6. Conclusion

This chapter has shown that the influence of culture on the economy extends well beyond the production and consumption of cultural goods in the field of media and the arts. Culture is concerned with the production and distribution of values and beliefs relating to fundamental issues. Cultural products are simply one of the means through which these values and beliefs are expressed. Identifying the fundamental issues addressed by culture is the key to analyzing its impact on economic performance. Values and beliefs of a suitable kind can improve economic performance – both materially, and by enhancing quality of life. Culture is therefore an economic asset. Culture is shared by communication between the members of a social group. It is, in fact, an intangible durable public good. Significant investment is required to create and maintain this public good. Competition between cultures, in terms of relative economic performance, is essentially competition between social groups in investing in appropriate public goods of this type.

By modifying five key assumptions of conventional neoclassical economics, and introducing a theory of leadership, it is possible not only to explain how culture influences performance but also to explain how cultures will adapt to changing local conditions. There are different levels of leadership, corresponding roughly to the size of the group that the leader controls. At any given level the nature of competition is strongly influenced by the media that leaders employ to recruit and retain their followers. The development of mass media disseminating visual images has had a profound effect on ideological competition between political leaders. Changes in the media have made the promotion of high-trust cultures extremely difficult, whilst a skeptical attitude towards leadership in general has diminished the supply of able leaders. Distorted incentives in the market for leadership mean that the most effective culture does not always prevail.

The ideal culture from an economic point of view is individualistic, pragmatic, high-trust and high-tension, though each of these attributes must be moderated to some degree by the need to adapt the culture to local requirements. A simple way of summarizing the advantages of this culture is to note that it is both entrepreneurial and moral. It is entrepreneurial because it encourages innovation and risk-taking, and it is moral because it discourages innovations or risky ventures that cause disproportionate damage to the



interests of others. It is moral because it encourages honesty and loyalty, but it is entrepreneurial because it does so without stipulating rigid conformity to specific practices.

The high-performance culture also encourages both freedom and responsibility. Freedom allows diversity of behavior and thereby facilitates innovation. It also decentralizes power: it allows decisions to be taken by people who have immediate access to relevant information, and so avoids the expense and delay of referring straightforward decisions to higher authority. However responsibility requires people to show consideration for others [Ellickson (1991)]. In respecting other people's freedoms, they accept constraints on their own. They consult with other people before acting in an unexpected way. Consultation is effected both formally and informally. A high-trust culture encourages people to honor informal agreements. A legalistic culture sets out rights and responsibilities, records them and enforces them. People are obliged to negotiate with people who hold the relevant rights before they act. Informal methods work well with members of a tightly-knit social group – friends, relatives and neighbors – whilst formal methods are more appropriate for more impersonal groups. A moral culture will rely on trust as much as possible but will underpin trust by the rule of law.

The high-performance culture respects both tradition and modernity. Embracing modernity promotes scientific research and the practical application of science in engineering and medicine. It also encourages economy through the systematic elimination of waste. Tradition on the other hand underpins many core moral values. Conflict can ensue when scientific discoveries appear to undermine traditional religious beliefs on which conventional morality is based. Some religions are more vulnerable than others on this score, however. An entrepreneurial culture is not devoid of religion, but rather involves religious beliefs which co-exist with a scientific view of the world.

## References

- Bardhan, P. (2000). "The nature of institutional impediments to economic development". In: Olson, M., Kahkonen, S. (Eds.), *A Not-so-Dismal Science: A Broader View of Economics and Societies*. Oxford University Press, Oxford, pp. 245–268.
- Barro, R., McCleary, R.M. (2003). "Religion and political economy in an international panel". Discussion Paper. Harvard University, Cambridge, MA.
- Baxter, J.L. (1988). *Social and Psychological Foundations of Economic Analysis*. Harvester Wheatsheaf, Brighton.
- Bolnick, B.R. (1975). "Towards a behavioural theory of philanthropic activity". In: Phelps, E.S. (Ed.), *Altruism, Morality and Economic Theory*. Russell Sage, New York, pp. 197–224.
- Buruna, I. (1999). *Voltaire's Coconuts, or Anglomania in Europe*. Weidenfield & Nicolson, London.
- Casson, M.C. (1990). *Entrepreneurial Culture as a Competitive Advantage*. Oxford University Press, Oxford.
- Casson, M.C. (1991). *Economics of Business Culture: Game Theory, Transactions Costs and Economic Performance*. Clarendon Press, Oxford.
- Casson, M.C. (1993). "Cultural determinants of economic performance". *Journal of Comparative Economics* 17, 418–442.
- Casson, M.C. (2000). *Enterprise and Leadership: Studies on Firms, Markets and Networks*. Edward Elgar, Cheltenham.

- Casson, M.C. (2002). "Leadership and cultural change: An economic analysis". *De Economist* 150 (4), 409–438.
- Earl, P.E. (1986). *Lifestyle Economics: Consumer Behaviour in a Turbulent World*. Wheatsheaf, Brighton.
- Easterlin, R.A. (1998). *Growth Triumphant*. University of Michigan Press, Ann Arbor.
- Easterlin, R.A. (2001). "Income and happiness: Towards a unified theory". *Economic Journal* 111, 465–484.
- Ekelund, R.B. Jr., Tollison, R.D. (1997). *Politicized Economies: Monarchy, Monopoly and Mercantilism*. Texas A&M University Press, College Station, TX.
- Ellickson, R.C. (1991). *Order without Law: How Neighbours Settle Disputes*. Harvard University Press, Cambridge, MA.
- Foreman-Peck, J., Federico, G. (Eds.) (1999). *European Industrial Policy: The Twentieth Century Experience*. Oxford University Press, Oxford.
- Frank, R.H. (1985). *Choosing the Right Pond*. Oxford University Press, New York.
- Ginsburgh, V., Michel, P. (1997). "Optimal policy business cycles". *Journal of Economic Dynamics and Control* 22, 503–518.
- Graham, J.L. (2001). "Culture and human resources management". In: Rugman, A.M., Brewer, T.L. (Eds.), *Oxford Handbook of International Business*. Oxford University Press, Oxford, pp. 503–536.
- Gwartney, J., Lawson, R. (2003). *Economic Freedom of the World: 2003 Annual Report*. Fraser Institute, Vancouver, BC.
- Hahnel, R., Albert, M. (1990). *Quiet Revolution in Welfare Economics*. Princeton University Press, Princeton, NJ.
- Hampden-Turner, C., Trompenaars, F. (1997). *Mastering the Infinite Game*. Capstone, Oxford.
- Hirschman, A.O. (1982). *Shifting Involvements: Private Interests and Public Action*. Martin Robertson, Oxford.
- Hofstede, G. (1980). *Culture's Consequences*. Sage, Beverly Hills, CA.
- Holmes, S., Sunstein, C.R. (1999). *The Cost of Rights: Why Liberty Depends on Taxes*. W.W. Norton, New York.
- Jones, E.L. (1981). *The European Miracle*. Cambridge University Press, Cambridge.
- Jones, E.L. (1988). *Growth Recurring: Economic Change in World History*. Oxford University Press, Oxford.
- Jones, E.L. (2000). "The case for a shared world language". In: Casson, M., Godley, A. (Eds.), *Cultural Factors in Economic Growth*. Springer, Berlin, pp. 210–235.
- Jones, S.R.G. (1984). *The Economics of Conformism*. Blackwell, Oxford.
- Knight, F.H. (1935). *The Ethics of Competition and Other Essays*. Allen & Unwin, London.
- Kohut, A., Green, J.C., Keeter, S., Tuth, R.C. (2000). *The Diminishing Divide: Religion's Changing Role in American Politics*. Brookings Institution, Washington, DC.
- Landes, D.S. (1998). *The Wealth and Poverty of Nations*. W.W. Norton, New York.
- Layard, R. (1980). "Human satisfactions and public policy". *Economic Journal* 90, 737–750.
- Leacock, S. (1998). "My Recollection of Chicago" and "The Doctrine of Laissez Faire". University of Toronto Press, Toronto. Edited by C. Spadini.
- Lucas, R.E. (1981). *Studies in Business Cycle Theory*. MIT Press, Cambridge, MA.
- Macfie, A.L. (1967). *The Individual in Society: Papers on Adam Smith*. Allen & Unwin, London.
- McClelland, D.C., Winter, D.G. (1969). *Motivating Economic Achievement*. Free Press, New York.
- Newman, K.S. (1983). *Law and Economic Organization: A Comparative Study of Pre-industrial Societies*. Cambridge University Press, Cambridge.
- O'Brien, G.E. (1988). "Work and leisure". In: Van Raaj, W.F., Van Veldhoven, G.M., Warnweryd, K.E. (Eds.), *Handbook of Economic Psychology*. Kluwer, Dordrecht, pp. 538–569.
- Olson, M. (2000). "Big bills left on the sidewalk: Why some nations are rich and others poor". In: Olson, M., Kahkonen, S. (Eds.), *A Not-so-Dismal Science: A Broader View of Economies and Societies*. Oxford University Press, Oxford, pp. 37–60.
- Protherough, R., Pick, J. (2002). *Managing Britannia: Culture and Management in Modern Britain*. Brynmill Press, Denton.
- Pryor, F.L. (1977). *The Origins of the Economy: A Comparative Study of Distribution in Primitive and Peasant Economies*. Academic Press, New York.

- Putnam, R.D. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press, Princeton, NJ.
- Reisman, D. (1990). *Theories of Collective Action: Downs, Olson and Hirsch*. Macmillan, London.
- Roberts, B., Holden, B.R. (1972). *Theory of Social Process: An Economic Analysis*. Iowa State University Press, Ames, IA.
- Rosenberg, N., Birdzell, L.E. (1986). *How the West Grew Rich: The Transformation of the Industrial World*. I.B. Tauris, London.
- Schlicht, E. (1995). "Economic analysis and organized religion". In: Jones, E., Reynolds, V. (Eds.), *Survival and Religion: Biological Evolution and Cultural Change*. John Wiley, Chichester, pp. 111–162.
- Schoeffler, S. (1955). *The Failures of Economics: A Diagnostic Study*. Harvard University Press, Cambridge, MA.
- Scitovsky, T. (1976). *The Joyless Economy: An Inquiry into Human Satisfaction and Consumer Dissatisfaction*. Oxford University Press, New York.
- Scully, G.W. (1992). *Constitutional Environments and Economic Growth*. Princeton University Press, Princeton, NJ.
- Sherman, A.L. (1997). *The Soul of Development: Biblical Christianity and Economic Transformation in Guatemala*. Oxford University Press, New York.
- Stark, R., Bainbridge, W.S. (1987). *A Theory of Religion*. Peter Lang, New York.
- Viner, J. (1972). *The Role of Providence in the Social Order: An Essay in Intellectual History*. American Philosophical Society, Philadelphia.
- Viner, J. (1978). *Religious Thought and Economic Society: Four Chapters of an Unfinished Work*. Duke University Press, Durham, NC. Edited by J. Melitz and D. Winch.
- Weber, M. (1930). *The Protestant Ethic and the Spirit of Capitalism*. Allen & Unwin, London. Translated by T. Parsons.

## CULTURE AND ECONOMIC DEVELOPMENT\*

PAUL STREETEN

*Boston University, USA*

### **Contents**

Abstract	400
Keywords	400
1. Culture in development thought	401
2. Diversity	402
3. The role of conflict	404
4. Globalization, development and culture	405
5. Tourism and development	407
6. Summary and conclusions	409
References	412

\* I am immensely grateful to David Throsby who has done a superb job in editing this chapter.

**Abstract**

This chapter begins by noting that culture as an element in economic development in the Third World has been largely neglected in traditional development economics, most writers either seeing culture as an obstacle to development or ignoring it altogether. Recently a shift in thinking has occurred whereby culture is now more widely seen as being more central to the development process, especially where a human-centered rather than a goods-centered view of development is taken. A particular aspect of culture that has been seen as important has been cultural diversity; it is argued that the beneficial aspects of diversity can only be realized when they are seen within a global ethical framework. The chapter goes on to consider the destructive and constructive role of conflict in bringing about social change, and discusses the pervasive effects of globalization on the economies and cultures of the world, arguing that international integration can lead to national disintegration. Next the chapter looks at the role of tourism as a significant economic and cultural force in developing countries. Finally the essay concludes with some recommendations for policy.

**Keywords**

culture, economic development, diversity, conflict, globalization, tourism

*JEL classification:* O10, Z10

## 1. Culture in development thought

The role of culture in our thinking about development has undergone a remarkable change in the last four decades. In spite of occasional declarations to the contrary such as those in the Universal Declaration of Human Rights of 1948, which states in Articles 22 and 27 that “everyone has the right to participate in the cultural life of their community, to enjoy the arts and to share in scientific advancement and its benefit” [United Nations General Assembly (1949)], most economists have either ignored culture completely or have found traditional cultural attitudes and practices obstacles to development. The neglect of culture by economists has been reflected until recently in the work of international institutions such as the World Bank.

Among those who took culture into account but saw it as a force working against development, Gunnar Myrdal (1968) is an outstanding example. In *Asian Drama*, for instance, he argued that traditional practices such as those relating to hygiene (or rather lack of hygiene), and traditional attitudes to work, leisure, saving, competition and co-operation must be eradicated before countries can embark on their modernization. He, in the company of many development experts in the 1950s and 1960s, regarded most traditional cultural “attitudes and institutions” as major obstacles to development. He thought that the people in developing countries were lethargic and work-shy, lacked habits of cleanliness and punctuality, had no desire to make money and lacked a willingness to cooperate, to experiment, to explore and to adopt a rational approach to life and work. In short, their traditional attitudes were hostile to development. The implication was that unless the people in developing countries adopted the culture, habits and customs of advanced Western countries, there was no hope of their developing. It is only a small exaggeration to say that Myrdal’s view was that not until the Indians could become like modern Swedes would successful development in India occur.

Even when some writers did not think that culture was an obstacle to development and even when they recognized it as of some value, they attached low priority to it. Culture can be taken care of, these authors thought, after more urgent needs such as food and health have been met. Others regarded some cultural features such as frugal consumption habits, an obedience to authority and a willingness to cooperate such as that displayed by the Japanese, as useful instruments of development. They advocated building modernity on features of tradition they saw as being useful for this purpose.

Since then the pendulum has swung in the opposite direction. Culture has become the fountain of progress and a key element in the development process. This change in outlook has been associated with a shift from thinking about development in purely material or commodity-centered terms to an approach orientated towards human development. The United Nations Development Strategy for the 1990s, for example, adopted human development as its key focus, reflected in the UNDP’s annual *Human Development Reports* which first appeared in 1991. It is no coincidence that the 2004 edition of the *Human Development Report* was entirely devoted to culture, under the title *Cultural Liberty in Today’s Diverse World* [UNDP (2004)].

Among those who argue that culture matters, Amartya Sen has played an outstanding part. He writes “[t]he issue is not whether culture matters . . . [t]he real issue, rather, is how culture matters. What are the different ways in which culture may influence development?” [Sen (2004)]. And in the process of exploring this question, he argues against the view that the fates of countries are effectively sealed by the nature of their respective cultures. It is wrong to see culture as independent, unchanging and unchangeable; it should be seen as non-homogeneous, non-static and interactive. We are now beginning to see that culture should neither be neglected, as it has often been in the past by economists, nor elevated to a homogeneous, stationary and deterministic force, as some models of anthropology have done.

There are two views of the relationship between development and culture. According to the first, economic growth is the objective and culture a means to promoting it. The value of culture is entirely instrumental. Protestantism and, by a later extension Confucianism and other religions, have been thought of as cultures that contribute to savings, hard work, discipline, punctuality, hygiene and healthy living habits, and are therefore to be welcomed. According to this view, as noted above, if cultural attitudes and habits hamper economic growth, they should be eradicated. This view of the relationship between culture and development is interesting and important, but one may question whether economic growth is the end.

The alternative view sees growth as the means to our freedom to live the way we value. And what we value and cherish is a matter of culture. Looked at this way, culture is the desirable end; it is what gives meaning to our existence. This dual role applies not only to growth, but also to sustaining our environment, preserving family values or protecting civil institutions. We value some cultural features as means to certain objectives. But when we ask why we value these objectives, culture enters not as a servant of ends but as the social basis of the ends themselves. Economics and politics should, ideally, not only use culture but also serve culture.

What if conflicts arise between the preservation of a culture and the attitudes and institutions that are needed for economic growth and development? Inevitably, growth and change bring with them the demise of traditional ways of life, customs, styles and artifacts. As Sen (1999) has argued, it is then for the people to decide whether to sacrifice material goods for the preservation of a culture or whether to sacrifice certain cultural features for greater prosperity. “[I]n the freedom-oriented perspective the liberty of all to participate in deciding what traditions to observe cannot be ruled out by the national or local ‘guardians’ – neither by the ayatollahs (or other religious authorities), nor by political rulers (or governmental dictators), nor by cultural ‘experts’ (domestic or foreign).” [Sen (1999, p. 32)]

## 2. Diversity

A feature of culture that has moved to center stage in recent considerations about culture in economic development is cultural diversity. Cultural differences are celebrated

not only as a means to development but also as one of the most important ends of the development process. The emergence of diversity as a focus for development thinking is reflected in the work of UNESCO. It was responsible for the World Commission on Culture and Development whose report, published in 1995, was entitled *Our Creative Diversity* [World Commission on Culture and Development (1995)]. Subsequently UNESCO formulated the *Universal Declaration on Cultural Diversity*, adopted in 2001 [UNESCO (2001)], and then initiated a process aimed at drawing up an international convention on cultural diversity which would have the effect of raising the profile of culture in national and international affairs and of locating culture as an essential element in Third World development.

The move from condemning or attaching a low priority to celebrating cultural diversity, both between and within cultures, and from considering it as an obstacle to development to regarding it as part of the end, has been an illuminating change. It reminds us that development can take many forms, that styles of development can differ and that we are not all destined to end up as uniform Californian-type mass consumers. Yet it is clear that not all traditional cultural practices are either desirable in themselves or contribute to development. For this certain universal principles are necessary. There are indeed ethical principles that are accepted by all cultures, and these can form the basis for a global ethics. They furnish the minimal standards any political community should observe; beyond them, there is scope for different political visions, influenced by different cultural heritages and historical experiences. Among these minimum ethical principles is the respect for life, the need to alleviate suffering whenever possible and to avoid inflicting unnecessary suffering, to treat others as one wants to be treated oneself, the idea of human rights, the protection of people's integrity and respect for the vulnerability of all human beings. Democratic government, respect for the rights of minorities and peaceful conflict resolution are also part of the global ethics, as is the basic principle of intergenerational equity, which calls on present generations to take care of and use the environment and cultural and natural resources for the benefit of all members of present and future generations.

In the early days of development economics it was thought that mobilizing resources, and particularly savings for investment in physical capital, was all that was needed for development. Today we know that development calls for much more: development means improving the lives of the poor by transforming societies. This inevitably means social and cultural change; differences within and between cultures mean that desirable processes of change will not be uniform but will vary from place to place and time to time depending on economic, social and cultural circumstances.

Although cultural diversity is in a sense inevitable, there is something to be said in favor of getting rid of it. For example, diversity can lead to inefficiency in the expenditure of time and effort in communication; it would be so much simpler to communicate in a world where that was only one language. Standardized, uniform methods of communicating and conducting business would save resources. Nevertheless, the weight of opinion and evidence suggest that cultural diversity is desirable and valuable, for several reasons. First, diversity is valuable in its own right as a manifestation of the creativity of



the human spirit. Second, it is required by principles of equity, freedom of choice, human rights and self-determination. Third, in analogy to biological diversity, it can help humanity to adapt to the limited environmental resources of the world. In this context diversity is linked to sustainability. Fourth, it is needed to oppose political and economic dependence and oppression. Fifth, it is aesthetically pleasing to have an array of different cultures; so, for example, people enjoy consuming cultural expressions such as music originating from cultures other than their own. Sixth, it stimulates the mind and encourages creativity. Finally, it can provide a reserve of knowledge and experience about good and useful ways of doing things.

### 3. The role of conflict

Of course diversity can give rise to conflict, as the “clash of civilizations” thesis asserts [Huntingdon (1996)]. It is readily observable that the politicized meaning of “culture”, in the sense of the lifestyle of minorities, can give rise to tension and violence. Small differences such as those between the Muslims and Christians in the old Yugoslavia or between the Albanians and the Serbs in Kosovo, or between the Protestants and Catholics in Northern Ireland, or between the descendants of the sons of Abraham – the Jews and the Arabs – in West Asia, can give rise to tensions that break out in violent conflicts. They are at heart narcissistic conflicts.

Conflict is normally viewed as destructive of the social order. But it can be argued that conflict is not necessarily an obstacle to successful development [Streeten (1953), Hirschman (1995)]. Heraclitus thought that “war is the father of everything” and Machiavelli entitled a chapter in the *Discourses* “How the Disunion between the Plebs and the Senate Made [the Roman] Republic Free and Powerful”. One may even go further. Conflict, or at least some forms of it, can also be regarded as a pillar of democratic societies, as the glue that holds them together. Conflicts can provide society with the “social capital” it needs to be kept together. Hirschman (1995) has made a beginning in distinguishing between when conflict is destructive and when constructive. He distinguishes between conflicts about more or less, such as the distribution of income, and conflicts about either/or, such as abortion. Conflict arises inevitably with change. Globalization and technical progress benefit some countries, some regions, some sectors, and some groups, and harm others. In free societies those who suffer will tend to organize themselves and attempt to regain their position. Those who agree with them from a sense of social justice or sympathy will support them. One group is motivated by self-interest, the other by solidarity or a sense of fairness or fellow feelings. The strength of democratic societies derives from this combination and from the conflicts to which it gives rise.

In poor countries, conflict arises from a number of causes. Poverty can be a significant factor contributing to conflict, especially in the competition for resources. How far inter-tribal or inter-ethnic conflicts can be explained by economic or by cultural considerations is a moot point; Collier and Hoeffler (2000) in their research for the World Bank

suggest that it is often “greed” rather than “grievance” that precipitates such conflicts, in other words that economic rather than cultural motives underlie the violence.

Paradoxically, the reduction of poverty can also lead to conflict. If poverty comprises many more dimensions than lack of income, and includes deprivation of education and health, social exclusion, lack of employment, discrimination against women, environmental degradation of soil, water, forests and climate, insecurity, violation of human rights, lack of voice in the counsels of society, and lack of cultural expression, the chances of conflict over the reduction and eradication of poverty are greatly increased. Income can be divided in different proportions and is therefore easier to negotiate and to compromise on than decisions that are subject to an either/or choice. Ethnic, linguistic, religious and gender divisions and disagreements on voting rights give rise to non-divisible conflicts. Unfortunately it seems that these types of conflict which are not readily amenable to negotiation and compromise are on the increase.

#### 4. Globalization, development and culture

It has become a cliché to say that international interdependence is great, has increased, and will continue to grow. Normally this is intended to refer to trade, foreign investment, the flow of money and capital, and the migration of people. Advances in technology such as the jet, telex, satellite TV, container ships, super tankers and super ore carriers, and technical progress in transport, travel and above all in communication and information, have shrunk the world. By reducing the cost of communication, technology has helped to globalize production and finance. Globalization, in turn, has stimulated technological progress by intensifying competition, and competition has forced the introduction of new technology. Globalization has spread its results widely through foreign direct investment. History may not have ended, but geography, if not coming to an end, certainly matters less. And the interaction of technology and globalization has presented new problems.

The international spread of ideological and cultural impulses is at least as important as that of economic impulses. Observe the young in the capitals of the world: from Ladakh to Lisbon, from Maine to Mozambique, from West Virginia to East Jerusalem, from China to Peru, in the East, West, North, and South, styles in dress, jeans, hairdos, T-shirts, jogging, eating habits, musical tunes, attitudes to homosexuality, divorce, abortion and so on have become global. Even crimes such as those relating to drugs, the abuse and rape of women and children, embezzlement and corruption have become similar everywhere. But although American cultural influences are important, there are many other influences. As *Sainath (2002)* argues, “(t)he super-rich are seceding from their nations. So what you have is not a Western or East Asian or Southeast Asian or Chinese model. We are building enclaves of super-privilege. What you’re having is not a global village but a series of global ghettos. The Western élite is not the sole villain”.

But the impression of global uniformity can be deceptive. Just as trade, foreign investment and the flow of money have affected only a few regions of the world and left

the rest comparatively untouched (except for some negative effects), so the globalization of culture is only partial. Large areas of the world do not participate in these globalizing tendencies. The effects of globalization are evident in the towns and suburbs and the more advanced countryside, but the poor in the rural hinterlands, in spite of the spread of transistors and television, have been largely bypassed. And in many lands there has been a return to tradition and tribalism. Global integration has provoked national disintegration; nations have broken up into smaller, ethnic groups. An assertion of their indigenous values is often the only thing that poor people can do. Traditional values bring identity, continuity and meaning to their lives. Between the two opposite forces, the assertion of peoples' identities on the one hand, and globalization on the other – between what Barber (1995) calls Jihad and McWorld – nation states have found their base undermined [Streeten (2001)].

There are five reasons why partial international integration can lead to national disintegration. First, downsizing, restructuring, “delaying”, and re-engineering have reduced the demand for low-skilled workers in the rich and middle-income countries and have kept the wages of those who succeeded in keeping their jobs low. This has led to growing inequalities in the rich countries. Second, preventing excessive growth in the brain drain from developing countries makes egalitarian incomes policies impossible in those countries. Third, tax revenues to pay for social services have been reduced, though the need for them has increased. Fourth, the élites in low-income countries are opting out of national commitments; this leads to the neglect of essential social services like education and health. Fifth, the culture of these élites is global and estranged from the culture of the local people. Finally, the tendency of minorities to break away from their country and to form independent states, resulting in the proliferation of states, can be explained by their desire to participate directly in the benefits of globalization.

It is therefore apparent that in the developing world, globalization makes national government more difficult. Monetary and fiscal policies run up against the impact of global tides as people, international banks and multinational corporations avoid the intended results by sending or spending their money abroad or attracting money from abroad. The obligations of extended families, government and religion disappear as people leave their rural communities to live in large cities. Recently-enriched members of the middle class with links to politicians and officials often use their newly acquired powers in corrupt ways that counteract traditional values. Impoverished, disoriented provincials overrun cities, once the seat of the privileged.

To sum up, globalization has brought benefits to some individuals or groups in the developing world, but has left out or harmed others. The same can be said of whole countries. Faced with these effects, governments may adopt one of three policy responses: they may give unqualified support to joining the global economy; they may aim for complete isolation and delinking from the global economy; or they may attempt some selective gearing into and selective delinking from the global economy. Common sense might indicate a preference for the last of these, so that the community can benefit from the good impulses propagated by the global system while excluding the bad ones.

The question is whether inadequate bureaucracies and politicians can be entrusted with this task.

## 5. Tourism and development

One aspect of the growth of the global economy that has both economic and cultural implications for developing countries is tourism. As a source of economic benefits, tourism has much to offer. Compared with agricultural commodities it is less volatile and compared with manufactured products it is less subject to import restrictions by advanced countries. It creates many jobs if the multiplier effects work; visitors to hotels create a demand for souvenirs, restaurant services, boat and taxi rides, car services, furniture for hotels, food, etc. It is a potentially large foreign exchange earner.

Furthermore, tourism raises the educational levels of the local population, encourages growth of a middle class (who will themselves become tourists) and promotes social mobility. American Indians' and Eskimos' art has been reinvigorated by tourist demand. Tourism encourages contact with the rest of the world through radio, TV, advertising and migration. It is a useful way of diversifying the economy. It is better to sell sugar, sea and sun than sugar alone and even better to sell safaris. There are thus many good arguments for encouraging tourism in developing countries, but there are also some arguments for caution in encouraging cultural tourism, especially if income disparities between tourists and the local population are large.

First, it is not entirely a play on words to say that rendering services for tourists such as waiting, hotel and restaurant services, driving, etc. tends to make for servile attitudes unless incomes per head are already quite high. For Switzerland, Bermuda, Malta and Cyprus tourism is all right because the local population has already achieved a decent standard of living, self-respect and a sense of identity. But in the Caribbean or in Africa or in the Pacific, the situation is different. Maurice Bishop, a former Prime Minister of Grenada, said, "It is important to face the fact that . . . most of the tourists who come to our country happen to be white, and this clear association of whiteness and privilege is a major problem for Caribbean people just emerging out of racist colonial history where we have been so carefully taught the superiority of things white and inferiority of things black" [quoted in [Pattullo \(1996, p. 64\)](#)].

Second, high-income tourists can corrupt a country in other ways. Cuba was used by Americans as a "playground", which was a polite word for America's brothel. The encouragement of prostitution, venereal disease and crime contributed to Castro's revolution. Third, the local population tends to want to copy the tastes of the tourists. There is a demonstration effect on consumption habits. If their incomes are low this will lead to frustration and unhappiness. Fourth, tourism can upset male-female relationships if the women get better-paid jobs in hotels. This may, of course, contribute to their libera-

tion but it will upset the domestic peace. Fifth, tourism will tend to increase crime and violence as the frustrations of the local people grow.

Finally, tourism can change or corrupt or destroy the culture tourists have come to observe. The sheer numbers of tourists destroy the very things they come for. In Egyptian sites such as the Valley of the Kings or Sakkara thousands of visitors mill about each day in cramped tombs that were designed for one occupant's afterlife. The deterioration of the paintings and reliefs on the walls is plain to see. Even the apparently indestructible pyramids of Giza are suffering. As Throsby (2001) notes: "The cultural impacts of mass tourism are well known, ranging from the physical pressures imposed by large numbers of tourists on heritage sites to the damage that may be caused to local community cultural values if an area is constantly being invaded by crass and insensitive visitors" (p. 129).

There are also economic problems with tourism. Although it is often believed that tourism is labor-intensive, it can be very capital-intensive if hotel rooms are unoccupied for long periods out of season. It can also be very import-intensive if the tourists drink imported orange juice while local fruits are available, or if money is borrowed from abroad to finance the industry, or if hotels and related businesses such as car hire and yacht chartering are owned by foreign firms. In underdeveloped small island economies a large share of tourist earnings will leak back abroad through purchases of imported goods such as food, or through dividends or interest on invested capital, expatriate employees' salary remittances, etc. Thus, whereas Kenya, for example, produces many inputs into its tourist industry itself, many small islands in the Caribbean and in the Pacific do not.

The income elasticity of demand for tourism is liable to be high. This can lead to sudden large losses when incomes drop. When rich countries are faced with balance of payments troubles, tourism tends to be the first thing that gets cut. Furthermore, a flourishing tourist industry can cause inflation, push up prices, wages and exchange rates, displacing farming and making it harder for other industries to develop.

There are also political problems with tourism. It is a highly volatile industry. It is sensitive to energy crises, to political turbulence, news of crime, and other bad news such as natural disasters or terrorism. Sri Lanka's tourism halved between 1982 and 1986 as a result of the ethnic conflict, Africa's has suffered from the AIDS scare and Egypt's from the slaughter of tourists at Luxor. An additional problem is that some critics regard tourism as a new form of colonialism.

Can tourism be conducted in a more efficient, less destructive way that at the same time does not deprive the local population of its benefits and revenues? Can tourism become "sustainable"? Codes of practice for sustainable tourism in developing countries have been promulgated, stressing the need for cultural sensitivity and respect, care for the environment, and equitable treatment of host communities. Moreover, in a reaction to the destructive aspects of mass tourism, more constructive or alternative forms have evolved such as ecotourism, agrotourism, nature tourism and cultural tourism.

## 6. Summary and conclusions

It is a truism to say that one of the principal aims of development efforts is the eradication of poverty so that all people can develop their full potential. Yet all too often in the process of development it is the poor who shoulder the heaviest burden. In the transition from subsistence-oriented agriculture to commercial agriculture, poor women and children are sometimes hit hardest. In the transition from a traditional society in which the extended family takes care of its members who suffer misfortunes, to a market society in which the community has not yet taken on responsibility for the victims of the competitive struggle, the fate of these victims can be cruel. In the transition from rural patron–client relationships to relations based on the cash nexus, the poor suffer by losing one type of support without gaining another. In the transition from an agricultural to an industrial society, the majority of the rural people are neglected by the public authorities in favor of the urban population. In the transitions that we are now witnessing from centrally planned to market-oriented economies, and from autocracies to democracies, inflation, mass unemployment, growing poverty, alienation and new crimes have to be endured. In spite of four decades of development efforts, poverty remains high in many areas of the world. Over a billion people are estimated to fall below the poverty line [World Bank (2001, p. 23)].

Globalization and growing international interdependence present new challenges and opportunities for culture and cultural policies throughout the world. These challenges are environmental, political, social, human and cultural (in the narrow sense). This chapter has suggested that the role of culture will be of increasing importance in the discussion of development policies, especially of alternative approaches to development and its different styles. Culture is often invoked to explain both the successes and failures of development. For example, some have attributed the economic miracle of the East Asian economies to Confucian culture and so-called “Asian values” and asserted that cultural diversity is not only useful but also essential for development. It has also been established beyond doubt that the Protestant ethic is not the only source of thrift and hard work and that many religions can breed economic development. At the same time, the recent turmoil in these economies has also been attributed to certain features of local cultures. Both positive and negative traits of economic development are said to be linked to culture. Neither view is persuasive. The discussion requires a deeper analysis of the links between culture and development and its crises. Cultures are not static but change all the time, adopting features from other cultures while communicating other features to them.

I conclude with several recommendations for policy. First, although it is argued that diversity contributes to creativity, it is not enough to advocate diversity in the abstract. We must show precisely how diversity enhances economic success, social opportunities, political stability and conflict resolution, as well as being valuable, beautiful and delightful in itself. In the global system of cultural exchanges some cultures are disappearing. But as some forms of culture disappear, new forms are created, and they are created locally. The disappearance of old cultural forms is entirely consistent with a

rich variety of new forms of human life. Attitudes should be encouraged and laws and institutions should be established that recognize multiple voices and actors and provide ways of handling differences and plural, partial interpretations of the world. We should learn to value the variety of human forms of social and cultural life, so long as these differences meet the principles of a universal global ethics, in particular the respect for basic human rights such as those to life, liberty, due process, free speech, free practice of religion, and so on. We should celebrate and rejoice in cultural variety.

Second, the evidence suggests that democracy is reinforced by certain cultural conditions. It is by building democratic institutions that a more participatory culture evolves which in turn strengthens democracy. The direction of causality is mainly from social and political institutions to political values and practices. The strong evidence suggesting that the existence of democratic institutions is not associated with culturally-defined differences implies a clear signal: policy makers cannot refuse democracy by claiming their own culture. In other words they cannot use cultural traditions and characteristics as an excuse for failing to institutionalize participatory and democratic political structures of decision-making, including diverse voices and interests. The countries that weathered the Asian financial storm of the late 1990s best were democracies – Taiwan, the Philippines and Japan. South Korea and Thailand have also recovered; they have got rid of their corrupt former regimes by democratic processes. The history of Indonesia's development should put to rest the myth that democracy and human rights are Western concepts hostile to Asia and economic growth. A government that is not answerable to its people is not likely to have the institutions required to impose discipline to overcome a financial crisis or to embark on successful long-term development.

Third, if national and local cultural values are to be recognized in the management of the economy, it will be necessary for cultural policies to be integrated with economic and political policies. There is much room for experiment with creative forms of conflict resolution that foster equality instead of discrimination, and conviviality instead of violence. The growing risk of violent ethnic conflicts that arises from the arrogant and intolerant assertion of cultural, indigenous, ethnic, racial, religious, linguistic, or minority rights is evident in many places. Isolating linguistic or cultural groups and “drawing the borders in blood” reflects not only a misapprehension of the nature of culture and the evolution of history, but is doomed to fail in societies that are becoming increasingly interdependent and multicultural. Indigenous and culturally distinct groups – i.e. ethnic, racial, or religious sub-groups that demand rights to express and to continue to develop their cultures – deserve to be supported, but their relationships with the wider societies, nations, and world community in which they are embedded must also be recognized in thought and action. The cure for separatist, exclusive ethnicity is multicultural ethnicity.

Fourth, policy makers have to rethink state, community and international institutions and policies to permit local populations to choose their languages, allegiances and ways of life, provided that the implementation of these choices is taken up by the local or regional communities themselves. At the same time, institutions should be created that encourage a dialog between leaders of different cultural groups to negotiate exchanges and promote a better mutual understanding. Intercultural dialog becomes a prime line

of policy action that should be implemented according to local ways of management and organization. The important thing is that local communities and their administrative arrangements – municipal, provincial, state or departmental governments – must take on the main responsibility for this dialog and ensure that no artificial walls are erected to stop the flow of discourse between cultures. This is especially relevant for the young generations who are open to many cultures from which they take symbols, icons and customs that allow them to rejuvenate their traditions, thereby making them better adapted to the changing conditions of a global and shrunken world. Policies to encourage intercultural productions in the arts, especially by the young and by women, should be given priority. Women, where they are allowed to participate fully in social and cultural life, will contribute much to creating the new societies of the 21st century.

Fifth, cultural policy should look beyond a purely national emphasis and take in addition a broader international, interregional and global perspective. New partnerships between governments, corporations, private voluntary associations and other stakeholders should be developed. The positive impact of globalization on local creativity, by opening up markets, should be identified and encouraged. The effects, both good and bad, of global markets on local cultural industries should be more clearly recognized, so that policy can protect and enhance their cultural and economic flowering.

Finally, policies that allow expression and development of cultural potential will also have repercussions on how people relate to their physical environment. Achieving environmental sustainability calls for democratization of expertise and participation of local communities. Cities, the most multicultural sites of the future, should evolve ways in which people who speak different languages and have different loyalties can live together in peace. Urban squalor, violence and crime are not the result of urbanization but of inadequate incomes, unemployment, poor education, overcrowded housing, insecure tenure, homelessness and lack of social support. City governments and municipalities can encourage peaceful and prosperous urban communities by strengthening their support for local initiatives. Apart from the necessary social services in health, education, housing, water and sanitation, they can encourage new artistic expressions that enhance the dynamics of a local–global dialog. Artists everywhere wish to express their personal and cultural identities, and in doing so they create global trends. The new site for the creation of art is the global market place. Policy makers should ensure that artists are able to participate in the expanding global markets.

In conclusion, the analysis in this chapter shows that culture goes far beyond the field traditionally assigned to Ministries of Culture. Culture is indeed concerned with artistic creation and with ethnic and indigenous issues, but culture has also social and political dimensions. It is relevant in designing and implementing models of economic development, constructing stable democracies, ensuring that diverse cultures can live together without violent conflict or war, and providing a sense of trust, partnership and solidarity that are needed in any society in which people cooperate for their well-being. For this, an education for world citizenship is needed, a citizenship that is rooted in local culture, and is directed at patriotic cosmopolitans or cosmopolitan patriots, loyal to their families, neighbors, local communities, countries and humankind. For we undermine the



case for multicultural respect by failing to make central to education a broader respect for all human beings.

## References

- Barber, B.R. (1995). *Jihad vs. McWorld*. Ballantine Books, New York.
- Collier, P., Hoeffler, A. (2000). "Greed and grievance in civil war". Policy Research Working Paper 2355. World Bank, Washington, DC.
- Hirschman, A.O. (1995). "Social conflicts as pillars of democratic market societies". In: *A Propensity to Self-Subversion*. Harvard University Press, Cambridge.
- Huntingdon, S.P. (1996). *The Clash of Civilizations and the Remaking of World Order*. Simon & Schuster, New York.
- Myrdal, G. (1968). *Asian Drama*. Twentieth Century Fund, New York.
- Pattullo, P. (1996). *Last Resorts: The Cost of Tourism in the Caribbean*. Cassell, London.
- Sainath, P. (2002). *Everybody Loves a Good Drought: Stories from India's Poorest Districts*. Penguin India, New Delhi.
- Sen, A. (1999). *Development as Freedom*. Alfred Knopf, New York.
- Sen, A. (2004). "How does culture matter?" In: Rao, V., Walter, M. (Eds.), *Culture and Public Action*. Stanford University Press, Stanford.
- Streeten, P. (1953). Appendix to G. Myrdal, *The Political Element in the Development of Economic Theory*. Routledge and Kegan Paul, London, pp. 208–217.
- Streeten, P. (2001). *Globalisation: Opportunity or Threat?*. Copenhagen Business School Press, Copenhagen.
- Throsby, D. (2001). *Economics and Culture*. Cambridge University Press, Cambridge.
- UNESCO (2001). *Universal Declaration on Cultural Diversity*. UNESCO, Paris.
- United Nations Development Programme (2004). *Human Development Report 2004: Cultural Liberty in Today's Diverse World*. UNDP, New York.
- United Nations General Assembly (1949). *Universal Declaration of Human Rights*. UN Department of Public Information, Lake Success.
- World Bank (2001). "Attacking poverty". *World Development Report 2000/2001*. World Bank, Washington, DC.
- World Commission on Culture and Development (1995). *Our Creative Diversity*. UNESCO, Paris.

## EMPIRICAL STUDIES OF DEMAND FOR THE PERFORMING ARTS\*

BRUCE A. SEAMAN

*Georgia State University, USA*

### Contents

Abstract	416
Keywords	416
1. Introduction	417
2. What do we know about arts audiences?	419
2.1. Audience profiles	419
2.2. Age	420
2.3. Education and income	421
3. An overview of econometric performing arts demand studies	423
4. Price and income elasticities	424
4.1. Price elasticity differences by level of aggregation	424
4.2. Conceptual issues in interpreting the price elasticity results	428
4.3. Income elasticity differences by level of aggregation	431
5. The modeling problem: Single versus simultaneous equations	432
5.1. Moore: "The demand for Broadway theater tickets" (1966)	433
5.2. Withers: "Unbalanced growth and the demand for the performing arts: An econometric analysis" (1980)	434
5.3. Luksetich and Lange: "A simultaneous model of nonprofit symphony orchestra behavior" (1995)	437
6. The multicollinearity problem: Education versus income	439
7. The taste cultivation problem and human capital: Habit formation, learning-by-consuming and rational addiction	441
7.1. Habit formation	442
7.2. Learning by consuming	443
7.3. Rational addiction	447
8. The product and geographic market problem: Substitutes and complements	449

\* This chapter is an abbreviated version of the full review of the performing arts demand literature, Seaman (2005), appearing in the Nonprofit Studies Program Working Papers series (NP05-03) of the Andrew Young School of Policy Studies, Georgia State University. This paper can be found at <http://aysps.gsu.edu/nonprofit/working/nspwp0503.pdf>.

9. The product quality problem	453
9.1. Modeling quality	453
9.2. Studies using objective quality criteria	455
9.3. Studies using subjective quality assessments	457
9.4. An extended model	458
10. The role of socioeconomic factors versus life-style determinants of arts demand	459
10.1. Lifestyle determinants	459
10.2. Racial differences	461
10.3. Sexual orientation	462
11. Data problems	463
12. Conclusions	465
References	467

### Abstract

While audience and participation surveys, as well as econometric demand studies, generally confirm that performing arts audiences are relatively elite, there are surprises. Education (despite conflicting causal interpretations) is a stronger determinant than income, but that evidence is more reliable from survey results than from econometric estimation, and arts training is often distinguished from formal education. The arts as luxury goods can only be confirmed by those rare studies controlling for the value of time, and price elasticities are often higher than expected, especially when more disaggregated data are examined. Price inelastic demand is more likely the result of low pricing strategies of non-profit arts managements rather than any inherent result of an acquired taste for the arts, while cross-price elasticity evidence is relatively weak, even within the performing arts. Arts demand cannot adequately be estimated without also considering “life-style” variables, or non-standard socioeconomic factors such as sexual orientation, gender and socialization processes, and even the role of age has been notably complex. Quality of arts performance or organization seems important, but the econometric results are mixed. Habit formation must be distinguished from learning-by-consuming and rational addiction in examining dynamic determinants. Sociologists, psychologists, and marketing specialists, as well as economists, have contributed to this literature, which remains unusually enigmatic despite about forty years of increasingly sophisticated analysis.

### Keywords

demand, elasticity, performing arts, pricing, product quality, firm versus industry, habit formation, learning-by-consuming, rational addiction

*JEL classification:* C1, C2, C3, D12, D13, L82, Z1

## 1. Introduction

When readers of *La Scena Musicale* in December 2001 were informed that “. . . the likelihood of money being spent on orchestral music is linked to consumers’ increasing age, education and income” [Ehrensaft (2001, p. 1)] they could hardly have been shocked.<sup>1</sup> The “high arts” are widely viewed as the domain of a minority of elites, long an argument used by opponents of government arts subsidization to characterize such support as regressive, and ironically also by proponents who stress the need to make culture more accessible to the general public. Furthermore, one might suspect that this consensus and conventional wisdom would render efforts to conduct empirical studies of the demand for the arts relatively useless – at best, carefully designed confirmations of the obvious.

This chapter is designed to evaluate that suspicion, and finds surprising evidence of contradictory results, personally held convictions that are inconsistent with the empirical evidence, and significant popular misconceptions about the findings in some of the most cited empirical studies. Lévy-Garboua and Montmarquette, themselves significant contributors to this literature, reflect this best when they observe: “It is likely that the demand for the arts is price-elastic and art is a luxury good. But this prediction stems more, as yet, from a theoretical conjecture than from well-replicated empirical estimates” (2003, p. 211). They also suggest that we have not yet clarified whether arts goods have close substitutes, hence suggesting that we are still “groping towards firm answers” to three of the most basic empirical questions regarding arts demand (2003, p. 201).

Corning and Levy (2002, p. 218) observe that “studies of demand for the performing arts typically take one of two basic approaches: survey studies which seek to characterize the demographics of theater [and other] patrons, and econometric studies which seek to quantify demand and income elasticities”, although it is notable that data for econometric studies are often derived, at least in part, from either audience or arts participation surveys.<sup>2</sup>

Regarding econometric studies, while “income and price elasticities . . . are the usual end-products of empirical demand analysis” [Barten (1992, p. 21)], a substantial portion of the performing arts demand literature does not derive such elasticities. For example, only 29 of the 44 regression-based studies cited in this chapter report some kind of demand elasticities; of these only 19 estimate both own-price and income elasticities, and fewer still also estimate any cross-price elasticities.<sup>3</sup> Thus, a notable part of this litera-

<sup>1</sup> Although they might have been surprised by the optimistic tone of the article, which cited demographer David Foote as demonstrating that “aging baby boomers” and the “graying of classical music audiences” will be a “valuable asset” that will ultimately lead to an increase in the classical music market [Ehrensaft (2001, p. 1)].

<sup>2</sup> For a good roundtable discussion of the difficulties in accurately conducting such surveys, see Horowitz (1985). Audience surveys are typically based on distributing questionnaires to audiences and collecting them upon departure, while participation surveys are designed to randomly sample the broader population, not limited to those who have been “self-selected” as part of an arts audiences.

<sup>3</sup> See further in Table 1.

ture is devoted instead to more broadly examining the competing determinants of arts attendance or participation patterns without any formal link to the neoclassical theory of consumer behavior and its related concerns with formal homogeneity or aggregation constraints.<sup>4</sup>

This chapter is a much shortened adaptation of Seaman (2005), focusing on the econometric literature and also briefly summarizing the results of the survey studies. The arts are defined primarily as the non-profit performing arts (orchestral and chamber music, opera, ballet and modern dance, and theater, but also including for-profit Broadway), although comparisons are made with museums and the largely for-profit media arts and other forms of recreation and entertainment, including sports.

The organization is as follows. Section 2 briefly describes performing arts audiences based on participation and audience surveys. Section 3 provides an overview of the econometric literature, while Section 4 reports on the estimation of price and income elasticities, focusing on the effects that different levels of aggregation and audience segmentation have on the empirical results. The technical challenges that have faced researchers in conducting empirical arts demand studies are addressed in Sections 5–9, while Section 10 evaluates the view held by some that “life-style” and various socialization measures are more important determinants of variation in arts consumption behavior than are the traditional socioeconomic determinants of age, income, education and occupation. “Mixed” factors such as gender, race and ethnicity, religious affiliation, and sexual orientation combine with other variables to complicate that analysis. Section 11 summarizes the data problems that have plagued economists in addressing the issues discussed in the previous sections. A concluding summary and evaluation (Section 12) focuses on an assessment of the claim that three main developments are required before more definitive answers can be given to questions about the demand for the arts:

- (1) more careful econometric work;
- (2) the increased use of large data sets; and
- (3) the “more intensive use of explicit models of the cultivation of taste” [Lévy-Garboua and Montmarquette (2003, p. 211)].

<sup>4</sup> Even studies that estimate demand elasticities are faced with sufficient data and econometric challenges such that few can afford the luxury of ensuring that estimated demand functions are homogeneous of degree zero while also meeting Engel and Cournot aggregation conditions. Pommerehne and Kirchgassner (1987) is a rare example of using almost-ideal-demand-system restrictions to estimate expenditure shares for cinema, theater and a composite good. The general absence of such restrictions can indeed complicate the interpretation of the results, especially regarding the controversial issue of price elasticities. However, many of the conclusions regarding other important issues are relatively unaffected by statistical technique (e.g., discriminant, cluster, factor or multivariate regressions) or by field of specialization (e.g., economics, sociology, arts policy, psychology, or marketing).

## 2. What do we know about arts audiences?

### 2.1. Audience profiles

Two of the earliest empirical observations in arts economics are, first, that performing arts audiences are elite in terms of income, education and profession and hence non-representative of the more general population, and second, there are only trivial differences in those audience characteristics across the various performing arts forms.<sup>5</sup> In fact, that “audiences from art form to art form are *very* similar” was viewed by Baumol and Bowen as “the most remarkable finding” of their path-breaking efforts to assemble credible data on arts consumption patterns, primarily in the United States (1966, p. 84).

Throsby and Withers (1979), evaluating 1976 Australian data, found essential similarity between Australian and American audiences, and also cited British and Canadian data. A more recent Canadian survey of theater patrons in Montreal found ongoing evidence of this elitism in that 54 percent were university graduates, 45 percent earned more than \$40,000 per year, and only 11 percent were employed in primary (manufacturing or construction) industries [Colbert and Nantel (1989)]. The most detailed ongoing source of survey data on performing arts audiences is the *Survey of Public Participation in the Arts* (SPPA), periodically published by the US National Endowment for the Arts. The most recent version applies to 2002 [National Endowment for the Arts (2004)], and continues to generally support the arts elitism hypothesis. These data have also served as the foundation for regression-based studies such as Peterson, Hull and Kern (2000) and Gray (2003).

Actually, the commonality of arts consumption patterns across many different countries, educational systems and cultures was not a universally anticipated result, as Baumol and Bowen discovered when they were told by British colleagues to anticipate much more egalitarian results in extending their survey to Great Britain [Baumol and Bowen (1966, p. 89)]. However, except for a slightly higher representation of lower middle-income groups in British audiences (p. 93), Baumol and Bowen found “remarkable” similarity in the British and American results (p. 89). Interestingly, Cwi (1985) attacked the Baumol and Bowen (1966) conclusion about arts audience elitism as fostering a “welfare economics mentality towards arts policy” that primarily serves the interests of a political agenda to justify government subsidies to bring “arts to the people” (p. 32). Cwi further argued (1985) that even if the basic audience profile were to remain relatively constant, substantial societal changes in education and occupational choice would progressively make that profile more reflective of the general population, and that the elitism of arts audiences had always been in part the result of an overly narrow definition of the arts. The evidence regarding whether the arts are becoming less elitist is decidedly mixed. For example, O’Hagan (1996) found absolutely no evidence

<sup>5</sup> Baumol and Bowen (1966); Ford Foundation (1974); Book and Globerman (1975); National Research Center of the Arts, Inc. (1976); DiMaggio and Useem (1978); Throsby and Withers (1979); West (1985).

in Irish, British and American data of any change in the access of the performing arts to a wider audience, while Heilbrun (1996), using a different measurement standard, found notable improvements in the accessibility of the arts to the general population between 1980 and 1990 in the United States.<sup>6</sup>

## 2.2. Age

Age represented the most unexpected results in the Baumol and Bowen (1966) audience survey, with those aged 20–24 the most over-represented in both American and British arts audience relative to the size of that age group in the general population, when attendance is imperfectly defined as “having attended at least one performance within the past year”. These “relative frequencies” decline systematically with age, indicating that performing arts audiences were dramatically younger than the general urban populations in both countries in the mid-1960s. However, when frequency of annual attendance is considered, the role of age in arts audiences changes. For those attending more than 10 times per year, 7.1 percent were over 60 years old compared to 2.4 percent under age 20 for Broadway Theater, with the “older age gap” a very high 17.9 percent vs. 3.9 percent for Major Orchestras, and a more moderate 7.0 percent to 5.8 percent for Regional Theater [Baumol and Bowen (1966, Appendix, Table IV-I)]. The most recent SPPA [National Endowment for the Arts (2004)] also reflects an ongoing aging of arts audiences, even though managements of many arts organizations have objected to this conclusion [Peterson, Hull and Kern (2000, p. 1)].

Japanese and German data from the early 1980s and mid-1990s reflect the surprising complexity of the role of age in performing arts audiences. Although the context of analyzing Japanese audiences regarding western classical music may influence the comparability of such results with western audiences, the Kurabayashi and Ito (1992) survey results are striking. With the sole exception of a notable trend toward older audiences for the NHK Symphony Orchestra between 1977 and 1981, these results indicate a remarkable bias toward younger audiences, especially for the Osaka Philharmonic, where fully 68.9 percent of females in the audience were younger than 30. In fact, more than 50 percent of each gender was younger than 30 for all non-NHK cases except for males attending the Tokyo Philharmonic and the Sapporo Symphony, and even in those cases the young group outweighed the older group.

Would anything approximating this youth bias be found in countries at the heart of western classical music such as Germany, where classical music is typically referred to as *ernste* or “E-Musik”, in contrast to less culturally “rich” popular or “U-Musik”? Wiesand (1995, Table 2) provides evidence of notable differences in the propensity of different age groups to consume four different types of concerts (music theater, E-concerts, U-concerts, and rock/jazz), but except for the strong youth bias for rock/jazz music, his findings show relatively similar consumption patterns among the youngest

<sup>6</sup> See Seaman (2005, Section 2.3).

age group (18–24) across the three other music types. The next youngest group (25–34) shows more variation, but has the highest propensity to experience “E-Konzerte”. The Wiesand data confirms the expected result that the youngest German age group has the highest overall participation rate in attending concerts and that the most “high-brow” music (E-concerts) is the least popular. The youngest group (18–24) is also notable for having relatively similar participation rates for all three non-Rock and Jazz music types (although lowest for the classical type E-music). The most significant result is that the second youngest age group (25–34) has the highest classical music participation rate, and the combined participation rates of the youngest groups (younger than 35) are higher than the comparable rates for the two oldest age groups (older than 50). Thus, the German evidence is also consistent with the earlier results that, at least when frequency of attendance is ignored, the performing arts should not automatically be thought of as dominated by older age groups.<sup>7</sup>

### 2.3. *Education and income*

Arts survey studies such as the [National Research Center of the Arts, Inc. \(1976\)](#), which served as the key data source for [DiMaggio and Useem \(1978\)](#) and the [Ford Foundation \(1974, Vol. II\)](#), have also contributed substantially to our understanding of the relative roles of education and income in determining performing arts attendance. While the positive causal relationship between education and income has plagued econometric efforts to separate their independent effects,<sup>8</sup> the early non-econometric literature was replete with evidence that the role of education was much stronger than that of income.<sup>9</sup> In fact, [Heilbrun and Gray \(2001\)](#) identify the [Ford Foundation \(1974, Vol. II\)](#) study as important evidence of the relative effects of education versus income, but also citing [Gray \(1998\)](#) as providing multivariate regression evidence supportive of a larger role for education than income, based on an analysis of 1997 SPPA data (Tables A17, A20 and A21).<sup>10</sup> The later National Endowment for the Arts 2002 SPPA survey [National

<sup>7</sup> This point is also consistent with the [West \(1985\)](#) Ontario, Canada audience survey finding that the same percentage of arts audience (20.1 percent) were 20–30 years old as were 40–50 years old, and those older than 50 constituted only a trivially higher 20.9 percent of audiences. The dominant age group was 30–40 (29.6 percent). He did confirm, however, the frequent finding that the under 20 age group was dramatically under-represented at only 2.6 percent of Ontario audiences in 1984–1985.

<sup>8</sup> See further in Section 6.

<sup>9</sup> [Globerman \(1989\)](#) also cites other US data from the [Association of College, University and Community Arts Administrators, Inc. \(1984–1985\)](#) as providing at least “suggestive” evidence that education is more important than income and occupation in determining arts attendance.

<sup>10</sup> Researchers who are selectively familiar with only the econometric literature seem especially prone to concluding that there is no coherent evidence of the separate roles played by these two strongly positively correlated variables. This conclusion has no doubt been reinforced by the fact that two of the best early econometric studies that did indeed confirm a relatively weak role for income, did not include education as a separate independent variable in their equations [[Moore \(1966\)](#); [Withers \(1980\)](#); with additional results also reported in [Throsby and Withers \(1979\)](#)], and the fact that an early study that did include both independent



Endowment for the Arts (2004)] continued to report that education, “more than any other demographic factor” is highly correlated with attendance at arts events and museums (p. 19). However, of more relevance to the issue of whether regression analysis has confirmed this result is the NEA sponsored study by Peterson, Hull and Kern (2000), which generally (but not universally) reported that education is the strongest predictor of arts attendance using data from the 1997 SPPA survey and basic OLS estimation.

DiMaggio and Useem (1978, Table 1) stressed the so-called education gap and the income gap in the self-reported “exposure” of various types of individuals to seven different “cultural forms”. For example, an education gap of 55 (percent) was reported for exposure to theater because the exposure rate of the most educated group (college graduates) was 73 percent while the exposure rate for the least educated group (< high school graduate) was only 18 percent (an absolute difference of 55 percent). Each “consumption gap” entry reported by DiMaggio and Useem (1978) reflected this absolute difference between the exposure rates of the most versus least educated, or the highest versus lowest income group. The most critical feature of these consumption gap data was the sizeable education gap for all seven cultural forms, even including popular music. The three high arts forms showed the greatest education gap, a result not duplicated by the income gap, which displayed the unexpected pattern of being lowest for arguably the most elitist art form (classical music), and almost as high for the more popularized cinema as it was for art museums and theaters.

However, the Ford Foundation study (especially Table 15) has been the most influential in confirming that “to a startling degree . . . it is indeed education rather than income that matters most” [Ford Foundation (1974, II, p. 16)]. For example, using theater attendance as an example, when income was held constant at either a high or low level, differences in education generated either a 21 percent differential in attendance rates (for high income), or a 25 percent differential in attendance (for low income). However, when this was reversed, and education was held constant at either a high or a low level, differences in income had much smaller effects – an attendance rate differential of only 8 percent for those with high education, and a 12 percent differential for those with low education. This pattern was also evident for symphony, opera and ballet. While much weaker, this apparent relative potency of educational differences in affecting attendance rates even extended to popular Broadway musicals, and the more contemporary music forms of jazz, rock and folk. By this measure, education only failed as the stronger factor compared to income in the case of movies, where their average “explanatory power” was equal.

We return to the question of education vs. income in our discussion of the results of econometric studies in Section 6 below.<sup>11</sup>

variables [Gruenberg (1975); see text below] was unpublished. Furthermore, similar to the case with Gray (1998), three other studies documenting some econometric support for the strength of education over income are relatively unknown [Gapinski (1981); Goudriaan and de Kam (1983); Ganzeboom (1989)].

<sup>11</sup> Other issues regarding arts audiences are discussed in Seaman (2005), including: the question of overlapping performing arts audiences (“co-patronage”); the debate regarding possible arts booms and how such

### 3. An overview of econometric performing arts demand studies

Since 1966 there have been at least 44 econometric studies of the demand for (or consumer participation in) the performing arts.<sup>12</sup> A few of these studies also included specific numerical estimates of other demand elasticities: cross price [Withers (1980); Throsby and Withers (1979); Touchstone (1980); Gapinski (1986); Bonato, Gagliardi and Gorelli (1990)]; leisure price [Withers (1980); Throsby and Withers (1979)]; donor price [Lange and Luksetich (1984)]; tourism attendance [Gapinski (1988)]; education [Globerman and Book (1977); Gapinski (1981)]; advertising [Luksetich and Lange (1995)]; “number of shows” [Moore (1966)]; and even “unpopularity of conductor” [Greckel and Felton (1987)]. Other studies derived coefficient estimates that were not translatable into elasticities, and/or evaluated a large number of additional independent variables, many of which lacked either economic or statistical significance.

Basic linear ordinary least squares (OLS), especially using the double-log form, has been the most popular primary estimation technique (used in 18 studies), but other related techniques have been used including: step-wise OLS [Globerman and Book (1977); Greckel and Felton (1987); Andreassen and Belk (1980)]; double-log weighted OLS [Felton (1992)]; two-stage least squares [Moore (1966), Lange and Luksetich (1984), Luksetich and Lange (1995), Jenkins and Austen-Smith (1987)]; conditional maximum likelihood estimation [Corning and Levy (2002)]; the almost ideal demand system [Pommerehne and Kirchgassner (1987)], Clawson–Knetsch distance modeling [Forrest, Grimes and Woods (2000)]; non-parametric linear regression [Schimmelpfennig (1997)]; and logit, tobit, or probit non-parametric estimation (eight studies).<sup>13</sup>

Of the 44 econometric studies, 22 relied upon US data. Approximately two-thirds of all studies reported time-series results (including some pooling of time series and cross-section data), but since a few of those studies also included separate cross-section analysis [Moore (1966); Goudriaan and de Kam (1983); Luksetich and Lange (1995)] about 42 percent of all studies involved cross-section estimation. Only nine demand

dynamic changes might affect the degree of audience elitism [DiMaggio and Mukhtar (2004)]; and the emerging marketing literature regarding “omnivores” (i.e., those whose music and leisure consumption is so broad and eclectic as to defy the label elitist) and “univores” (by contrast, persons with decidedly narrower favorites among music and other leisure options) as considered by Peterson (1992), Bryson (1997), van Eijck (2000), Fisher and Preece (2002, 2003), and Lopéz Sintas and García Álvarez (2004).

<sup>12</sup> Seaman (2005) further clarifies which types of studies are omitted. For example, highly specialized marketing forecasting models such as Weinberg and Shachmut (1978), whose “ARTS PLAN” model inspired some similar efforts to predict attendance at specific (usually university) performing arts events are omitted. See also Weinberg (1986) and Putler and Lele (2003).

<sup>13</sup> This choice of approaches is in no way at odds with standard practice in empirical economics. DiNardo and Tobias (2001) begin their overview of non-parametric techniques by observing: “Even a cursory look at the empirical literature in most fields of economics reveals that a majority of applications use simple parametric approaches such as ordinary least squares or two-stage least squares accompanied by simple descriptive statistics” (p. 11).

studies constituted essentially a case study of one or two specific arts firms, with the rest involving some degree of aggregation among many organizations, with seven reporting results for some version of the aggregate performing arts [although three of these also reported less aggregated results for separate art forms: Houthakker and Taylor (1970), Throsby and Withers (1979), Pommerehne and Kirchgassner (1987)]. The most popular separate art form aggregation has been theater (including a few studies of the for-profit Broadway theater), identified in 20 studies, followed by symphony orchestras, which were the primary or secondary focus of 16 studies. Separate results were reported for aggregated opera companies eight times and for dance/ballet companies seven times.

Owing to the important role played by the concepts of own price and income elasticity of demand in economic analysis generally and in discussions of the demand for the arts in particular, the next section focuses on those studies that have estimated such elasticities.

#### 4. Price and income elasticities

Table 1 more fully documents the 29 studies that have reported either own price or income elasticities, or both. Note that the table omits any regression study that does not derive elasticities, even if it includes income and/or price as variables.

A review of Table 1 confirms that the Lévy-Garboua and Montmarquette suspicion that the arts really are luxury goods with own-price elastic demands (2003, p. 211) has not yet been justified by the econometric evidence. Regarding estimates of the own-price elasticity of demand, 12 studies found that the demand for the arts is price inelastic while only four found strong evidence of price elastic demand. Krebs and Pommerehne (1995) reported low short-run but high long-run price elasticity. However, five other studies found mixed results for the price elasticity of demand, especially when data allowed a more disaggregated analysis of different price ranges, audience characteristics, or type and sizes of individual arts organizations.

Let us consider the differences among these results in more detail.

##### 4.1. Price elasticity differences by level of aggregation

In interpreting the price elasticity results reported in Table 1, we can observe that, regardless of technical sophistication, the price *inelasticity* result is much more prominent in those studies that used very aggregative data across all performing arts groups in contrast to studying individual arts organizations,<sup>14</sup> and/or that used a measure of ticket price (such as total revenue divided by attendance) that does not measure the actual

<sup>14</sup> Of course, this aggregation problem is hardly confined to the arts. The common estimation of supply and demand functions “using uniform prices and quantities across products, yielding a single industry-wide demand elasticity estimate” is criticized in a study of the personal computer market as especially misleading when firms produce differentiated rather than homogeneous goods, since “each product is likely to face

Table 1  
Summary of performing arts own price and income elasticity estimates: sorted by year

Study	Price	Income	Study	Price	Income
Moore (1966)	-0.33 to	0.35 to 0.43;	Carson and	Prior month	-4.74 to 5.78
Moore (1968)	-0.63	1.03 cross-sec.	Mobilia (1989)	-0.38	seasonally
Houthakker and	-0.18 short;	0.74 short;	Bonato, Gagliardi	-0.38	0.78
Taylor (1970)	-0.31 long	1.26 long run	and Gorelli (1990)		
Globerman and	None	0.76 to 1.07 by	Throsby (1990)	-0.41 (not sig.)	None
Book (1977)		art form			
Withers (1980)	-0.90 to	0.64 to 1.55	Oteri and	Not	Not
Throsby and	-1.19; but	conventional	Trimarchi (1990)	statistically	statistically
Withers (1979)	-0.62 to	1.43 to 2.78		significant	significant
USA data	-0.67 adjust.	adjusted			
Throsby and	-0.62 to -1.0	Not significant	Felton (1992)	-0.13 to -0.95	0.77 to 3.09
Withers (1979)	conventional	in either model		by art	varies by art
Australia data	-0.61 to			form and size	form and size
	-1.17 adjust.				
Touchstone	Imputed:	None	Abbé-Decarroux	-0.31 full	None
(1980)	-0.09 to		(1994)	price (insig.);	
	-0.13 by art			-2.45 low	
	form			price not sig.	
				diff. -1.0	
Gapinski (1981)	None	0.36	Felton (1994/1995)	-0.85 total	1.40 total
				-0.24 subscrp.	0.82 subscrp.
Goudriaan and	None	0.10 to 1.02 by	Luksetich and	-0.16 to -0.42	Not significant
de Kam (1983)		art form	Lange (1995)	by orch. size	
Gapinski (1984)	-0.66	1.33	Krebs and Pom-	-0.16 short;	0.1 not stat.
			merehne (1995)	-2.6 long	sig.
Lange and	-0.49 to	None	Lévy-Garboua	-1.00 to -1.47	None; wealth
Luksetich (1984)	-1.26 by orch.		and Montmarquette	by experience	proxies +
	size		(1996)		
Gapinski (1986)	-0.07 to	0.06 to 0.27 by	Schimmelpfennig	-1.34 to -5.56	None
	-0.29 by art	art form	(1997)	by ballet, seats	
	form				
Pommerehne and	-1.22 to	1.50 (not sig.)	Ekelund and	Inelastic (no	0.78 normalize
Kirchgassner	-1.65 by	to 2.44 by	Ritenour (1999)	numerical)	
(1987)	income	income			
Jenkins and	+1.1 to +2.5	0.26 to 0.54	Forrest, Grimes	-1.24 point	None
Austen-Smith		insignificant	and Woods (2000)	-1.11 arc	
(1987)					
Greckel and	-0.34 to	2.26 insig. to	Coming and Levy	-0.05 to -4.87	1 of 3 > 1.0;
Felton (1987)	-2.33 insig.	6.13 by org.	(2002)	by venue	3 > 0 but only
	by art org.				2 sig.
Felton (1989)	-0.64 to	None			
	-1.62 opera	significant			

Note: "None" indicates that no elasticities were estimated, due largely to an absence of data.

prices paid by different types of consumers. Notably, [Jenkins and Austen-Smith \(1987\)](#), one of a few studies to find statistically significant but positive own-price elasticities, explained the paradox of that finding in part by suggesting that price is serving as a proxy for quality, but also implicating their overly aggregated measure of price (i.e. an average of total box office revenues over the entire season divided by total season attendance; p. 170).<sup>15</sup>

Given the expected importance of having more disaggregated and targeted price data, what can be said about the results of studies that segmented the audience more carefully, or that focused on less aggregated measures of the performing arts? Caution is essential in making any generalizations about this literature. For example, [Moore \(1966\)](#) found consistently low own-price elasticities for Broadway theater tickets even when studying a relatively disaggregated segment of the performing arts (aggregated to be sure across seven Broadway houses and 18 performances, but at least not focusing on the overall performing arts). His somewhat unconventional use of a list price (i.e. “an average of the cost of the most expensive seats for a regular performance of each production”; p. 83) to proxy average price paid can be criticized, but may not have seriously biased his time-series results. And [Gapinski \(1986\)](#) is widely lauded for deriving separate theater, opera, symphony and dance estimates of cross-price as well as own-price elasticities using quite disaggregated data specific to thirteen individual arts companies in London. Yet he found generally low own-price elasticities for each of the individual organizations (varying from  $-0.05$  to  $-0.70$ ). However, in constructing his nominal price variables Gapinski was forced to resort to dividing attendance into box office revenues (including value-added tax), hence failing to fully capture actual transaction prices or to differentiate among consumer groups.

Less aggregated studies generally segmented audiences in various important ways that allow for a more precise examination of pricing, as well as other likely demand determining characteristics. [Pommerehne and Kirchgassner \(1987\)](#) segmented by income of consumers, with price elasticity lower for high-income than for average-income consumers (but with both being greater than one in absolute value); they also found uncommonly low own-price elasticities for cinema. Seating section was the key for [Schimmelpennig’s ballet study \(1997\)](#), which found generally elastic demand for Orchestra, Grand Tier, and Rear Amphitheater sections, with surprisingly high price elasticities even for the more expensive Orchestra seats that supposedly serve higher-income patrons. By contrast, the [Abbé-Decarroux \(1994\)](#) results for a Geneva theater were more expected, with a finding of a high price elasticity of  $-2.45$  for his “reduced price” consumer group (although not clearly statistically different from 1.0 at the 0.05 level), but inelastic (and not statistically significant) price elasticity for the “full price” group. He

a different demand elasticity”, ideally requiring a focus on “individual products’ attributes and their market position” in estimating demand elasticity [[Stavins \(1997, pp. 347–348\)](#)]. For a comprehensive review of heterogeneity and aggregation problems in economics, see [Blundell and Stoker \(2005\)](#), with applications to demand modeling (pp. 350–364).

<sup>15</sup> [Seaman \(2005\)](#) provides a detailed analysis of the various ways to measure price (Part 3.1.1).

was also unique in explicitly arguing that such results “weaken” the conclusions of other major studies that use an overly aggregated average price to derive price-inelastic results for the performing arts (p. 105).<sup>16</sup>

Felton’s (1994/1995) study of 25 large US orchestras found lower price elasticity for subscribers ( $-0.24$ ) compared to price elasticity for the combined “total attendance” (although still less than unity at  $-0.85$ ). However, that result contrasted with her pooled time-series study of 13 opera companies [Felton (1989)], where subscribers appeared more responsive to ticket-price changes than single-ticket purchasers (i.e. her only statistically significant results were for subscribers, although the magnitude of the price elasticities varied widely across organizations). Even though subscriber reactions to ticket price changes differed somewhat between her 1994/1995 orchestra and 1989 opera samples, she decided to limit her 1992 study of orchestra, opera and ballet companies to subscriber demand, basing this decision on her conclusion that her previous work with opera data had revealed “that season subscribers do react to ticket price changes while non-subscribers do not” [Felton (1992, p. 2)].

Lévy-Garboua and Montmarquette (1996) imputed rather than directly estimated price elasticities (see Section 7 below) but derived the surprising result that more experienced French theater-goers actually had higher price elasticity of demand ( $-1.47$ ) compared to those less experienced (price elasticity close to unity), inconsistent with the view that price elasticity will be low for an acquired taste like the arts.<sup>17</sup> Since the more educated typically have more arts experience, Forrest, Grimes and Woods (2000) found contrary strong evidence for price inelastic demand in those regional zones with particularly high educational levels and higher price elasticities elsewhere.<sup>18</sup>

Segmenting by organization location has also generated differing own-price elasticities, with Corning and Levy (2002) finding that a three-location Southern California theater group faced price elasticities varying from inelastic to elastic depending on the specific venue. Lange and Luksetich (1984) found higher own-price elasticities for smaller orchestras than for large budget major orchestras, but their 1995 simultaneous

<sup>16</sup> Throsby (1990) found supporting evidence in his equations estimating the consumer “valuation” of a play. These utility function estimations revealed that the strongest influence of price was on the mainly young and less affluent audience of one of his Sydney theaters, whereas his least price-sensitive group was the audience of the most conservative of his theaters (p. 79). Huntington (1991) showed a kinked demand curve based on different price points. Kirchberg (1988) found that low-income groups regarded museum entrance fees as a barrier five times as often as do those in higher income groups with education, occupation, and “lifestyle” variables further broadening this gap.

<sup>17</sup> Doubtless the most novel interpretation linked to this finding is that of Köster and Marco-Serrano (2000, p. 8, Footnote 7), who cite Lévy-Garboua and Montmarquette (1996) for the finding that “the satisfaction degree is bigger among the occasional attenders than in the frequent attenders”, suggesting to them that the satisfaction of the occasional attenders stems more from the “sensation of having completed a duty” than from any direct arts consumption “sensorial rewards”.

<sup>18</sup> Ulibarri (2005) lends further theoretical support to the idea that more experienced arts consumers should have lower price elasticities in his application of an adaptive utility choice model the arts, lending further support to the notion that markets for arts goods will be segmented (p. 140).

equation orchestra model (where they also incorporated interactions with donations) found generally low price elasticities regardless of organization size (see Section 5 below).

Thus, while it can be said that own-price elasticities estimated with more disaggregated data are more likely to be greater than one in absolute value (whereas with aggregated data they almost never are), the extent and significance of those findings have sometimes been overstated. For example, Felton (1992) is regularly cited for finding low price elasticities of demand for aggregated groups of arts organizations, but significantly higher price elasticities for individual arts organizations within those groups [e.g., Throsby (1994, p. 8)]; however, she actually found price elastic demand in only 21 percent of her orchestra sample, 7 percent of her ballet sample, and 16 percent of her opera sample. Furthermore, while that paper is sometimes portrayed as distinguishing between industry and firm price elasticities, her pooled data did not include multiple companies in any one artform in any one city,<sup>19</sup> and in some cases did not even have more than one company of any artform in any one city. So Felton (1992) is not really a study well-designed to distinguish market or industry price elasticities from firm price elasticities. Furthermore, Felton's earlier study of individual opera companies (1989) reported only three opera companies with statistically significant subscriber own-price elasticities, with one being elastic (San Francisco), one unity (San Diego), and one inelastic (Houston). So there is no unambiguous finding of price elasticity greater than one even when the focus is on individual organizations.

#### 4.2. *Conceptual issues in interpreting the price elasticity results*

Regardless of how the empirical price elasticity results vary by level of aggregation or with the sophistication and “accuracy” of econometric technique, it is surprising how little attention has been paid to interpreting these results in light of economic theory. For example, little focus has been directed to the standard argument that non-profit arts organizations often charge lower than revenue-maximizing ticket prices.<sup>20</sup> In a related literature, it has been explicitly argued that profit-maximizing sports teams strategically under-price tickets; such non-revenue-maximizing behavior has become a standard explanation for the low price elasticities frequently found in empirical studies of sports

<sup>19</sup> Admittedly not possible to do if there is only one local professional opera company or even orchestra; Gapinski (1986) is a notable exception.

<sup>20</sup> Of course, the convenient and very commonly used double-log linear equation specification generates constant elasticities that do not vary with price, in contrast to a cubic log equation that can generate price elasticities that vary with prices [for an example in sports, see García and Rodríguez (2002)]. If the range and level of available performing arts price data is “artificially low” due to either non-optimal pricing, or systematic efforts to make the arts “more accessible”, perhaps not just as a public service but as part of a longer run strategy of encouraging people (especially the young) to develop the kind of human capital that can lead to various forms of consumption addiction, we would naturally expect to find relatively low estimated constant price elasticities.



demand.<sup>21</sup> In fact, sports economists have generally refused to accept any empirical evidence of price-inelastic demand.<sup>22</sup> Even when there is dissent, the issue of the price level at which price elasticity is estimated remains critical. For example, Coates and Harrison (2005) express amazement at their finding that the demand for baseball attendance in the US is strongly price inelastic. They argue that the significant local market power of all baseball franchises should induce them to “operate on the elastic portion of the demand curve”, and consider the quest for an explanation for why teams are operating in the inelastic portion of their demand curve “an important question for future research” (p. 298).

By stark contrast, the most common reaction among arts economists has been to find econometric evidence for low price elasticities to be consistent with theoretical expectations and non-econometric survey evidence.<sup>23</sup> For example, Throsby (1994) repeats his earlier distinction between “immediately accessible” popular entertainments and the higher arts that reflect an acquired taste, in defending his view that the arts will have lower price elasticities among established consumers “for whom qualitative characteristics of performances are likely to be decisive” (pp. 3 and 7–8).<sup>24</sup> While this argument seems consistent with a view that demand for the lively arts is “inherently” price insensitive, it is important to note the Throsby reference to “established” consumers rather than all consumers.

In defense of an “inherently” low price elasticity of demand for the arts, it is universally recognized that ticket price is only one component of the explicit expense of attending a live performance, and an even smaller share of the total expense when the implicit opportunity cost of time is included in the “full price”.<sup>25</sup> Furthermore,

<sup>21</sup> See, for example, Marburger (1997), Fort (2004); see Seaman (2005, Part 3.1.2) for a detailed review.

<sup>22</sup> In a study of the Spanish Football League García and Rodríguez (2002) found confirmation of low price elasticities of demand for all league teams when using a linear model and not adjusting for the possible endogeneity of price, which is inconsistent with “clubs acting as profit maximizers and costs not depending on attendance in a standard monopolistic model” (p. 28). However, they found good evidence for their contention that econometric specification issues should play a bigger role in estimating sports demand; their cubic specification using instrumental variables to correct for price endogeneity yielded 11 teams out of 27 with average price elasticities above one in absolute value, although for all teams the null hypothesis of unitary elasticity could not be rejected.

<sup>23</sup> Examples include Globerman and Book (1977), Ryans and Weinberg (1978), Kolb (1997, 2002), and Scheff (1999).

<sup>24</sup> The presumed greater availability of effective substitutes for more accessible forms of entertainment compared to the more “esoteric” forms of the higher arts would certainly be a key factor that would suggest that demand for the performing arts would be less price elastic than the demand for, say, cinema at comparable prices or when evaluated at competitive equilibria with limited market power. The necessity of distinguishing the price elasticity of demand at *observed* prices from the price elasticity at *competitive* prices is an iconic feature of antitrust economics.

<sup>25</sup> In fact, Globerman (1989) would extend this point to popular culture as well, noting that the admittedly less consistent evidence is that demand is price inelastic for those arts as well, consistent with the idea that the opportunity cost of time is usually the largest cost of attending any live events (his Note 14). However, it is interesting that Cameron (1990) extending his similar earlier results, has found relatively high price



viewing consumers as producers would distinguish between purchasing a ticket for a performance (a market good serving as a productive input) and actually consuming the individually-produced “art appreciation”. Once that fundamental distinction is made, there is no theoretical inconsistency between finding a “shadow price-elastic demand for art appreciation” and a “market price-inelastic demand for art consumption” as revealed in ticket price data.<sup>26</sup>

Another aspect of the full price of arts consumption that can reduce the economic significance of the explicit admission price is the interaction between admission expenses and voluntary donations. The potential for price discrimination strategies in the arts is further enhanced by their non-profit status that encourages donations. As argued by Lange and Luksetich (1984), the total price of attending a symphony concert includes both the price of the ticket and any contributions of that patron to the orchestra.

Finally, it is useful to remember the textbook list of factors determining price elasticity variations across different products (always measured for the industry or market, and not for the individual firm or consumer):

- substitution possibilities;
- budget share;
- direction of income effect; and
- time.

Frank (2006) presents a comparison of empirically estimated price elasticities of demand for seven very aggregated product groupings ranging from “green peas” (elasticity of  $-2.8$ ) to “theater, opera” [ $-0.18$ , citing the short-run price elasticity from Houthakker and Taylor (1970)]. In explaining why the price elasticity of the demand for green peas is more than 14 times larger than for theater and opera performances, he cites two factors: first, the likely small real income effect that would accompany any change in price for arts consumers who are expected to have much larger than average incomes, and second, the many more close substitutes for green peas than there are for theater and opera performances [Frank (2006, p. 128)]. This explanation is consistent with the view that there is something inherent in the performing arts that would yield very low price elasticities, i.e. limited substitutes. Yet again, any such explanation ignores the question of “substitutes at what price” and the issue of localized competitive conditions [Seaman (2004)]. It also fails even to consider why such firms (assuming some market power) would choose to operate so far from the seemingly rational average price level (i.e., the one approximating unit price elasticity when marginal costs approach zero).

In summary, estimated arts price inelasticity may simply reflect pricing in the inelastic range of demand curves or, when using excessively aggregated data, the failure

elasticities of demand for cinema in the United Kingdom (in the range of  $-1.53$  to  $-1.6$ ) using pooled cross-sectional and time series data, but when later explicitly testing for rational addiction in the demand for cinema (1999), he found neither strong support for rational addiction, nor statistical significance in his price variable. Fernández-Blanco and Baños Pino (1997) estimated an even higher long-run price elasticity of demand for cinema in Spain ( $-3.51$ ), but did not address the issue of price levels.

<sup>26</sup> See Lévy-Garboua and Montmarquette (1996, p. 206) and further discussion in Section 7 below.

to capture the range of differing arts consumer segments. At the same time, theoretical clarifications as to the full price of arts consumption, either by distinguishing arts attendance from arts appreciation or by clarifying the possibly low weight that ticket price alone has relative to the full cost, could suggest that the arts do indeed have lower price elasticities than other goods and services when properly evaluated at comparable prices.<sup>27</sup>

#### 4.3. Income elasticity differences by level of aggregation

Despite the somewhat mixed results for price elasticity, the results are even more ambiguous for income elasticity. *Pommerehne and Kirchgassner (1987)* found income elasticities greater than 1.0 in German theater for both average- and high-income consumers, but with higher-income people having lower income elasticity than do average-income people. However, while they used this differential income result in their summary, they warned about the low level of statistical significance in their income results and noted “income elasticities are not much larger than one” (p. 48).

Felton’s results were more indicative of the varying income elasticities derived in more refined studies. *Felton (1989)* found no statistically significant income elasticities for her individual opera companies, and found (1992) income elasticities of either less than one or about one for her two largest orchestra groups (again, while this is still fairly aggregated, it at least distinguishes orchestra size). *Felton (1992)* did, however, find luxury good evidence for her highest and second highest budget samples of ballet companies (but nothing significant for opera). *Felton (1994/1995)* also found a higher income elasticity of 1.4 for her “total attendance” group vs. an elasticity of only 0.82 for her “subscriber” database for US orchestras. By contrast, *Luksetich and Lange (1995)* found no statistically significant income elasticities at all in their large market vs. small market study of orchestras. Unfortunately, neither *Schimmelpennig (1997)* nor *Abbé-Decarroux (1994)*, who did such useful work regarding more disaggregated price elasticities, were able to estimate any income elasticities since they lacked income data.

Superficially, the strongest evidence for income elasticity estimates being notably different when less aggregated data are utilized is *Greckel and Felton (1987)*, who derived a statistically significant income elasticity of 6.13 in their second demand equation for the Louisville Orchestra, although their other orchestra income elasticity of 2.66, as well as the 2.26 estimate for the Bach Society were not statistically significant. Furthermore, this suspiciously high income elasticity of 6.13 is derived in an equation with only ticket price and concert hall capacity as control variables. Even more importantly, that high

<sup>27</sup> However, since other forms of entertainment with lower ticket prices share these characteristics, this conclusion is not inevitable. Furthermore, a focus on the weight of the admission price relative to the full consumption price would suggest that a sport like cricket, where a match can last for up to five days, would have the lowest comparable price elasticity of demand of nearly all live entertainments; regrettably, despite their insightful analysis of county cricket, *Paton and Cooke (2005)* do not estimate price elasticities.

income elasticity estimate dropped to the 2.66 noted above and lost statistical significance when a proxy variable was added to account for a four-year period in which the conductor of the Louisville Orchestra and his successor were widely unpopular. Hence, this evidence for high income elasticities of demand when using organization-specific data is not compelling.<sup>28</sup>

Abbé-Decarroux and Grin (1992) also studied individual (Swiss) arts organizations, and estimated logit models that included pre-tax monthly personal income, but the estimated attendance probability coefficients were quite small (although statistically significant for two of three organizations). Only one of the Corning and Levy (2002) income elasticities for their target theater company across three geographical venues exceeded one while also being statistically significant. Gapinski's (1984) study of production and demand functions for the Royal Shakespeare Company (RSC) did generate a statistically significant income elasticity above one (1.33) in an equation with only price and the constant term. While he was willing to conclude that "an RSC cultural experience is a luxury good", he faced considerable difficulty in eliminating autocorrelation and heteroskedasticity problems and, even after correcting for those problems, was forced to eliminate seven of his eight demand equations "because of intercept or substitute-price insignificance" (p. 463). Thus, the evidence is mixed that estimating demand at the individual organizational level will reliably result in higher income elasticities compared to more aggregated data.

In the following sections we discuss a range of problems encountered in econometric studies of arts demand, beginning with that of model selection.

## 5. The modeling problem: Single versus simultaneous equations

Economists have wrestled with the important issue of whether to model arts demand using a single equation, i.e. by treating the performing arts market as essentially a recursive process in which supply decisions are not determined in the same time period as demand, or to model arts demand as one of several equations in which key endogenous variables are determined simultaneously, hence requiring more complex estimation techniques.<sup>29</sup> In this section we consider these measurement and modeling issues by de-

<sup>28</sup> Carson and Mobilia (1989) also found high standardized income elasticities for Broadway that were highly sensitive to seasons, with highly positive income elastic demand for the fall, winter and spring seasons, but highly negative income elastic demand during the summer season. Given the outlier nature of their estimates (+5.78 to -4.74) and the lack of any modeling foundation compared to, say, Moore (1966), who found dramatically different results (although not adjusting for seasonality), it is difficult to know how much weight to place on these income elasticity results.

<sup>29</sup> Another modeling issue that has been less prevalent in the arts demand literature, but is nevertheless important, is the functional form to choose when estimating a system of demand equations for differentiated goods so as to make parameter estimation feasible. For example, both logit and the almost-ideal-demand-system [e.g., Pommerehne and Kirchgassner (1987)] can be viewed as imposing constraints on substitution patterns so as limit the number of parameters that would have to be estimated.

scribing the key features of three representative studies: Moore (1966), Withers (1980), and Luksetich and Lange (1995).

### 5.1. Moore: "The demand for Broadway theater tickets" (1966)

Moore did not present an explicit utility maximizing framework, but focused on developing a defensible econometric model to estimate demand elasticities that may solve the puzzle of low Broadway attendance growth. Since ticket prices were fixed in any one time period, an explicit time-series model was necessary to derive price elasticity, and a more broadly defined full attendance cost elasticity of demand.<sup>30</sup> However, he also examined the determinants of the cost of an evening of entertainment per person beyond the cost of tickets by using cross-section analysis from data derived from a survey of seven Broadway houses and 18 performances. He also used the cross-section survey results to estimate an income elasticity of 1.03 based on relative frequency of attendance as a function of income.

His time-series model included three equations related to the  $i$ th time period, with  $A$  representing attendance,  $Y$  a measure of permanent income,  $C$  the cost of attending the theater,  $S$  the number of shows,  $P$  ticket prices,  $M$  a dummy variable for sound movies,  $T$  the transportation cost to the theater, and  $O$  the other costs of attending a Broadway play:

$$A_i = f(Y_i, C_i, S_i), \quad (1)$$

$$S_i = g(A_i, P_i, M_i), \quad (2)$$

$$C_i = h(P_i, T_i, O_i). \quad (3)$$

He then postulated that average attendance per show was probably a constant, so that a long run equilibrium condition could be expressed as  $A = \alpha S$ , yielding a four equation system with four endogenous variables,  $A$ ,  $S$ ,  $P$  and  $C$ . However, since it could not be assumed that the market was necessarily in long-run equilibrium, price was treated as exogenous. Furthermore, given the lack of adequate data regarding travel expenses and other expenses of attending the theater, he assumed that such costs, including those related to population movements within the New York area, had no trend over his 1928–1963 time period and were uncorrelated with the other variables, allowing him to drop Equation (3) and substitute price  $P_i$  for cost  $C_i$  in Equation (1).

These modifications allowed him to estimate three basic variations of the model:

- (1) a "naive" approach that assumed that the supply of shows was determined outside the system, estimated both as linear and multiplicative;
- (2) a simultaneous equation approach using two-stage least squares (again both linear and multiplicative); and

<sup>30</sup> Throsby and Withers (1979, p. 111) observed that time-series analysis is more appropriate "for past analysis and for prediction" since time-series data are capable of describing the effect on behavior of changes in a variable, whereas a cross-section elasticity can only describe the effects of differences in that variable.

- (3) two-stage estimation, but with the income elasticity constrained to equal its value estimated from the cross-section data (i.e. 1.03) as a way to limit the bias in the income elasticity estimate created by the absence of population and travel expense variables, which are likely to be correlated with income.<sup>31</sup>

The three most important results are:

- the price and income elasticities were less than one in absolute value, and those results were largely independent of the specific econometric specification;
- the low income elasticity results in all of the time-series estimations were surprising to Moore and prompted him to observe that he had not been capable of controlling for the opportunity cost of time; and
- despite the fact that Moore believed that it was a “dubious” assumption to consider the number of shows (his Equation (2)) as being determined “outside the system”, the fact that shifting to two-stage estimation increased the price elasticity results only trivially compared to the naive single-equation cases has led most later researchers to justify the use of single-equation recursive techniques.

5.2. *Withers: “Unbalanced growth and the demand for the performing arts: An econometric analysis” (1980)*<sup>32</sup>

The basic model was a straightforward application of theory postulating quantity demanded as a function of relative prices and income, adjusted to reflect the realities of using aggregated data applicable to the entire performing arts for the period 1929–1973 (largely from the US *Survey of Current Business*). Hence, the general estimating equation, defined for all time periods  $t$ , was:

$$(Q/\text{Pop}) = f(P_A, P_S, I, D), \quad (4)$$

where  $(Q/\text{Pop})$  is the number of attendances per capita in the population, the  $P$  terms are the price of attendance and the price of substitutes, respectively,  $I$  is income, and  $D$  is a measure of the distribution of income (defined to increase with inequality favoring the wealthy). Expected partial derivatives are positive for all variables except own-price. Withers’ normalization of the attendance dependent variable by dividing by population became conventional in many arts demand studies.

Withers’ critical contribution was to adapt the Owen (1969) approach to adjusting hourly wage rates by the unemployment rate so as to better measure leisure price, and then to utilize the Becker (1965) concept of full income (defined over all available hours, not just working hours) so as to impute leisure time as part of this full income, while incorporating the price of leisure into the consumer price index deflator. This generated

<sup>31</sup> An approach recognized as problematic by Moore and others, such as Throsby and Withers (1979, p. 111).

<sup>32</sup> Withers’ study was a development and application of a model originally proposed in Throsby and Withers (1979, Chapter 3).

the alternative “time allocation” estimating model defined for all time periods  $t$ :

$$(Q/\text{Pop}) = f(P_A, P_S, P_L, F, D) \quad (5)$$

with the new variables  $P_L$  defined as the price of leisure, and  $F$  as full income. In turn, those two variables are defined as:

$$P_L = w(1 - U_R), \quad (6)$$

where  $w$  is the hourly wage rate and  $U_R$  is the unemployment rate, and full income is the standard Becker formulation:

$$F = T_C P_L + T_W w + Y \quad (7)$$

with  $T_C$  and  $T_W$  defined as hours of consumption and hours of work ( $1 - T_C$ ), respectively, and  $Y$  defined as property income. An adjusted price index was also derived incorporating the leisure price. The expected signs on the first partial derivatives are the same as in the non-time allocation model for own-price, substitute price and income distribution ( $D$ ), and negative for the two new leisure price variable ( $P_L$ ), while positive for the full income variable ( $F$ ). The arts are considered a priori to be a “superior time-intensive good” in this formulation.

Withers utilized single-equation estimation, justified in part by the Moore (1966) results, but also by his belief that the performing arts market is inherently recursive in structure, with price in one time period affecting attendance in that same period, but any supply of new performances having an observable effect no earlier than the next time period. The latter hypothesis is attributable to the “advance planning and announcement of production and seasons and their prices that is typically required in this industry” [Withers (1980, p. 737)].<sup>33</sup> Thus, ordinary least squares was used with the double-log transformation, which he also justified by prior econometric evidence that this form is preferable for non-necessities in single good estimation (pp. 737–738).

There are two broad reasons for the popularity of the Withers (1980) results. First, the largely successful decomposition of the effect of rising income in the time allocation model into a relatively high “full-income” effect offset by a smaller real leisure price effect is consistent with a priori expectations that the arts can indeed be considered luxury goods that are time-intensive in consumption. Second, the low price elasticities in the more structurally sound time-intensive model (which increase to approximately unity or are only modestly elastic in the conventional model) are also consistent with a priori notions among many arts economists. Of course, these price elasticities are applicable to an extremely aggregated measure of the performing arts, and do not apply to specific organizations in specific product and geographic markets, nor does their interpretation

<sup>33</sup> Heilbrun (1984, 1996) stressed the difficulties in smoothly increasing the quantity supplied of arts services and emphasized the role that periodic supply shifts can play in later stimulating observed increases in arts attendance. However, that issue is never linked in the econometric literature to the choice of single or simultaneous equation methods.

reflect any sensitivity to the issue of the level of prices at which such “industry-wide” elasticities are estimated, as discussed above.

Interestingly, Withers’ (1980) cross-price elasticity results for reading and recreation are rarely cited, even though they are generally higher than the cross-price elasticities estimated by Gapinski (1986). Of course, Gapinski’s data were much more disaggregated, so that he was also able to estimate cross-price elasticities faced by individual arts organizations (three of which were indeed quite high). Furthermore, the substitute prices were much better defined as the ticket prices charged by the other art forms.<sup>34</sup> Another relatively ignored feature is Withers’ conclusion that the results potentially weaken the case of government support of the performing arts on the grounds of financial distress. That is, the high income elasticities only partially offset by the elasticity of the price of leisure together with the relatively low price elasticities of demand, suggested to Withers that “the potential for continued growth of private market support for the performing arts should be recognized” (p. 742). It is ironic that this conclusion would be downplayed inasmuch as the very title of his paper announces his interest in exploring the unbalanced growth issue in the performing arts.<sup>35</sup>

Ekelund and Ritenour (1999) represents a rare subsequent attempt to focus on this problem and isolate the effect of the time costs on US symphony concert demand using a less aggregated unit of analysis than Withers, although still an aggregation of anonymous individual orchestra data.<sup>36</sup> They estimated a single linear OLS equation, which they suggested may be even more justified for orchestras than for Broadway as studied by Moore (1966), also independently testing for possible simultaneity bias. They regressed annual per capita symphony orchestra concert attendance on average ticket price, the price of audio recordings, annual real disposable income, and their key variable, the cost of time (as measured by the annual average real hourly wage rate). Only the coefficient on the substitute price variable behaved poorly (negative in sign rather than positive, and significant at only the 0.10 level). The own-price coefficient was negative and strongly statistically significant (although low in magnitude consistent with a

<sup>34</sup> See further in Section 8. This aspect of the Gapinski (1986) results can easily be missed. For example, Fernández-Blanco and Baños Pino (1997) observe that Gapinski “shows that the best substitute for a theater play is not a film, but a different theater play” (pp. 62–63). In fact, the substitute price for the two theaters in the Gapinski database is the average of prices of opera, symphony and dance only [Gapinski (1986, p. 21)].

<sup>35</sup> Another forgotten feature of the Withers results is that when his model was applied to both Australian and Canadian data in Throsby and Withers (1979), the results were not as strong statistically. In the Australian case, while the ticket price elasticities were largely consistent with the American results, no significant income effect was found using either the conventional or the time allocation model (pp. 115–117). While Throsby and Withers attributed some of these problems to weaknesses in their Australian data, it is interesting that those data were at least less aggregated than in the US case (i.e. applying to seven major professional performance companies, although over a shorter time period of 1964–1974; p. 115). Any degrees-of-freedom and related problems were even more severe in the Canadian data, and the model could not be estimated in that case at all [Throsby and Withers (1979, p. 112)].

<sup>36</sup> Ekelund and Ritenour acknowledged this aggregation problem by noting that results may be different if city or SMSA data were to be used for specific orchestras, citing some panel data evidence supporting that possibility.

low ticket price elasticity of demand), while income had a positive effect on attendance per capita (but with a normalized coefficient suggesting less than unit elasticity), and the wage rate (value of time) had a negative coefficient.

Despite its modeling limitations and remaining aggregation problems, the Ekelund and Ritenour (1999) results are consistent with the fundamental idea that any positive income effect on arts demand will be partially counterbalanced by the time-intensive nature of live performances and the opportunity cost of that time. They were duly cautious in evaluating their findings, but interestingly tended to stress the threats to the arts resulting from their results in contrast to the more optimistic assessment provided by Withers (1980).

### 5.3. Luksetich and Lange: “A simultaneous model of nonprofit symphony orchestra behavior” (1995)

Luksetich and Lange had previously employed two-stage least squares methods to estimate orchestra demand.<sup>37</sup> Their key findings in 1984 were:

- price elasticities varied by orchestra size, becoming more elastic as the size of the orchestra fell, with major orchestra demand quite inelastic and metro orchestra demand modestly elastic;
- the price elasticities became less elastic with the inclusion of a donor price variable (measured as total donations divided by attendance), although the donor price elasticities themselves were not statistically significant;
- their total sample price elasticity estimate of about  $-0.48$  both with and without donor price was notably close to estimates from major prior studies; and
- they concluded that “in general” when comparing the OLS and the 2SLS results there was support for the latter, and the elasticities of the non-price determinants were stable regardless of procedure.<sup>38</sup>

The superior database for the 1995 paper allowed for a more thorough exploration of the relationship between factors under managerial control and various orchestra performance measures, which they modeled as a six-equation system with attendance, average price, administrative expenses, orchestra quality (using non-administrative orchestra spending as a proxy, an expansion of the “wages” variable they had used in 1984), number of concerts, and donations simultaneously determined. The model was

<sup>37</sup> See Lange and Luksetich (1984). A related paper [Lange, Luksetich and Jacobs (1986)] confirmed that equations for orchestras of different size and classification should be estimated separately; see also Luksetich and Lange (1995, p. 52).

<sup>38</sup> However, the case for the superiority of the more complex approach was not overwhelming. The authors noted that the estimated equations of “price” from the single-equation estimation of 2SLS varied widely regarding the adjusted  $R^2$  results across the orchestra sub-samples, possibly suggesting that the instrument price in the second stage was not necessary because the firms were not price takers so that no supply curve existed [Luksetich and Lange (1984, p. 43)].



estimated using 2SLS regression, the second stage being estimated using pooled cross-section, time-series techniques with estimates corrected for heteroskedasticity and serial correlation.

Luksetich and Lange (1995) is perhaps the best example of estimating arts demand as part of a multi-equation simultaneous equation system using 2SLS estimation techniques to date. Their findings are summarized here. While a surprising omission in the demand equation was any measure of regional educational level, the most important findings related to the low price elasticities of demand for each orchestral group. Although this is a familiar result with aggregated data (although here at least segmented by orchestra size), it has especially interesting implications for orchestral full-income-maximizing (ticket-plus-donation) pricing strategies and for the interpretation of price elasticity results when organizations have multiple interrelated revenue sources, as discussed in Section 4.2 above.

Regarding the individual components of the demand equation, ticket price performed well with a derived price elasticity of  $-0.33$  (p. 56). Surprisingly, quality had a negative effect on per capita attendance, although their quality proxy of artistic personnel and total concert production expenses per concert was shown by Tobias (2004) to be an unreliable predictor of *expert* opinions regarding the quality of arts organizations. The result that the number of concerts per capita in the region was a strong predictor of regional attendance per capita was consistent with arguments that variations in the sheer availability of artistic resources is critical to explaining attendance variations [Heilbrun (1984, 1996), Gold (1980), Khakee and Nilsson (1980)], but in a model expressly designed to correct for simultaneity bias, it is surprising that attendance did not also appear as an independent variable in the concert equation.

The most noteworthy results of the Luksetich and Lange (1995) analysis relate to the interaction of ticket buying and voluntary donations. Not only were the implied price elasticities low for all three orchestra types, but the interactions between the ticket price and gifts per capita further clarified the fact that orchestras have been following an excessively low price strategy. That is, even when adjusting for the possibility that higher ticket prices would induce arts patrons to partially reduce the donated portion of their “full price” of attendance, orchestras would generate more total income if they were to substantially increase average ticket prices (even ignoring more targeted price increases that might be justified if the audience could be further segmented). For major orchestras, the mean elasticity of gifts per capita to ticket price was  $-1.19$  (p. 56). But, even considering that interaction between revenue sources, the authors calculated that the major orchestras could double the average ticket price from its current mean value in order to maximize ticket revenues, but more importantly could increase ticket prices as much as 62 percent and still maximize the sum of ticket-plus-donated revenues (p. 58).

Of particular importance was their calculation that following such a ticket price increase, the resulting average-ticket-price elasticity of demand at those higher prices was still well within the inelastic range at  $-0.65$ , fully consistent with the sports literature result that the optimal price for such “performance firms” is to actively price in the

inelastic range of their demand curves, hence explaining the “paradox” of low price elasticities estimated in empirical demand studies.<sup>39</sup>

## 6. The multicollinearity problem: Education versus income

The non-econometric evidence considered in Section 2.3 above demonstrated that education is likely the most important single variable in explaining variations in performing arts attendance. However, regression-based models have not generally succeeded in confirming this result. While the typically high correlations between education and other key independent variables (class, occupation, age, and even ticket prices in some models) is a key reason, multicollinearity is not the only complication.<sup>40</sup> Competing measures of education, distinctions between arts training and general education, and conceptual conflicts regarding the underlying causal relationships are also to blame.

The issue of how to measure education and whether to distinguish it from arts specific training at home, arts specific training in school, or just from past experience is highly problematic. As McCaughey (1989) asserts, “the reasons for the positive association of general educational attainment with participation in the arts are not fully understood; and how specifically arts education fits into this association is not clear” (p. 48). For example, Orend and Keegan (1996) and Relish (1997) stress the socialization or network effects of education in affecting arts participation rates, while Globerman and Book (1977) put education into a consumer production function model, although with only limited success in isolating a unique “consumption efficiency” effect on arts demand. It is also possible that education reduces search costs and generates more accurate perceptions of performing arts prices, a possible link to attendance established by Globerman (1978).

The Globerman and Book (1977) contribution is unique. While education is most commonly cited as merely a taste-determining variable, cultural economists are perfectly comfortable with viewing education as one form of specific consumption capital affecting relative shadow prices, and hence the constraints, facing arts consumers. However, no one else has actually tested for an explicit education consumption efficiency

<sup>39</sup> Metro orchestras appeared to have come closer to their optimal pricing strategy, requiring only about a 10 percent price increase to maximize combined ticket and donation revenues (the price elasticity of gifts was quite small for those orchestras), while small market orchestras were in the unique but bizarre position of being able to maximize ticket revenues with only a 31 percent price increase, but would actually generate even higher combined ticket and gift revenues if they were to raise ticket prices by 57 percent (the result of having an estimated positive elasticity of gifts with respect to ticket prices); see Luksetich and Lange (1995, p. 60).

<sup>40</sup> For example, due to the high correlation between incomes and prices over time, Goudriaan and de Kam (1983) could not get statistically significant elasticity estimates for both in their time-series demand equations for theater and concerts without imposing a constraint on the income elasticity from their separately estimated cross-section equations. While they were aware of the weaknesses to that approach [citing Kuh and Meyer (1957) on problems with extraneous estimates], they observed that “there was no alternative to get significant results” (p. 39).

link to arts attendance. Despite its creative application of the consumer household production model [Michael (1972); Michael and Becker (1973)], the Gliberman and Book (1977) effort to isolate this effect yielded only mixed results.<sup>41</sup>

Regardless of the exact causal connection between education and arts consumption, the various classifications of formal schooling are common measures of “education” in econometric studies. The specific way such formal education enters those equations depends on the data being used and the unit of analysis being studied. In what might be called “Type A” (aggregative) studies, the aggregate attendance or participation behavior of the population is known, but the data are not based on the attendance patterns of specific individuals with known personalized demand-determining characteristics.<sup>42</sup> The dependent variable is typically some version of attendance at specific arts organizations normalized by the regional population in which they are located. Consequently, the education independent variable must reflect the aggregative characteristics of the population in that region (e.g., “percent of the population with a college degree” or “median years of schooling”, etc.). The same measurement issues arise regarding income in Type A studies (i.e. “percent of the population earning above \$30,000” or “per capita disposable income”, etc.); but some localized measures of income are typically incorporated into those demand equations. These studies sometimes also incorporate measures of local performing arts ticket prices or even prices of substitutes and organizational quality proxies.

Alternatively, in “Type P” (personalized) studies, the attendance or participation behavior of particular individuals can be matched to their unique demand-determining characteristics. Dependent variables can be continuous (e.g., “number of arts events attended”, or “number of attendances”) or binary (e.g., “1 if attended at least one time, 0 if did not attend at all”). In Type P studies a wider variety of educational variables can be included beyond the usual formal education measures, and in fact, distinctions can be made between own education, parents’ education (sometimes distinguished by gender), specific types of education (e.g., art training at home versus at school), or diverse measures of exposure to various art forms or practical training in those fields. However, they are not capable of incorporating arts organization prices or organizational quality. At best, a variable can be inserted such as “lives in city over 50,000” [Lewis and Seaman (2004)] to try to capture some of those localized “fixed effects”. In this sense, Type P studies are not estimations of well-defined demand functions.

The results of 12 econometric studies that include both income and education variables are fully documented in Seaman (2005, Table 14). Gapinski (1981) viewed

<sup>41</sup> Seaman (2005, pp. 90–94) provides the details.

<sup>42</sup> As usual, the variability in databases can create anomalies. For example, Bajic (1985) had access to subscriber lists (not audience survey data) for both the St. Lawrence Theatre and the Toronto Free Theatre. While this provided information on the residential location of individual subscribers and hence allowed the construction of an independent variable measuring commuting distance to the theaters, there was no personalized information on education or income – hence requiring the construction of an aggregated measure of education and income per “zone”.

his own elasticity estimates as only tentative, but paid homage to the Ford (1974, Vol. II) non-regression based findings by noting their general similarity to his results. More importantly, the overall results of the 12 studies provide only relatively weak confirmation that education measures reliably outweigh income measures as determinants of arts demand. Only five of the studies (41.7 percent) found evidence for the dominance of education; strong pro-education evidence is in Ganzeboom (1989) and Peterson, Hull and Kern (2000). Gapinski (1981), Gray (2003), and Lewis and Seaman (2004) provide modestly strong evidence that education outweighs income, while two studies generated opposite results favoring income [Bajic (1985); Bonato, Gagliardi and Gorelli (1990)]. Although Globerman and Book (1977) focused on the separate issue of whether the role of education is consistent with the household production model, their estimated elasticities were higher for income than for education, providing some evidence against the education dominant hypothesis.

The remaining four studies yielded ambiguous results. Goudriaan and de Kam (1983) did not calculate education elasticities, making direct comparisons difficult; however their t-statistics on the education variable were very high (but with the non-normalized coefficients, difficult to compare in economic significance across determining variables). Meanwhile their income elasticities were only high when age and education are dropped from the equations. Jenkins and Austen-Smith (1987) generated a higher education-demand elasticity in their two-stage least squares equation but a lower elasticity relative to income using OLS. But in both cases, the standard errors were quite high when both variables were included along with “social class”. The income coefficient in Dobson and West (1989) was negative, but the standard educational variable was also weak and statistically insignificant. Their strongest results were for two formal education substitutes: childhood participation in non-school theater performances as opposed to school-related performances, as well as childhood theater participation rather than just exposure to the theater, both strongly influenced adult theater attendance. Finally, Abbé-Decarroux and Grin (1992) did not utilize a formal education variable, but found similar evidence to that of Dobson and West (1989) that “arts training” variables performed well (although their variables did not distinguish between home and school exposures). But they also found fairly strong evidence for the importance of income (although less strong than for the arts training variables).

## **7. The taste cultivation problem and human capital: Habit formation, learning-by-consuming and rational addiction**

Cultural economists have always stressed that current arts demand (whether for live performing arts services or the stock of tangible art works) is especially influenced by past arts exposure, and hence that inter-temporal dynamics should be incorporated into a well-specified demand model. Nevertheless, except for indirectly addressing this issue via the inclusion of age, education, or childhood exposure [e.g., Morrison and West (1986)], explicit inclusion of lagged dependent variables as separate determinants

has been relatively rare. Only Houthakker and Taylor (1970), Goudriaan and de Kam (1983), Oteri and Trimarchi (1990), Krebs and Pommerehne (1995) and Urrutiaguer (2002) included a one-year lagged endogenous dependent variable, and in each case it was strongly positive and statistically significant.<sup>43</sup> Carson and Mobilia (1989) defined the lag differently. Their much shorter weekly lagged dependent variable had negative effects on current attendance, a result they explained as capturing the effect of infrequent arts consumption such that very recent attendance actually reduces the likelihood of attending again in the current period. With a longer lagged dependent variable of even one year, what seems to be captured is not infrequent and discrete arts consumption patterns, but a proxy for the cumulative effect of some version of what might be called the cultivation of taste. It is not always appreciated that there are multiple versions of this phenomenon, with each having somewhat different implications for our understanding of arts demand and for optimal arts marketing strategies.<sup>44</sup> In the following sections we consider three such explanations: habit formation, learning by consuming and rational addiction.

### 7.1. *Habit formation*

The most “passive” explanation for past consumption affecting current and future consumption is simple habit formation, the behavioral inertia so characteristic of much of human behavior [Houthakker and Taylor (1970); Pollak (1970)]. If habit formation is the primary reason for the strong performance of lagged dependent variables in arts demand equations, arts managers should go to great lengths to introduce the arts to young audiences with regularly scheduled targeted programs to get them into the habit of attending, regardless of any particular human capital formation effects. In fact, the

<sup>43</sup> Krebs and Pommerehne attributed the difference between their estimated low short-run price elasticity ( $-0.16$ ) and the higher long-run elasticity ( $-2.6$ ) to the stronger impact of their one-year lagged attendance variable (actually mislabeled as a lagged quality variable in their Equation (1A), p. 26). They interpreted this result, following Houthakker and Taylor (1970), as reflecting “habit formation”, although their partial adjustment model (where consumers are assumed to partially adjust to long-run equilibrium) was not identical to their predecessors’ “stock adjustment” model [see Krebs and Pommerehne (1995, p. 25, and their Note 25, p. 30)].

<sup>44</sup> Lévy-Garboua and Montmarquette (1996) emphasized the two most interesting reasons identified by Pollak (1970) for a distinction between long- and short-run demand functions, ignoring his more institutional “contractually fixed commitments” explanation [Pollak (1970, p. 745)]. Thus, while Pollak originally did have a three-fold classification scheme, only two approaches were cited by Lévy-Garboua and Montmarquette. They defined the simple “habit forming” phenomenon identified by Pollak (his third case) as a “deterministic approach”, by which any type of habit formation or the creation of consumption capital, whether myopic or farsighted, is considered an inevitable reality of the human condition (p. 28). This was contrasted with Pollak’s second approach stressing the ignorance of consumers about their own preference orderings “outside the range of . . . past consumption experience” who rely upon personal experience through “a time consuming learning process” [quotations from Pollak (1970, p. 745)]. Lévy-Garboua and Montmarquette (1996) also cited two other antecedent related approaches (Note 2), but seem to have been the first to actually use the phrase “learning-by-consuming”.

important role of specific consumption capital is closely associated with the notion of rational addiction discussed below.

Houthakker and Taylor (1970) remains the classic “state-adjustment” dynamic model linked to habit formation, estimated using OLS as part of a comprehensive study of consumer demand in the United States (reporting results for 82 separate spending categories). While they also reported results for motion pictures and spectator sports, their estimated equation for “legitimate theater and opera” is often cited, both for the specific derived short-run versus long-run “relative price” elasticities ( $-0.1827$  and  $-0.3109$ , respectively) as well as for their conclusion that while theater and opera are subject to habit formation, it “wears off quite rapidly” (p. 131). While not as frequently cited, they also derived short- and long-run elasticities with respect to total expenditures, ( $0.7407$  and  $1.2604$ , respectively). Their equations for all spending categories were estimated over the period 1929–1964 using the *Survey of Current Business* as the principal source of data.

The Globerman (1978) approach to indirectly isolating the effect of past consumption on present consumption is novel in the arts demand literature. He investigated the determinants of public perceptions about performing arts prices, including the role played by past attendance in generating more accurate price perceptions that can be viewed as stimulating attendance. His approach might be viewed as a more traditional price search model in contrast to the “utility function search” model of Lévy-Garboua and Montmarquette (1996); see also Brito and Barros (2005). Furthermore, Globerman’s finding that higher levels of formal education are at times related to lower perceived minimum arts prices is suggestive of an important and generally ignored link between education and attendance that requires neither a “taste development” explanation, nor even a beneficial productivity effect in a household production context [as in Globerman and Book (1977)].

## 7.2. Learning by consuming

Another version of taste cultivation has become known as learning-by-consuming, in which consumers are characterized as uncertain about their utility functions but learn their own subjective preference structures through a process of consumption experiences generating either positive or negative feedback [Lévy-Garboua and Montmarquette (1996); Brito and Barros (2005); Ulibarri (2005)]. Abbé-Decarroux and Grin (1992) presented a “hybrid” type of learning-by-consuming model, in which consumers already have well-defined utility functions (with older consumers being especially risk averse), but poor knowledge about the product characteristics of different suppliers which can be clarified by positive and negative feedback from a kind of “lottery” of actual consumption. These models can have implications for the programmatic choices of arts managers.<sup>45</sup> Globerman (1978) put forward a search model that has some sim-

<sup>45</sup> Usually biased toward less esoteric and less risky programming; see, for example, Abbé-Decarroux and Grin (1992).

ilarities to both passive habit formation and a broader interpretation of learning-by-consuming that relies upon a process of price discovery rather than taste discovery.

Lévy-Garboua and Montmarquette (1996) has become the most cited example of the learning-by-doing approach. Their approach to the dependence of current consumption on past consumption did not include a lagged endogenous dependent variable in their estimated equations, but focused instead on the contrast between the various predicted effects of key variables in testing their learning-by-consuming model (1996, p. 39). They confirmed the methodological challenge of how best to incorporate past consumption into studies of current consumption by observing that despite the general consensus among economists and sociologists that such effects are important for a wide variety of goods, “it is seldom possible to directly verify this assertion . . . using individual data and after controlling for many wealth, price and taste variables” (pp. 27–28).

Their approach fully accepts the non-Beckerian premise that changes/differences in behavior can be linked to taste changes/differences rather than subtle constraint variations and incorporates two key elements: first, any new experience of a good to the consumer reveals an unexpected positive or negative “increment in his taste” for the good, with this increment treated *ex ante* as a random variable with zero mean; and second, the concept of someone developing a taste for an art form such as the theater can be viewed as their having experienced repeated pleasant surprises when attending plays and hence revising expectations upward. Lévy-Garboua and Montmarquette (1996) claimed three theoretical advantages for this representation:

- it is compatible with an assumed strong heterogeneity of tastes and the independence of individual choices;
- it allows for extensive differentiation of cultural goods further magnified by the “unique nature of each ‘cultural’ experience provid[ing] new possibilities for surprises and implies long learning periods”; and
- it maintains intertemporal separability of the utility function conditional on past consumption by viewing consumers as having uncertainty regarding their preferences that prevents them from rationally anticipating the future taste (utility) that will be acquired over time (p. 28).

A full description of their quite complex model is beyond the scope of this chapter,<sup>46</sup> however, the especially important features can be identified. The utility function includes the quality-adjusted quantities for all goods  $i = (1, \dots, r)$  whose consumption “may give rise to non-systematic cultivation of taste” (1996, p. 28):

$$U = u(s_1 x_1, \dots, s_r x_r), \quad (8)$$

where  $x_i$  = the quantities consumed of market goods and the  $s_i$  weights represent “subjective qualities” anticipated before the decision is made to consume the goods; these weights depend on previous consumption experiences. A similar utility function can

<sup>46</sup> See Seaman (2005) as well as their own more concise description in Lévy-Garboua and Montmarquette (2003).



be defined for each time period. If we now define  $x_a$  as attendance at an arts performance, the effect of consumption experiences that yield additional information about the subjective quality of that good,  $s_a$ , can be represented in period  $t$  as:

$$s_t = E_{t-1}(s_t) + \gamma_t \quad \text{if } x_t > 0, \quad (9)$$

where  $E_{t-1}$  represents the expectation operator before period  $t$ 's choice, and  $\gamma_t$  is the "taste surprise" experienced in period  $t$  (i.e.  $E_{t-1}(\gamma_t) = 0$ ). This can be called the "experienced taste for arts consumption in period  $t$ ". As noted above, consumers who find themselves "developing a taste" for music (for example), will generally experience pleasant surprises, i.e.  $\gamma_t > 0$  is more common than the reverse. Consumers are viewed as basing their expectation of taste solely on their past experiences. Therefore, taste expectations are identical across all time periods (as noted, preserving the intertemporal separability of the utility function). Lévy-Garboua and Montmarquette then define a taste-constant Frisch demand function (i.e. marginal utility of income or wealth constant in contrast to Marshallian demand functions) that suggests that a good such as arts attendance that is assessed by a consumer as having a high quality will have a low "personalized price". However, analogous to the household production model of the new consumer theory when the marginal product of any unit of attendance is quite high in producing "arts appreciation", this greater potency of each unit of attendance in generating quality also implies that less quantity of the good  $x$  is required to achieve a given utility level (see Equation (8) above).

This generates a key insight from the learning-by-consuming approach, and allows Lévy-Garboua and Montmarquette to "impute" a price elasticity of demand from their empirical analysis even though they are missing a price variable in their vast database. That is, if the price elasticity of  $x$  is  $> 1$  in absolute value, the experience of consuming the arts will have a positive effect on current consumption when the good was "enjoyable overall", but a negative effect when it was not (1996, p. 30). They clarify this relationship by modifying the demand function so as to isolate the marginal effect of quality on the quantity of the arts consumed,  $\delta x / \delta s$ , which they derive after several steps to be:

$$\frac{\delta x}{\delta s} = -\frac{x}{s}(1 + e), \quad (10)$$

where  $e$  designates the price elasticity of demand for  $x$ . This directly implies that

$$\frac{\delta x}{\delta s} \frac{s}{x} = -(1 + e),$$

i.e. that the elasticity of arts attendance with respect to perceived arts quality (which they call the "taste elasticity") =  $-(1 + \text{own price elasticity of demand for arts attendance})$ . That elasticity is  $> 0$  if  $e < -1$ , i.e. if the price elasticity is elastic. If the price elasticity is inelastic ( $e > -1$ ), the quality (or taste) elasticity of the demand for the arts becomes negative, and unitary price elasticity implies a zero quality elasticity. As interpreted by Lévy-Garboua and Montmarquette, this allows them to measure the price elasticity of



demand when the model measures accumulated experience and taste for consumption (1996, p. 30). As noted above in Section 6, normally price elasticity is not capable of being estimated in Type P studies due to the absence of data on ticket prices.

Since their Ministry of Culture survey database was rich in opinion type questions, some of the key empirical results referred to those variables. For example, the authors viewed opinions referring to greater appreciation of the actors and the quality of the text of the play as indirectly measuring the taste for the theater. Two of their constructed variables, were designed to measure taste more directly. They used the “appreciation scores” from 0 to 10 that respondents had assigned to a list of 56 (23 theatrical writers and 33 actors/directors), and designated high scores of 9 and 10 as indicating “evidence of a taste for the theater” (p. 39). But they then found that writers were seemingly treated differently than actors and directors (based on the performance of these variables in their estimated tobit model), so they identified a taste for reading as a substitute for live theater among those who showed a strong preference for writers, and vice versa for those giving especially high ratings to actors and directors. They found support for this view in the negative effect of reading journals and magazines on the frequency of theater attendance.

They found that the variable measuring the percentage of actors and directors known was the best measure of “the degree of familiarity with or experience of” the theater, and they designated a person who claimed to know more than 80 percent of the names put to him or her as knowing the theater well. They found it noteworthy that the performance of their variable designating knowledge of writers performed more poorly than their variable measuring knowledge of actors and directors, and concluded that “it is necessary to have attended the theater personally in the past in order to know the actors and directors whose talent can only be appreciated on the stage and in action” (p. 39). They thus identified this variable as their prime measure of previous theater attendance and predictor of current attendance (and based on their tobit estimation the probability of someone knowing more than 80 percent of the actors and directors *not* attending the theater fell from 0.49 to 0.02).

This finding then becomes the key to their imputing price elasticity from their survey data. They estimated the price elasticity of demand (see their Note 7) for their experienced group of theater consumers as  $-1.47$  based on an application of Equation (10) above and the assumption that the average experienced consumer knows 85 percent of the actors and directors (yielding in their model an  $s$  value = 0.85), and that this type of consumer attends the theater an average of 3.87 times per year (so that  $x = 3.87$ ). These parameters along with the estimated logit coefficient of 2.1262 on their knowledge of actors and directors variable (interpreted from above to be  $\delta x / \delta s$ ), yielded the calculation (which they do not explicitly show) of  $e = -1.467$  for the more experienced consumer [see Equation (20) in Seaman (2005)] and  $-1.0$  for the less experienced theater-goer.

It is easy to question the many steps and assumptions required to impute this result. But Lévy-Garboua and Montmarquette viewed it as fully expected from their model when one views experienced theater-attenders are those who have “completed their learning process” after experiencing many cases of pleasant surprises generating high  $s$

values, high quality adjusted quantities of arts consumption and low personalized prices of consuming the arts. In their view, had theater experiences generally led to unpleasant surprises and a reduction rather than increase in the subjective quality assessments, the result would have been price inelasticity.

This approach of endogenizing the discovery of consumer utility functions via a learning-by-consuming process and testing the model without the use of a lagged consumption variable is unique, but is not the only empirical approach that might be taken. For example, Brito and Barros (2005) modeled the dynamics of demand and prices in a learning-by-consuming model and suggested that the standard empirical approach of using a lagged consumption variable is fully consistent with their version of that model. The unique finding of price-elastic demand for experienced consumers is also not an inevitable result of learning-by-consuming processes inasmuch as Brito and Barros (2005) derived the contrary expectation of both low income and own-price elasticities (p. 104) due to the way that flows of cultural good consumption interact with the decay of the stock of culture in response to an exogenous “shock” in the relative price of cultural goods in their particular dynamic model (pp. 97–98).

### 7.3. *Rational addiction*

Finally, past consumption can positively influence present and future consumption through rational addiction.<sup>47</sup> The key assumption here is consistent forward-looking behavior where consumers maximize an intertemporal utility function and are willing to sacrifice current utility for future utility by making investments in human capital (either general education, or more targeted training). The opposite of myopic habit formation, this model actually requires the least of arts managers, who in the extreme need only maintain generally high quality standards; in so doing, it is suggested, they will not endanger the perception that the arts are one of those goods capable of yielding future utility through the sacrifice of current consumption in order to invest in the creation of human consumption capital. In contrast to the learning-by-consuming approach, the rational addiction model combined with specific consumption capital in a household production framework has a different modeling setup and can generate different implications, some of which were described earlier in the Section 6 discussion of Gliberman and Book (1977).

Despite its popularity in the arts literature, attempts to verify a more technically precise version of this framework have not always succeeded. For example, while applied to cinema rather than the performing arts, Cameron (1999) found his results in testing for this phenomenon problematic and offering “little support” for the rational addiction model (p. 619). This is in large part due to the additional requirements that must be met in finding evidence for addiction in the more technical version of that model.

<sup>47</sup> See Stigler and Becker (1977), McCain (1979, 1981, 1995), Spinnewyn (1981), West and McKee (1983), Becker and Murphy (1988), Villani (1992), Cameron (1999), Lévy-Garboua and Montmarquette (2003).

That is, the household production element of the rational addiction model generates an important distinction between shadow-price elasticities linked to arts appreciation and market-price elasticities related to observed arts attendance. But the intertemporal non-separability of the utility function feature of the model stresses the rate of time preference of consumers, i.e. their ability to be far-sighted rather than short-sighted. In this context, findings of “too-high” a rate of time preference is inconsistent with the premise behind the rational addiction framework.<sup>48</sup> Tests for rational addiction also involve investigating the relationship between the interest rate and the discount rate. Lévy-Garboua and Montmarquette (2003) provided a concise description of the rational addiction perspective, and contrasted it to their own learning-by-consuming approach.<sup>49</sup> A key result derived from the formal analysis is that the “relative shadow price” of arts appreciation (as distinct from ticket purchases) will normally decline over time with the accumulation of arts-specific capital, be it from consumption or training [Smith (1998)].

There are two especially important implications of the rational addiction framework:

- While it might be thought that with a declining relative shadow price of art appreciation over time its quantity will grow as well, that will only definitely occur when the rate of time preference does not exceed the interest rate; more impatient consumers, therefore, may actually reduce their consumption of arts appreciation over time even in the face of a declining relative shadow price.
- As is true of any variation in the household-production-consumer-choice framework, even if consumption of the unobserved art appreciation ( $X$ ) rises over time, there is no guarantee that the observed attendance at arts events ( $x$ ) will similarly increase over time. As usual this is because with the cultivation of taste, increasing arts experience makes each unit of  $x$  more productive in generating a unit of  $X$ . Since less  $x$  can generate a given  $X$ , the net result is uncertain, although again  $x$  is more likely to rise the lower is the discount rate and the higher is the interest rate.

Finally, this critical distinction between arts appreciation  $X$  and arts attendance  $x$  provides one of the theoretical justifications as to why one might expect an inherent bias toward relatively low price elasticities of demand for arts attendance relative to other goods at comparable explicit prices. That is, there is a distinction between the shadow price elasticity of  $X$  in any time period ( $E$ ) and the market elasticity of arts attendance ( $e$ ). Specifically, the shadow price elasticity is always higher than the market price elasticity due to the effect of positive addiction. The key point is that there is nothing inconsistent between having a price-elastic demand for art appreciation  $X$  and a market-price-inelastic demand for arts attendance  $x$  [Lévy-Garboua and Montmarquette (2003, p. 206)].

<sup>48</sup> McCain (2003, p. 448) notes that efforts to test the rational addiction model empirically have not been “entirely satisfactory”, and that at times findings of high rates of time preference have been found to be “implausible” by advocates of the rational addiction model.

<sup>49</sup> Borrowing also from the original contributions of Spinnewyn (1981), Stigler and Becker (1977), and Becker and Murphy (1988).

In summary, while many arts demand studies have improved the performance of their estimated equations by including a one-year lagged dependent variable capturing the effect of past consumption on future consumption, that is not the only, or even necessarily the preferred, way to capture the dynamic effect of taste cultivation in arts demand analysis. Even if such a lagged variable is introduced, the underlying theoretical justification for its inclusion is a complex subject, with competing approaches having somewhat different implications, especially for arts management strategy and for the future growth of performing arts demand.

## 8. The product and geographic market problem: Substitutes and complements

In addition to the unresolved issue of whether the arts are price-inelastic luxury goods, Lévy-Garboua and Montmarquette (2003) have found no definitive evidence in the literature regarding close substitutes for the performing arts. While Withers (1980) estimated fairly strong cross-price elasticities for his “reading and recreation component of the CPI” as noted earlier, the t-statistics were not especially strong, and a time-series study of the aggregate performing arts in the United States is not well-adapted to capture localized competitive effects, or to distinguish among the sub-categories of the performing arts. Corning and Levy (2002) stated the dilemma perfectly when they observed that the most direct competitors of their three southern California theater venues are the Santa Barbara City College Theater Group, the Ensemble Theater Company, and Civic Light Opera, but “unfortunately insufficient price data were available to construct a useful variable” (Note 5, p. 234).<sup>50</sup> They were thus forced to fall back on the “recreation component of the CPI” and found that none of those variations had any “measurable effect in any configuration and was dropped” (p. 227).

Some suggestive evidence on substitutes was developed by Lévy-Garboua and Montmarquette (1996), although no variable in their vast database really captured the price of substitutes. One variable reflected the perceived quality of available substitutes, and the relatively strong negative coefficients for this variable reflected a potentially broader product market than theater alone. This conclusion is also supported by their finding of a negative effect of regular magazine and journal reading on the frequency of live theater attendance, suggesting that those two forms of intellectual stimulation and entertainment are partial substitutes. Certainly, evidence regarding substitutes and complements need not stem solely from cross-price elasticities. Available quantities of potential substitutes like television programming or proxies reflecting technical improvements in the quality of such substitutes [e.g., Bonato, Gagliardi and Gorelli (1990), Pommerehne and Kirchgassner (1987)] may be revealing indicators of substitution relationships. Heilbrun

<sup>50</sup> In a non-regression case-study of marketing strategies for the Los Angeles Music Center, Kaali-Nagy and Garrison (1972) identified eight potentially competing Southern California attractions: Marineland, Busch Gardens, Disneyland, Knott’s Berry Farm, the L.A. Zoo, the San Diego Zoo, Huntington Library and the L.A. County Museum.

(1997) found evidence that the popular arts have had notable negative effects on the high arts by examining the press coverage of both art forms over time in the *New York Times*.

Regarding cross-price evidence, it was not always presumed that arts and entertainment alternatives would be substitutes as opposed to complements for the high arts. Nevertheless, Felton's (1992) explicit expression of neutrality regarding the expected signs of such variables is rare<sup>51</sup> and most discussions of the empirical results reflect the expectation that gross substitution should dominate.<sup>52</sup> For example, Ekelund and Ritenour (1999) were troubled by the unexpected negative coefficient on their variable for price of audio recordings, and stressed that it was statistically significant at "only" the 0.10 level.

While some version of movie price is the most popular cross-price choice,<sup>53</sup> such variables have performed very poorly.<sup>54</sup> In fact, it can be safely concluded that there is little empirical evidence that movies are effective substitutes for the performing arts. Only Withers (1980) and Gapinski (1986) had any real success in capturing positive cross-price effects for the performing arts, with Withers being the only remotely successful application of an aggregated recreation or entertainment price index. However, Felton (1992) generated evidence of a positive cross-price effect of symphony prices on attendances for large-budget but not small-budget ballet companies, and Goudriaan and de Kam (1983) found evidence of a positive effect of education and recreation expenditures on symphony, but not theater, attendance in their time-series model.

While Gapinski (1986) has understandably received all of the attention regarding cross-price evidence, it is noteworthy that his earlier attempt to find cross-price effects between cinema and recreation price indices and attendance at the Royal Shakespeare Company was not successful [Gapinski (1984)]; moreover his later study of resident versus tourist demand using the same 13 arts companies as in his 1986 study (and with a more inclusive variation of his 1986 substitute-price variables) also failed to generate any significant results. It is interesting that the Gapinski substitute prices that omitted any "intra-art form" components (1986) were more successful than the versions that

<sup>51</sup> She stated that "it was deemed equally likely that the two [alternative arts] experiences would be substitutes or complements" [Felton (1992, p. 4)].

<sup>52</sup> At times, this expectation was clearly dependent on the particular measure being used. For example, in Greckel and Felton (1987) their "poor" measure of substitute prices was dropped (i.e. the entertainment component of the CPI), not only due to weak and statistically insignificant coefficients but because in most equations the "the sign of the coefficient was negative, indicating complementarity instead of substitution" (p. 64).

<sup>53</sup> The enthusiasm for cinema prices in performing arts demand equations is not reciprocated in cinema demand analysis, where performing arts prices never appear. One possible reason, beyond mere measurement problems and data availability, is that while cinema and theater may appear to be excellent substitutes, the greater availability of cinema in all parts of a country compared to the more concentrated location of theater in only the major urban centers, may reduce the practical degree of such substitutability. Fernández Blanco and Baños Pino argued this for the case of Spain (1997, p. 62).

<sup>54</sup> Only Touchstone (1980) came close to an expected result in her symphony equation and the coefficient in the opera equation was negative.

included such intra-art form prices (1988), although the two studies were not otherwise identical. Also, it is puzzling that citations of [Gapinski \(1986\)](#) almost always claim that he found strong evidence of substitution across artforms and stress the fact that his cross-price elasticities are greater than 2.0 in some cases, despite the fact that the cross-price elasticities are below 0.20 for four of his 13 companies, and 0.65 or lower for all but three of his companies (with two of the dance companies being as low as 0.21 and 0.28).<sup>55</sup>

Furthermore, [Gapinski](#) himself stressed that the two modern dance companies having those unusually high cross price elasticities were those with the lowest attendance and most “heavily contemporary” of all the companies in his study (p. 22), and that the “clearest pattern to emerge” was that “a price change by a single company alone has minor impact on a second company” and that the “greatest attendance response to a price maneuver occurs for the initiating firm itself” (p. 23), which is especially noteworthy given the quite inelastic own-price elasticities that average less than  $-0.30$  for all art forms [see [Gapinski \(1986, Table 1\)](#)]. However, among the rarely discussed results are his projections of the effects of price changes by rivals acting together rather than unilaterally (p. 24), with one of his examples being a loss of 4800 patrons annually for one of his theaters if all other non-theater companies reduced their prices by 10 percent (an attendance decline that he argued is almost twice what could be generated by own-price changes by that particular theater itself).<sup>56</sup> Nevertheless, the overall message of his path-breaking study is that price interdependencies among performing arts firms in specific geographic markets are potentially important, and that the focus should clearly be on further efforts to find evidence of the degree to which “the lively arts substitute for the lively arts”. To date, no successful replications or extensions of the [Gapinski \(1986\)](#) analysis have appeared.

There is little direct cross-price elasticity evidence regarding one aspect of complementary goods, i.e. the effect of higher prices for transportation (for given distances from the venue), parking, child care, dining and other components of the full price of an arts performance visit. The only exceptions are [Carson and Mobilia \(1989\)](#) and [Lévy-Garboua and Montmarquette \(1996\)](#), where some indirect evidence is presented that such costs (or perceived costs) can have negative effects on performing arts consumption.

The relative success of [Gapinski \(1986\)](#) in studying the particular geographic area of London raises an issue not normally addressed: what is the geographic scope of the market for the performing arts? Complementary television programming or other forms

<sup>55</sup> The rare exception of a more accurate portrayal of the [Gapinski \(1986\)](#) results is [Abbe-Decarroux \(1994\)](#), who lamented his inability to find sufficient real income and substitute price data to include in his study of Geneva theater, but noted that various studies have found that performing arts demand is income insensitive, and that [Gapinski \(1986\)](#) also found that it is substitute-price insensitive as well (Note 6, p. 103).

<sup>56</sup> It is possible, but unlikely, that all other firms would independently change price by roughly the same magnitude, and [Seaman \(2004\)](#) provides evidence that successful collusion, whether on price or other issues, has no doubt been rare in the performing arts.

of the media arts [Waterman, Schechter and Contractor (1991); Heilbrun (1993)], as well as touring arts companies [Escalera (2002)], can extend the geographic scope of the relevant market far beyond any localized geographical region.<sup>57</sup> Furthermore, while Verhoeff (1992) confirmed the negative effect of distance on performing arts attendance [see also Zuzanek and Lee (1985)], he found surprising variability in the distances that people travel to performances in the Netherlands.

Despite this potential evidence that the geographic scope of effective performing arts markets need not be as localized as is generally assumed, Forrest, Grimes and Woods (2000) provide the most sophisticated econometric evidence regarding the important negative role of distance traveled in performing arts demand. They applied a zonal travel cost model to data regarding the Royal Exchange Theater in Manchester. Their price variability was derived from the notion that all potential consumers faced different effective prices due to the varying costs of traveling to the venue. Using distance as a proxy for travel costs, they also assumed that an increase of 2.66 km is equivalent to an increase of £1 in the cost of attendance, and that symmetrically a £1 increase in ticket prices is the equivalent of shifting the population of each zone outwards from the theater by 2.66 km. They then calculated for each of 20 zones the change in total visitor numbers with respect to a change in ticket price, and derived a price elasticity of demand of  $-1.24$ , which they interpreted as being reasonably close to the revenue-maximizing pricing strategy when marginal cost in the non-capacity-constrained case is nearly zero. Interestingly, calculations of elasticity for individual zones (with prices different from the mean) showed inelastic demand in areas with the largest values for their variable controlling for educational level but elastic demand elsewhere.<sup>58</sup>

Finally, there is a small literature that addresses the substitutes/complements issue not by inserting the prices or quantities of consumption alternatives into a regression equation on arts attendance, but by examining the correlation between the estimation errors corresponding to consumers' demand equations for different types of entertainment in an effort to answer the question: "Are high arts and popular arts (or sports) consumers the same people?" These studies by Prieto-Rodríguez and Fernández-Blanco (2000) regarding classical and popular music, and Fernandez-Blanco and Prieto-Rodríguez (2000) and Montgomery and Robinson (2005) regarding live sports and the live arts, generate some evidence that the high and low arts may be complements instead of substitutes, but provide conflicting results regarding the relationship between arts and sports consumption [Seaman (2005, pp. 120–122)].

<sup>57</sup> Seaman (2004) reviews this evidence.

<sup>58</sup> They also addressed the issue of the possible endogeneity of residential location, whereby people with strong arts demands would locate in close proximity to arts venues. While it is known that arts attendance is higher for those living in urban areas or in locations with relatively large populations, that issue is rarely addressed as a simultaneity problem. Bajic (1985) also found some evidence of theater location being a factor in the housing choices of those with especially strong theater demand in Toronto, but this result is hardly typical of arts consumers, much less the general population.



## 9. The product quality problem

### 9.1. Modeling quality

Throsby's attempt to address the quality issue in arts demand studies (1990; a reprint of a 1982 paper) represented an important step in correcting a prior mis-specification of arts demand equations. He developed the distinction between an *objective* measure of individual plays (based on repertoire classification), and so-called technical variables (standards of source material, production, acting, and design) that actually depended upon the *subjective* assessments of press reviews. There is another critical distinction in the arts quality literature: whether the primary focus is on a time-series or pooled analysis of the quality of a large number of performances by a small group of arts organizations, or on the overall quality of a large number of arts organizations themselves. Two things are clear. First, the most studied artform by far has been theater [Throsby (1990), Jenkins and Austen-Smith (1987), Dobson and West (1989), Abbé-Decarroux (1994), Corning and Levy (2002), and Urrutiaguer (2002), with Krebs and Pommerehne (1995), a mixed case that primarily focused on theater but with a database that also includes opera and other art forms]. Second, the overwhelming focus has been on attendance per performance related to quality variables applied to the individual repertoire as opposed to the overall quality of arts organizations. Interestingly, the few studies that did not focus on theater [Greckel and Felton (1987), Felton (1989), and Luksetich and Lange (1995)] primarily used non-Throsby type variables that did apply to the organizations themselves (see below).

The Throsby (1990) model postulated a subsidized non-profit theater management choosing price, season length, and quality attributes of its productions, so as to maximize a managerial utility function containing those quality attributes and the percentage of seats in its venue that are filled with paying customers. Paid attendance demand is a straightforward function of average price per seat, season length, venue capacity, and an  $n$ -vector of quality characteristics. Data were available from three Sydney theater companies. The key innovations were in generating the variables to include in the quality characteristics vector. Five characteristics  $q_1, \dots, q_5$  were defined as repertoire classification, standard of source material, standard of production, standard of acting, and standard of design, respectively. The repertoire classification variable,  $q_1$  could be defined using objective criteria based on four groupings of plays that would be essentially non-controversial: A = a "classic" written before 1900; B = written after 1900 by a well-known author (from the audience perspective); C = written after 1900 by little or unknown authors (from the audience perspective); D = entertainment, revue, and musical. Class D was defined as the omitted benchmark if all three other repertoire classifications enter the equation, and a dummy variable was created for each of the A, B, and C groupings which was set = 1 if a play fell into that particular class, and 0 if it did not.

The development of variables  $q_2$  through  $q_5$  was more challenging and was based on an assessment of the "subjective" opinion of press reviews defined in terms of how



Table 2  
Demand function estimates for three Sydney companies, 1974–1978

T	Con	Price	Cap	Repertoire class			Technical standard				$\sum q_i$	$R^2$
				Cl A	Cl B	Cl C	Mat	Prod	Act	Set		
1	1.67 (3.2)	-0.41 (-0.7)			0.04 (0.7)		0.18 (0.7)	-0.09 (-0.4)	0.90 (2.2)	0.21 (0.6)		
1	1.14 (2.0)	-0.20 (-0.3)			0.03 (0.5)						0.87 (2.1)	0.21
2	-0.85 (3.3)	0.66 (1.9)	0.92 (11.4)	-0.01 (-0.1)	-0.04 (0.7)	-0.12 (2.2)	0.12 (0.8)	0.20 (1.1)	0.20 (1.1)	0.30 (1.8)		0.51
2	-1.38 (5.1)	0.58 (1.9)	0.94 (13.2)	-0.04 (0.6)	-0.05 (0.9)	-0.12 (2.3)					0.88 (4.2)	0.51
3	0.43 (1.3)	0.30 (0.8)	0.63 (5.9)	0.06 (1.2)	0.07 (1.6)		0.06 (0.4)	0.04 (0.4)	0.04 (0.4)	0.04 (0.3)		0.71
3	0.31 (0.9)	0.28 (0.8)	0.63 (6.2)	0.06 (1.7)	0.07 (1.8)						0.21 (1.3)	0.71

Source: Throsby (1990, Table 1).

Notes: Theater (T) 1 = Ensemble (180 seat theater-in-the-round in a converted boat shed in a harbor-side suburb); 2 = Nimrod (300 seat converted factory in inner suburbs); and 3 = Old Tote (principal state drama company in several venues; already was closed when the study was completed); Con = constant term; Cap = venue capacity; Cl (class) A, B and C are as defined in the text prior to the table; and the technical standards of source material (Mat), production (Prod), acting (Act) and design (Set) are the  $q_2$  through  $q_5$  variables. Absolute values of the t-statistics are in parentheses.

well any play met high standards defined over the four “technical” dimensions listed above. A cardinal scale 1 to 5 was created. Importantly, he also summed the  $q_2$  through  $q_5$  variables to obtain a single “composite standard” (p. 73). This was rationalized as an attempt to overcome some of the variability in individual assessments, but can be justified on its own merits as reflecting something akin to the overall impression that a play would make on a viewer who may not even be thinking in terms of the four separate criteria. This entirely separate variable (defined as “sum” =  $\sum q_i$ , for  $i = 2, \dots, 5$ ) was substituted for the individual  $q_2$  through  $q_5$  variables in an alternative specification of the model for each of the three theaters. The results are reported in Table 2, although coefficients are rounded to two decimals and the column headings are slightly modified compared to the original. Except for the repertoire dummies, the double-log specification was used.

Since only the Nimrod theater (theater 2) offered plays in all three of the “non-popular” (i.e. class D) repertoire classifications entered as dummies, its results regarding those three types of plays should be interpreted relative to entertainment, revue and musical plays. Its audiences revealed a strong distaste for class C plays (those by little known authors), but neither classics nor twentieth century plays by well-known authors had significant effects. The type of play also had no effect on Ensemble audiences (theater 1), but Old Tote audiences reacted favorably to both classics and well-

known plays relative to those by little-known authors (type C, the omitted base case for theater 3). Thus, there is some evidence that the type of play is an important variable to include in theater demand equations.

There were also mixed but generally supportive results regarding the more subjective rankings of the four technical performance criteria. The most noteworthy finding was that for all three theaters the aggregated quality variable generated much more statistical significance in the positive coefficient estimates than for each quality variable entered separately (although still not significant for theater 3, the Old Tote). The weak results for the Old Tote possibly reflected the high proportion of subscription seats sold by that theater, producing what Throsby called a “captive audience” effect (p. 75).<sup>59</sup> Considered individually, the standard of acting had by far the strongest effect on the Ensemble theater (theater 1), with set design dominating audience choices for the Nimrod (theater 2). Throsby found this first result fully consistent with the Ensemble’s known commitment to acting (p. 75). Taken as a whole, these “tentative and qualified” (p. 81) findings have been viewed as confirmation of the importance of controlling for quality in arts demand studies.<sup>60</sup>

Following Throsby (1990), the inclusion of some form of both objective and subjective measures of program classification into arts demand studies has become relatively common, although the results continue to be mixed. These approaches and results are reported below in the following sections.

## 9.2. Studies using objective quality criteria

We review here a number of studies that have assessed quality on the basis of objectively observable criteria, beginning with that of Jenkins and Austen-Smith (1987) who had only limited success in finding a positive effect on English provincial theater attendance of “less esoteric” programming (defined as comedies, thrillers and musicals, in contrast to so-called serious drama). While the effect was positive and statistically significant, it was not especially economically significant. They estimated that an increase in the mix of less esoteric programming from 50 percent to 60 percent would increase demand by only one percent. On the other hand Greckel and Felton (1987) did not attempt to characterize the content of programs, but included even more objective measures without success. Their dummy variable to capture the shift of the Louisville Symphony

<sup>59</sup> This is reminiscent of the Hjorth-Andersen (1992) suggestion that if subscription sales dominate total ticket sales (as he found in his forecasting equations for Danish theater), the financial success of a whole season is known before it even starts.

<sup>60</sup> In addition to the widely varying  $R^2$ s for the three equations, the primary discordant note was in the behavior of the ticket price coefficients, not statistically significant and with conflicting signs for the Ensemble and the Old Tote, but positive and significant for the Nimrod. Throsby suggested that this reflected a demand shift for the Nimrod during a time when the real price of admission was only gently rising due to the Nimrod’s policy of keeping prices low to encourage audience development rather than commercial success [Throsby (1990, p. 75)]. This justification is an extension to the positive elasticity case of the downward bias in estimated negative price elasticities created by low pricing strategies.

Orchestra to a full-time orchestra, and variables for the number of programs, and the number of performances (both intended to reflect desirable product variety) failed and were dropped from their reported equations. Two variables that did work well were idiosyncratic to the Louisville Orchestra: (1) a dummy variable to capture the effect of an “unpopular conductor” (i.e. Jorge Mester) was introduced to assess the negative impacts of the frequent absences of this primary conductor during his final two years, and the unpopularity of Akiro Endo, his successor (the authors surprisingly identify an estimated elasticity of  $-0.30$  despite this being a dummy variable); and (2) a concert hall capacity variable with the unique interpretation of capturing the positive effect of the shift of the orchestra to new and far superior facilities that also included a larger hall. In subsequent studies *Dobson and West* (1989) found no significant effect on Atlanta theater audiences of the type of play or the day of the week of the performance; similarly, *Felton* (1989) found statistical significance for her opera popularity ratings variable in only one of her opera company equations (the Kentucky Opera Association), and concluded that in other cities programmatic content had little effect on subscriber attendance. *Luksetich and Lange* (1995) had no success with their attempt to link total expenditure per symphony performance to attendance via a quality argument.

While *Abbe-Decarroux* (1994) included a variable measuring subjective press reviews, he perhaps followed *Throsby* (1990) as closely as anyone in also introducing eight quite objective dummy variables (although not aggregating them into a single variable). He also estimated his seven-year (64 productions) Geneva theater demand function for both full-price and reduced price audiences (as well as reporting the total demand equation). Some variables performed well; others did not. “Home productions” (those produced by the institution itself) had significantly negative effects on per performance admissions (explained as being due to the higher frequency of such plays being performed), while well-known authors, producers and casts had separately positive and generally statistically significant effects across all equations. But surprisingly, the fame of the play, and whether it was a classic (written before 1900) or a modern play (after 1990; deceased author) had no real effect on either audience segment, but audiences clearly rewarded novelty, with “atypical” plays (circus, revue, collection creation or other) having strongly positive effects on per performance attendance.

*Krebs and Pommerehne* (1995) considered but rejected various quality measures, and tried to capture the popularity of various arts productions by measuring the “share of productions with many performances out of all performances” (p. 25). Specifically, they constructed a proxy for highbrow versus lowbrow theater productions by measuring the “ratio of works with more than 75 performances to all works played in a season”, which they viewed as a proxy for more popular lowbrow productions that would survive longer than highbrow plays. Given the recognized weaknesses of this construction, it is remarkable that this variable performed fairly well – a positive coefficient, although not quite statistically significant (unless income is omitted from the equation). The derived “lowbrow” elasticity of arts demand implied that a 10 percent increase in the share of such works would increase paid attendance per capita of the population by one percent. This lowbrow inelasticity result is consistent with the low “less-esoteric-programming”

elasticity result in [Jenkins and Austen-Smith \(1987\)](#) and suggests that even when this dimension of repertoire quality has an effect on attendance, it is not as high as has been generally expected. However, since the [Abbé-Decarroux \(1994\)](#) definition of “atypical play” seems to include some more popular rather than eccentric content, its strong positive effect in that study is potential counter-evidence.

A large number of objective quality variables were introduced by [Corning and Levy \(2002\)](#) in their study of demand for live theater with market segmentation and seasonality. They found evidence that programmatic content had limited effects on single-ticket theater attendance, varying notably across individual theaters and strongest for the unique case of the musical *Tommy*, which in the case of one Southern California theater location would increase per performance attendance by over 200 in contrast to a performance of Shakespeare *ceteris paribus* (p. 230). They also found more evidence than did [Dobson and West \(1989\)](#) that scheduling has some effect on attendance, with the weekend dummy variable having the strongest positive effect across all venues. However, their seasonality results were generally weak. Their most fascinating objective variable is one that would seem to pre-judge the issue: “Flop” was a dummy variable = 1 when the average total attendance for a play at its first location prior to its shifting to another venue was less than 50 percent of capacity; remarkably, the coefficient on this variable was positive in two venue equations (statistically significant in one) and only negative and significant in the outdoor summer season venue. The fact that seemingly clear evidence that a play is a failure cannot reliably reveal later negative effects on attendance may be the best evidence of the challenge in capturing the effects of quality in arts demand studies.

### 9.3. *Studies using subjective quality assessments*

In contrast to the above efforts to capture objective quality effects, some subsequent studies focused on the determination of subjective perceptions of quality, typically focusing on the role that various expert critics have on the perceptions of the lay public. For example, [Abbé-Decarroux \(1994\)](#) supplemented his four repertoire and four “fame” (author, play, producer, cast) objective variables with a qualitative press review variable scaled nearly identically to [Throsby \(1990\)](#), adding a sixth category for “excellent”. This variable was positive and highly statistically significant in both full-price and reduced-price equations. [Corning and Levy \(2002\)](#) also supplemented their inventory of programmatic, scheduling and seasonal variables with a newspaper review variable on a five-point scale [adding some additional detail to the parsimonious [Throsby \(1990\)](#) ratings; their Table II, p. 227]. Due to some format changes in the two publications used in their analysis, some productions were not reviewed at all, which was registered with a “no-review” variable. The quality of such press reviews had a positive and statistically significant effect on per performance attendance in only one case (the summer outdoor festival venue). Interestingly, the total absence of any review of a production actually had a positive effect in each of the three venue equations, with reasonably high although not statistically significant t-statistics.

#### 9.4. An extended model

Urrutiaguer (2002) has provided the most important extension of the Throsby (1990) analysis. He modified the Throsby model in two ways, one somewhat technical, but the other more important conceptually. The technical adjustment was in changing the dependent variable from a measurement of seasonal attendance for each year to per performance attendance, by dividing the Throsby dependent variable by the number of performances. Urrutiaguer justified this in his case as necessary to avoid heterogeneity in the size of theatrical institutions among his much larger database of 87–104 theaters, and to reduce the risks of heteroskedasticity (p. 187). More fundamentally, he noted that when he applied his model to the demand for individual shows in 1995, the adjusted  $R^2$  was quite low (about 0.13; the Throsby  $R^2$ s ranged from 0.21 to 0.71). He suggested that the weakness of this result should not be surprising inasmuch as the theater itself rather than just individual shows is a “much more appropriate” level at which to examine variance in demand, which can be caused by an organization’s program, auditorium comfort and the overall image of an individual theater.

Urrutiaguer’s four repertoire classifications were updated variations of Throsby (1990) adapted to France. An important modification was required to make the resulting dummy variables applicable to a theater rather than an individual play. A theater-reviews variable was based on an analysis of the opinions in three publications, again aggregated to apply to an individual theater rather than just specific performances. The final step was to recognize that unambiguous reviews have more impact than those with “nuances”, and any dummy variables entered into the estimating equation should reflect this. Urrutiaguer also developed a “weak and high centrality” measure to capture the potential effects of a unique quality proxy in France: the weight in a theater’s programs of shows produced by what he calls “directors-cum-managers”. Finally, assuming that public recognition of a theater via tax-financed subsidies is a quality proxy, he specified a dummy with value = 1 if the yearly growth of subsidies was more than 8 percent, and = 0 otherwise. Two such dummies were defined, one for state and one for local subsidies.

Despite the sophisticated technical innovations in constructing the drama review and centrality dummy variables, and the important conceptual shift toward a focus on the theater rather than the individual play, Urrutiaguer (2002) originally had limited success in generating significant results for his quality variables, an outcome consistent with the frequently mixed experience of prior studies. So he went on to test the hypothesis that different portions of theater audiences have contrasting perceptions of reputation, especially in how they weigh the reputations of drama critics and those of artistic directors. Thus, the full theater sample was segmented into two groups: a Group I of 40 institutions for which the audience shares the scale of judgments of drama critics and a Group II of 47 institutions whose audience have more trust in the artistic reputation of directors-cum-managers and who either ignore or do not share the judgments of theater critics. Therefore, the signs on certain estimated quality coefficients were expected to differ between Group I and Group II regressions.

The effect of this disaggregation into two groups was dramatic. The adjusted  $R^2$  for both groups was very high ( $>0.824$ ). While the price coefficient was enigmatic, positive in both equations and strongly significant for Group I, and only one objective repertoire quality variable was significant (strongly negative for foreign contemporary plays for Group I), the other quality variables generally performed well. Also, consistent with other studies using a one-year lagged dependent variable, that effect was strongly positive in both equations, as is the venue capacity coefficient (although stronger for Group I). The regressions confirmed the expected opposition of the signs on the reputation variables for the media and art director, except for the weak centrality variable, with both groups have highly significant but opposite coefficients on the drama-reviews variables. This suggests that the effect of drama critics can be negative as well as positive, at least for some segments of the audience, which is a more encouraging result than the lack of significance that such variables have sometimes had in other studies.<sup>61</sup>

Despite the delay relative to the appearance of Throsby's work (which was first published in 1983), Urrutiaguer (2002) significantly advances our understanding of the arts quality issue, even if he did not incorporate Throsby's aggregated composite of different technical standards variable into his extensive analysis.

## 10. The role of socioeconomic factors versus life-style determinants of arts demand

### 10.1. Lifestyle determinants

Andreasen and Belk (1980, 1981) are best known for asserting that life-style factors, attitudes and socialization to the arts are more reliable predictors of attendance among "marginal attenders" (at least for theater and symphony in the southern US) than are demographic and socioeconomic variables. A notable feature of Andreasen and Belk's provocative assertion is that it was not only founded upon the derivation of statistically significant univariate correlations of 56 independent variables with the likelihood of attendance, but was further confirmed using step-wise regression analysis. As expected, multicollinearity problems were severe among their 56 independent variables that included all standard socioeconomic determinants (education, gender, income, occupation, and age), but also variables as diverse as six "general life-style" dimensions (e.g., optimism/hedonism and traditionalism), six "leisure life-style" group characterizations (e.g., passive homebody, culture patron, inner-directed self-sufficient), life-cycle variables, and various socialization proxies.<sup>62</sup> As a result, only six variables in the step-wise regressions were found to add significantly (at the 0.05 level) to the prediction of

<sup>61</sup> Including Urrutiaguer (2002) prior to segmenting the theater sample by audience reputation perceptions.

<sup>62</sup> Levantal (1989) included a related list of six "psychographic factors" including "eclecticism" and "independence of opinion" along with "determination to see a particular play".

theater attendance and five to the prediction of symphony attendance, with only modest adjusted  $R^2$  values of 0.279 and 0.289, respectively [Andreasen and Belk (1980, Table 2)].

However, despite those limitations<sup>63</sup> the essential results were striking: not one of those standard socioeconomic variables was a significant predictor of future arts attendance when controlling for attitude, and general and specific life-style factors. Instead, what mattered most as positive predictors of future attendance in the theater equation (based on the standardized beta weights) were: attitude toward attending the theater; being a “culture patron”, a leisure life-style characteristic; interest in live theater when growing up; and theater attendance during the past year. The most important negative predictors were two of the “general” life-style characteristics: “traditionalism”; and “self-confidence/opinion leadership”. There were no significant negative predictors of symphony attendance, but quite similar positive predictors: attitude toward attending the symphony; culture patron; “socially active”, a leisure life-style trait; interest in classical music when growing up; and symphony attendance during the past year. While these results again confirm the importance of previous attendance as a predictor of current attendance, the elimination of variables like education, income and age as significant independent variables was novel.

But have those findings been replicated? Some studies have cited important socialization effects within the family [e.g., Ganzeboom (1989), van Eijck (1997)], and others have carefully distinguished between childhood arts experiences with parents in contrast to those obtained in school [e.g., Abbé-Decarroux (1995), Morrison and West (1986)]. Furthermore, among the determinants used by Lévy-Garboua and Montmarquette (1996) were such attitudinal variables as “appreciates humanity” (which intriguingly had a negative effect on the probability and frequency of attending the theater, but a positive effect on the “satisfaction” derived from attending; Table 1, p. 36). And when one adds variables from non-regression based studies, the variety of such factors expands further. A mere sampling would include: the ticket purchasing habits of friends [Bamosy and Semenik (1981), Kolb (1997)]; years spent residing in the geographical area being studied [Ryans and Weinberg (1978)]; divorced versus widowed status in addition to the more standard single versus married designations, as well as who within the family makes the decision to attend various types of entertainment events [Kaali-Nagy and Garrison (1972), and Upright (2004)]; and networking among women [Kane (2004)].

However, there is no real evidence that the Andreasen and Belk (1980) conclusion regarding the dominance of the socialization “error-term” variables over the other standard demand variables has been replicated. There are three reasons. First, even Andreasen and Belk (1980) found that of their daunting list of non-traditional variables, only a very few survived the step-wise regression pruning process (as described above). Second, other studies of “socialization”-type variables have found that income,

<sup>63</sup> And more; see Seaman (2005, p. 135).



education and age are not eliminated from those equations and often perform quite well [DiMaggio and Ostrower (1990); Lewis and Seaman (2004)]. Finally, from the purely limited perspective of explaining the variance in the dependent variable, extremely parsimonious time-series or pooled studies sometimes do quite well without adding such so-called taste-adjusting variables.

### 10.2. Racial differences

While race has been a surveyed characteristic of the SPPA since 1982, race and ethnicity have been nearly absent from econometric arts demand studies, with only Dobson and West (1989), Gray (2003), and Lewis and Seaman (2004) as exceptions. It is also difficult to find examples using non-US data, with Trienekens (2002) being a rare exception (i.e. the Netherlands).

Dobson and West (1989) found “ethnic background” to have a modestly negative effect on Atlanta theater attendance; Lewis and Seaman (2004) found mixed results, with “Black” being a negative and statistically significant determinant of classical music attendance, but with competing and not statistically significant effects on arts museum (negative) and dance (positive) attendance. Gray’s (2003) logistic regression results based on the 1997 SPPA are of interest inasmuch as, after adjusting for age, income, education, gender, work hours, and music and art lessons, the “Black” variable coefficient was statistically significant and negative only for classical music, opera, and ballet, while being positive for jazz, musical theater, non-musical theater, dance, and museum. Also, he reported *ceteris paribus* positive and statistically significant coefficients for “Hispanic” across all categories, negative “Asian” coefficients for all arts categories except dance and museum, and negative “Indian” coefficients for all arts categories except opera and musical theater. While not all of those coefficients were economically significant (e.g., six coefficients were below 0.10 in absolute value), no other study incorporated that degree of ethnic variety.

The surprisingly weak and mixed performance of racial/ethnic variables in these few studies is noteworthy in light of DiMaggio and Ostrower’s (1990) conclusion that “given the degree of racial oppression and exclusion to which black Americans have been subjected, they participate in the arts at rates and in ways remarkably similar to those of white Americans”, a pattern they call “*differentiation without segmentation*” (p. 772). Utilizing the 1982 SPPA as the data source, DiMaggio and Ostrower posed the key question: since whites are more likely than blacks to exhibit key characteristics linked to demand for the performing arts (higher levels of education, income and prestigious occupations), is there really an independent role for race in explaining the lower arts participation rates of blacks that are revealed by survey data? The four major results of their empirical analysis were:<sup>64</sup>

- the unique negative effect of race on Euro-American high culture arts participation was modest, but statistically significant [roughly consistent with Gray (2003)];

<sup>64</sup> The unique methodological features of DiMaggio and Ostrower (1990) are discussed in Seaman (2005).



- race effects were stronger for arts consumption than for arts production behavior, and stronger for public arts consumption than for private arts consumption via television watching, where in fact the racial effect was nearly absent – a fascinating result consistent with their view that racial discrimination may adversely affect public but not private participation in the arts and that arts consumption but not production is affected by “status competition”;
- being black had strong positive effects on both attendance and enjoyment measures for the non-European-based artforms of jazz, soul, blues and R&B, confirming the obvious point that viewing black arts participation rates as relatively low depends in large part on one’s definition of the arts; and
- the behavior of interaction terms with race and the other control variables was complex and highly variable.

Finally, DiMaggio and Ostrower conducted a detailed analysis of two hypotheses that might explain the remaining racial differences in arts participation. These were a “cultural convergence” model by which any such differences would be expected to diminish with increased interracial peer contact, and a “cultural resistance” model by which increased black/white economic competition would create opportunities for younger, well-educated blacks to embrace minority cultural norms [e.g., (1990, p. 773)]. However, despite their best efforts, they eventually found little systematic evidence to support either of these perspectives.

### 10.3. *Sexual orientation*

The Lewis and Seaman (2004) study of the role of sexual orientation in affecting arts demand relied upon the 1993 and 1998 US General Social Survey (GSS) for data. They confronted the challenging problem of defining sexual orientation by using the self-reported number of male and female sex partners since age 18, and whether one’s recent sex partners have been male, female or both. While they experimented with a variety of definitions (all yielding similar results), they chose the one yielding the largest sample size (5 percent, or 180 of 2188 respondents) in order to reduce standard errors. They coded the LGB (lesbian, gay and bisexual) variable as 1 for those who reported at least one same-sex partner since their 18th birthday, and 0 for everyone else.

Just as a large part of racial arts consumption differences may actually reflect educational, occupational and income disparities, similar control variable problems (along with urban location factors and differential family obligations) complicate the effort to isolate a unique role for sexual orientation.<sup>65</sup> The descriptive data used by Lewis

<sup>65</sup> While there is evidence that LGB’s are more educated, urbanized, and more likely to be childless than heterosexuals, the common notion that their average incomes are also higher is due largely to non-representative samples of wealthy gay men and lesbians. In fact, controlling for education, gay men earn 15 to 30 percent less than straight men of the same age, and the evidence on lesbian versus heterosexual women is mixed. Yet, there is indeed evidence that gay male couples have higher income (especially disposable income) than married straight couples, with lesbian couple earnings the lowest of the three groups; see Lewis and Seaman (2004, p. 525 and related citations).

and Seaman yielded powerfully suggestive results. With no demographic or other controls, substantially higher percentages of LGBs compared to straight respondents had visited or attended an art museum/gallery, or ballet, dance, classical music or opera performance in the prior year, yielding attendance differentials of between 16 and 19 percentage points, with LGBs being about twice as likely to have attended a classical music or dance performance and almost three times as likely to have attended a museum, musical and dance performance, i.e. 17 percent compared to 6 percent [Lewis and Seaman (2004, p. 529)]. Even after controlling for demographic and other variables, the percentage differences between LGB and heterosexual attendance patterns dropped only to 12 (museum), 10 (dance) and 14 percent (music) from the unadjusted respective differences of 17, 16 and 19 percent, respectively (p. 531). The logistic consumption regression equations<sup>66</sup> estimated by Lewis and Seaman revealed strong and statistically significant positive effects on attendance due to being LGB, higher education, higher parents' education, higher income, and living in a city over 50,000 population, with a moderately positive effect due to being Jewish (but not statistically significant in the classical music equation). Being a fundamentalist Protestant had a modestly negative and significant effect for dance performances (and negative but not significant for museums, and almost dropping out of the classical music equation entirely).

As with DiMaggio and Ostrower (1990), after establishing the unique role of sexual orientation, Lewis and Seaman (2004) confronted a similarly frustrating problem in explaining why it exists. Little compelling evidence was found from supplemental tests for any of four explanations: pure demographics; an innate "gay affinity for the arts", tested by also examining arts production behavior such as "make art, play music, perform live or identity as a professional artist"; a reaction to the historical repression of homosexuality; and an ongoing more welcoming environment for LGBs in arts venues than in other public entertainment environments – an explanation that was at least weakly supported.

Summarizing the various studies discussed above, we can conclude that the Andersen and Belk (1980) hypothesis that "life-style" and socialization-type variables eliminate from significance the standard socioeconomic variables of education, income, age and occupation (not to mention own- and cross-price effects) has found little support from subsequent studies. However, any full understanding of the demand for the performing arts clearly cannot ignore this larger variety of complex determinants.

## 11. Data problems

One of the most common laments of researchers in the economics of art and culture deals with the inadequacies of data.<sup>67</sup> Historically, cultural economists may have had

<sup>66</sup> Their dependent variable = 1 for "attended" and 0 for "did not attend" in each of three equations for art museum, dance performance, and classical music. Results did not change using ordered logit when the three dummy variables were summed to get a proxy for "intensity of arts attendance".

<sup>67</sup> Luksetich and Lange (1995) observed that in all of their earlier work on symphony orchestras they had been severely constrained by data limitations, including an absence of usable data for more than one year.

a strong case that their job was made especially challenging by the traditional lack of business savvy among artists and arts organizations that hindered coherent data collection and reporting, together with the tedious but perennial question “what is art?”. Yet without doubt, the seminal impact of Baumol and Bown (1966) was in large part the result of their uniquely rich database in addition to the provocative quality of their analysis. Even then, empirical work in the performing arts in the US was nearly impossible prior to the Ford Foundation’s extensive survey (1974),<sup>68</sup> followed by another by the National Research Center of the Arts, Inc. (1976); their surveys stimulated efforts in other countries to further improve the quality of their own data. It is therefore not surprising that a characteristic feature of many topics in the arts economics literature has been the degree to which basic data gathering and attempts to develop consistent standards of measurement have competed for attention with modeling and analysis.

Conditions have certainly improved in terms of government-financed studies such as those done by the venerable Arts Council of Great Britain but also extending to other countries. Even in the United States the NEA Research Division, despite desperate under-funding, has frequently generated data that have become international benchmarks, as with the *Surveys of Public Participation in the Arts*. These government efforts have been supplemented by an increasingly rich array of private non-profit and university centers devoted to arts research.<sup>69</sup> Furthermore, despite the frustrations of obtaining reliable data from arts organizations, it is also common for authors of arts demand studies to cite the unusual degree of cooperation (sometimes after failed efforts) given to them by individual or groups of arts organizations. Examples include Felton (1989) paying homage to Opera America for rescuing her after receiving assistance from only five of 20 personally contacted opera companies; Schimmelpfennig (1997) getting unusual cooperation from the Royal Opera House Covent Garden; and Abbé-Decarroux and Grin (1992) working closely with three major cultural organizations in Geneva.

One dilemma that is not unique to demand studies in the arts is that a particular data set or research agenda that may overcome one type of missing information problem, or that may address one deficiency in previous research, may be incapable of incorporating other critical demand-determining variables whose exclusion is undesirable in a fully specified model. For example, Globerman and Book (1977) used audience survey data and an explicit household production model to try to improve the estimate of the income elasticity of demand by directly incorporating the role of education in increasing productivity in the consumption of arts activities. However, they lacked ticket price data and “borrowed” long-run price elasticity results from Houthakker and Taylor (1970) to draw certain inferences about the relative productive effect of education on the arts

However, owing to the generosity of the American Symphony Orchestra League in providing them with truly extensive data on a proprietary basis, they were now able to seriously address those limitations (p. 51).

<sup>68</sup> In addition to the descriptive data published by Ford Foundation (1974), more extensive Ford data were also critical for some econometric demand analysis [e.g., Touchstone (1980), Gapinski (1981)].

<sup>69</sup> This phenomenon is especially notable in the United States, where the more modest governmental role in arts financing and policy-making has created a void to be filled by such organizations.

compared to other activities in the aggregate. Lange and Luksetich (1984) obtained data that allowed them to derive price elasticities that varied across three different types of orchestras, but the lack of consumer income data in their regressions raised questions as to the reliability of those results.<sup>70</sup>

Other examples of studies affected by data limitations include:

- Throsby (1990), who addressed the absence of systematic quality variables in previous arts demand studies, but was missing income as well as education data in his analysis;
- Forrest, Grimes and Woods (2000), who effectively incorporated distance traveled into their theater demand analysis but were able to use only education and age as control variables in their equations;
- Schimmelpfennig (1997), who confronted the issue of price elasticity variations as a function of differing seating sections in a ballet theater, but had no other control variables and was forced to use a quite restrictive assumption about product homogeneity across two different ballets;<sup>71</sup>
- Lévy-Garboua and Montmarquette (1996), who provided an unusually rich analytical framework to assess “learning-by-consuming”, supplemented by a database with 58 independent variables, that did not allow the direct measurement of either consumer income or arts admission prices; and
- Lewis and Seaman (2004), who were able to use unique data from the US General Social Survey to address the previously ignored topic of sexual orientation (and religious affiliation) in arts demand controlling for many other demand determining variables, but could not incorporate any arts pricing or quality variables given that particular data source.

Thus, despite the notable improvements in data availability and quality, a weakness of empirical performing arts demand studies has been the difficulty of any one study having adequate measures of all of the critical determining variables.

## 12. Conclusions

Several reasonably clear conclusions can be drawn from the research reviewed in this chapter, including the following:

<sup>70</sup> For example, see Green, Hassan and Johnson (1992) on the bias created by omitting income in demand function estimation.

<sup>71</sup> The Royal Ballet Summer Season that was examined consisted of 16 performances (nine of which included in the estimations) of two full-length works (for 1995, Giselle and Sleeping Beauty). Schimmelpfennig used the highest prices for (eventually) three seating categories. Such prices were the same for each performance of any one ballet, but that price schedule was higher for Sleeping Beauty than for Giselle. Therefore, he was forced to assume that both ballets are homogeneous in order to get the necessary price variation across his observations, an assumption that he recognized (p. 121) was contradicted by the very existence of two different price schedules, but that he viewed as acceptable for his purposes of examining the relationship between actual prices and revenue-maximizing prices.

- Income elasticities of demand for the arts cannot be adequately estimated without separating the real income effect from the opportunity cost of leisure pure substitution effect;
- Quality matters, although the mixed results from various objective and subjective measures suggest that we are not yet sure how best to capture this important determinant of variations in arts attendance and participation;
- Some version of dynamics in consumption, be it passive inertia, learning-by-consuming, or the more complex rational addiction, should be incorporated to better understand arts demand, especially because these distinct notions have quite different implications for the optimal marketing strategies of arts managers;
- The survey study evidence favoring formal education as the most powerful determinant of variations in arts attendance has not been reliably strengthened by regression analysis.

Nevertheless is hard to dispute the [Lévy-Garboua and Montmarquette \(2003\)](#) assessment that critical issues remain unresolved in the performing arts demand literature. This is especially the case regarding the robustness and interpretation of the price inelasticity “consensus”, the limited evidence regarding substitutes and complements, and the complex relative roles of the traditional socioeconomic demand-determining variables and the vast array of specialized lifestyle and socialization factors (or non-traditional socioeconomic variables). Regarding research improvements, Lévy-Garboua’s and Montmarquette’s faith in the importance of well-defined taste cultivation models, larger databases and more sophisticated econometric modeling in improving our empirical models is understandable, but such analytical improvements to date have largely failed to generate substantial changes in the results or new insights about performing arts demand. In the case of sample sizes, for example, despite the fact that some of the studies discussed in this chapter have used very large databases,<sup>72</sup> the nature and validity of some key results in arts demand analysis do not show any systematic relationship to the sample sizes from which they were derived, *ceteris paribus*. Furthermore there is no firm evidence that more complex econometric techniques are necessary to resolve the remaining enigmas in arts demand studies especially when compared to obtaining more disaggregated data. On the other hand it is clear that applying additional theoretical structure to the derivation and interpretation of empirical results is warranted. This is especially necessary regarding the debate about the price elasticity of demand for the arts, and the choice of model specification will be particularly important in future studies exploring this issue.

More generally it can be concluded that more carefully done modeling, when combined with efforts to apply such models to less aggregated data, will doubtless be important to making further improvements in our understanding of the demand for

<sup>72</sup> For example, [Lévy-Garboua and Montmarquette \(1996\)](#) themselves used a sample of 8000 individuals; other examples include [Peterson, Hull and Kern \(2000\)](#), [Globerman and Book \(1977\)](#) and [Montgomery and Robinson \(2005\)](#) with sample sizes of 17,135; 13,750; and 8000, respectively.

the performing arts. Despite the initial suspicion that the determinants of the typical performing arts demand function are obvious, there are surprisingly few arts-demand axioms beyond

- (1) the confirmation that demand curves are negatively sloped (i.e., assuming any contrary results reflect own price being a proxy for quality),
- (2) the performing arts are normal (but not necessarily luxury) goods, and
- (3) that some positive cross-price elasticities can be identified.

As stated by *Lévy-Garboua and Montmarquette (1996, p. 26)* after noting some general limitations of past arts demand studies, “it is nevertheless known that the performing arts are not exempt from the law of demand”. Empirical performing arts demand studies remain a rich area for further sound economic analysis.

## References

- Abbé-Decarroux, F. (1994). “The perception of quality and the demand for services: Empirical application to the performing arts”. *Journal of Economic Behavior and Organization* 23, 99–107.
- Abbé-Decarroux, F. (1995). “Demand for arts and endogenous tastes”. *Revue Économique* 46, 983–992. In French.
- Abbé-Decarroux, F., Grin, F. (1992). “Risk, risk aversion and the demand for the performing arts”. In: Towse, R., Khakee, A. (Eds.), *Cultural Economics*. Springer, Berlin, pp. 125–140.
- Andreasen, A., Belk, R.W. (1980). “Predictors of attendance at the performing arts”. *Journal of Consumer Research* 7, 112–120.
- Andreasen, A., Belk, R. (1981). “Audience development: An examination of selected analysis and prediction techniques applied to symphony and theatre attendance in four Southern cities”. Research Division Report #14. National Endowment for the Arts, Washington, DC.
- Association of College, University and Community Arts Administrators, Inc. (1984–1985). *The Professional Performing Arts: Attendance Patterns, Preferences and Motives*, vols. I and II. Association of College, University and Community Arts Administrators, Inc., Madison, WI.
- Bajic, V. (1985). “Determinants of theatregoing: The effects of the choice of residential location”. *Journal of Cultural Economics* 9, 60–70.
- Bamossy, G., Semenik, R.J. (1981). “Tangible and intangible motivators of performing arts patronage”. In: Bernhardt, K. (Ed.), *The Changing Marketing Environment: Theories and Applications*. American Marketing Association, Chicago, pp. 154–157.
- Barten, A.P. (1992). “Income and price sensitivity in consumer demand analysis”. In: Philips, L., Taylor, L.D. (Eds.), *Aggregation, Consumption and Trade: Essays in Honor of H.S. Houthakker*. Kluwer Academic, Dordrecht.
- Baumol, W.J., Bowen, W.G. (1966). *Performing Arts: The Economic Dilemma*. Twentieth Century Fund, New York.
- Becker, G.S. (1965). “A theory of the allocation of time”. *Economic Journal* 75, 209–227.
- Becker, G., Murphy, K.M. (1988). “A theory of rational addiction”. *Journal of Political Economy* 96, 675–700.
- Blundell, R., Stoker, T.M. (2005). “Heterogeneity and aggregation”. *Journal of Economic Literature* XLIII, 347–391.
- Bonato, L., Gagliardi, F., Gorelli, S. (1990). “The demand for live performing arts in Italy”. *Journal of Cultural Economics* 14, 41–52.
- Book, S.H., Globerman, S. (1975). *The Audience for the Performing Arts*. Ontario Arts Council, Toronto.
- Brito, P., Barros, C. (2005). “Learning-by-consuming and the dynamics of the demand and prices of cultural goods”. *Journal of Cultural Economics* 29, 83–106.

- Bryson, B. (1997). "What about the univores? Musical dislikes and group-based identity construction among Americans with low levels of education". *Poetics* 25, 141–156.
- Cameron, S. (1990). "The demand for cinema in the United Kingdom". *Journal of Cultural Economics* 14, 35–47.
- Cameron, S. (1999). "Rational addiction and the demand for cinema". *Applied Economics Letters* 6, 617–620.
- Carson, A., Mobilia, P. (1989). "Broadway as an industry: The determinants of Broadway attendance 1975–76 through 1987–88". In: Shaw, D.V., et al. (Eds.), *Cultural Economics '88: An American Perspective*. Association for Cultural Economics, Akron, pp. 105–118.
- Coates, D., Harrison, T. (2005). "Baseball strikes and the demand for attendance". *Journal of Sports Economics* 6, 282–302.
- Colbert, F., Nantel, J. (1989). "The market for cultural activities: New approaches for segmentation studies". In: Chartrand, H.H., et al. (Eds.), *Cultural Economics '88: A Canadian Perspective*. Association for Cultural Economics, Akron, pp. 133–140.
- Corning, J., Levy, A. (2002). "Demand for live theater with market segmentation and seasonality". *Journal of Cultural Economics* 26, 217–235.
- Cwi, D. (1985). "Changes in the US audience for the arts". In: Waits, C.R., et al. (Eds.), *Governments and Culture*. Association for Cultural Economics, Akron, pp. 32–42.
- DiMaggio, P., Mukhtar, T. (2004). "Arts participation as cultural capital in the United States, 1982–2002: Signs of decline?" *Poetics* 32, 169–194.
- DiMaggio, P., Ostrower, F. (1990). "Participation in the arts by black and white Americans". *Social Forces* 68, 753–778.
- DiMaggio, P., Useem, M. (1978). "Social class and arts consumption: The origin and consequences of class differences in exposure to the arts in America". *Theory and Society* 5, 141–161.
- DiNardo, J., Tobias, J. (2001). "Nonparametric density and regression estimation". *Journal of Economic Perspectives* 15, 11–28.
- Dobson, L., West, E. (1989). "Performing arts subsidies and future generations". In: Chartrand, H.H., et al. (Eds.), *Cultural Economics '88: A Canadian Perspective*. Association for Cultural Economics, Akron, pp. 108–116.
- Ehrensaft, P. (2001). "The expanding market for orchestral music". *La Scena Musica* 7 (4), December.
- Ekelund, R.B. Jr., Ritenour, S. (1999). "An exploration of the Beckerian theory of time costs: Symphony concert demand". *American Journal of Economics and Sociology* 58, 887–899.
- Escalera, J. (2002). "Conjectural variations oligopoly model in open performing arts market: Statistical data for the Portuguese case". Presented at the 12th biennial conference of the Association for Cultural Economics, International, Rotterdam, June 13–15.
- Felton, M.V. (1989). "Major influences on the demand for opera tickets". *Journal of Cultural Economics* 13, 53–64.
- Felton, M.V. (1992). "On the assumed inelasticity of demand for the performing arts". *Journal of Cultural Economics* 16, 1–12.
- Felton, M.V. (1994/1995). "Evidence of the existence of the cost disease in the performing arts". *Journal of Cultural Economics* 18, 301–312.
- Fernández-Blanco, V., Baños Pino, J.F. (1997). "Cinema demand in Spain: A cointegration analysis". *Journal of Cultural Economics* 21, 57–75.
- Fernández-Blanco, V., Prieto-Rodríguez, J. (2000). "Are live sports substitutes for cultural consumption? Some evidence from the Spanish case". Presented at the 11th biennial conference of the Association for Cultural Economics, International, Minneapolis, May 28–31.
- Fisher, C.G., Preece, S.B. (2002). "Evaluating performing arts audience overlap". *International Journal of Arts Management* 4, 223–241.
- Fisher, C.G., Preece, S.B. (2003). "Evolution, extinction, or status quo? Canadian performing arts audiences in the 1990s". *Poetics* 31, 69–86.
- Ford Foundation (1974). *The Finances of the Performing Arts, Vol. II: A Survey of the Characteristics and Attitudes of Audiences for Theater, Opera, Symphony, and Ballet in 12 US Cities*. Ford Foundation, New York.

- Forrest, D., Grimes, K., Woods, R. (2000). "Is it worth subsidizing regional repertory theatre?" *Oxford Economic Papers* 52, 381–397.
- Fort, R. (2004). "Inelastic sports pricing". *Managerial and Decision Economics* 25, 87–94.
- Frank, R. (2006). *Microeconomics and Behavior*, sixth ed. McGraw Hill, Boston.
- Ganzeboom, H.B.G. (1989). "International comparisons of culture consumption data: An elementary model". In: Waits, C.R., et al. (Eds.), *Cultural Economics '88: A European Perspective*. Association for Cultural Economics, Akron, pp. 109–116.
- Gapinski, J.H. (1981). "Economics, demographics, and attendance at the symphony". *Journal of Cultural Economics* 5, 79–83.
- Gapinski, J.H. (1984). "The economics of performing Shakespeare". *American Economic Review* 74, 458–466.
- Gapinski, J.H. (1986). "The lively arts as substitutes for the lively arts". *American Economic Review* 76, 20–25.
- Gapinski, J.H. (1988). "Tourism's contribution to the demand for London's lively arts". *Applied Economics* 20, 957–968.
- García, J., Rodríguez, P. (2002). "The determinants of football match attendance revisited: Empirical evidence from the Spanish Football League". *Journal of Sports Economics* 3, 18–38.
- Globerman, S. (1978). "Price awareness in the performing arts". *Journal of Cultural Economics* 2, 27–42.
- Globerman, S. (1989). "What we know and don't know about the economics of culture". In: Chartrand, H.H., et al. (Eds.), *Cultural Economics '88: A Canadian Perspective*. Association for Cultural Economics, Akron, pp. 3–20.
- Globerman, S., Book, S.H. (1977). "Consumption efficiency and spectator attendance". *Journal of Cultural Economics* 1, 15–34.
- Gold, S. (1980). "Determinants of arts demand: Some hypotheses, evidence, and policy implications". In: Hendon, W.S., et al. (Eds.), *Economic Policy for the Arts*. Abt Books, Cambridge, MA.
- Goudriaan, R., de Kam, C.A. (1983). "Demand in the performing arts and the effects of subsidy". In: Hendon, W.S., et al. (Eds.), *Economic Research in the Performing Arts*. Association for Cultural Economics, Akron, pp. 35–44.
- Gray, C.M. (1998). "Trends in arts participation, 1982–1997". Report to the National Endowment for the Arts. Washington, DC.
- Gray, C.M. (2003). "Participation". In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham, UK, pp. 356–365.
- Greckel, F., Felton, M.V. (1987). "Price and income elasticities of demand: A case study of Louisville". In: Grant, N.K., et al. (Eds.), *Economic Efficiency and the Performing Arts*. Association for Cultural Economics, Akron, pp. 62–73.
- Green, R., Hassan, Z.A., Johnson, S.R. (1992). "The bias due to omitting income when estimating demand functions". *Canadian Journal of Agricultural Economics* 40, 475–484.
- Gruenberg, B. (1975). "How free is free time? Analysis of some determinants of leisure activity pattern". Unpublished manuscript. Middletown, CT.
- Heilbrun, J. (1984). "Once more with feeling: The arts boom revisited". In: Hendon, W.S., et al. (Eds.), *The Economics of the Cultural Industries*. Association for Cultural Economics, Akron, pp. 34–46.
- Heilbrun, J. (1993). "Innovation in art, innovation in technology, and the future of the high arts". *Journal of Cultural Economics* 17, 89–96.
- Heilbrun, J. (1996). "Growth, accessibility and the distribution of arts activity in the United States: 1980 to 1990". *Journal of Cultural Economics* 20, 283–296.
- Heilbrun, J. (1997). "The competition between high culture and popular culture as seen in the New York Times". *Journal of Cultural Economics* 21, 29–40.
- Heilbrun, J., Gray, C.M. (2001). *The Economics of Art and Culture*, second ed. Cambridge University Press, Cambridge.
- Hjorth-Andersen, C. (1992). "Thaliometrics – a case study of Copenhagen theatre market". In: Towse, R., Khakee, A. (Eds.), *Cultural Economics*. Springer, Berlin, pp. 257–266.



- Horowitz, H. (1985). "Panel discussion: Measuring arts participation". In: Owen, V.L., Hendon, W.S. (Eds.), *Managerial Economics for the Arts*. Association for Cultural Economics, Akron, pp. 195–206.
- Houthakker, H.S., Taylor, L.D. (1970). *Consumer Demand in the United States*. Harvard University Press, Cambridge.
- Huntington, P.A. (1991). "Buying behavior over a defined product". *Journal of Cultural Economics* 15, 55–71.
- Jenkins, S., Austen-Smith, D. (1987). "Interdependent decision-making in nonprofit industries: A simultaneous equation analysis of English provincial theatres". *International Journal of Industrial Organization* 5, 149–174.
- Kaali-Nagy, C., Garrison, L.C. (1972). "Profiles of users and nonusers of the Los Angeles music center". *California Management Review* 15, 133–143.
- Kane, D. (2004). "A network approach to the puzzle of women's cultural participation". *Poetics* 31, 403–421.
- Khakee, A., Nilsson, G. (1980). "Is the supply of cultural events an indicator of their demand". In: Hendon, W.S., et al. (Eds.), *Economic Policy for the Arts*. Abt Books, Cambridge, MA, pp. 272–282.
- Kirchberg, V. (1988). "Entrance fees as a subjective barrier to visiting museums". *Journal of Cultural Economics* 22, 1–13.
- Kolb, B.M. (1997). "Pricing as the key to attracting students to the performing arts". *Journal of Cultural Economics* 21, 139–146.
- Kolb, B.M. (2002). "Pricing vs. social factors as the key to attracting young audiences". Presented at the biannual conference of the Association for Cultural Economics International, Rotterdam, June.
- Köster, P.R., Marco-Serrano, F. (2000). Consumers, non-consumers, cultural agents and cultural policies. Working Paper. Prepared for the FOKUS/ACEI Symposium, January.
- Krebs, S., Pommerehne, W. (1995). "Politico-economic interactions of German public performing arts institutions". *Journal of Cultural Economics* 19, 17–32.
- Kuh, E., Meyer, J. (1957). "How extraneous are extraneous estimates?" *Review of Economics and Statistics* 39, 380–393.
- Kurabayashi, Y., Ito, T. (1992). "Socio-economic characteristics of audiences for Western classical music in Japan: A statistical analysis". In: Towse, R., Khakee, A. (Eds.), *Cultural Economics*. Springer, Berlin, pp. 275–287.
- Lange, M., Luksetich, W.A. (1984). "Demand elasticities for symphony orchestras". *Journal of Cultural Economics* 8, 29–48.
- Lange, M., Luksetich, W., Jacobs, P. (1986). "Managerial objectives of symphony orchestras". *Managerial and Decision Economics* 7, 273–278.
- Levental, I. (1989). "Play genre selection: An important factor in developing audiences for live theatre". In: Chartrand, H.H., et al. (Eds.), *Cultural Economics '88: A Canadian Perspective*. Association for Cultural Economics, Akron, pp. 117–125.
- Lévy-Garboua, L., Montmarquette, C. (1996). "A microeconomic study of theater demand". *Journal of Cultural Economics* 20, 25–50.
- Lévy-Garboua, L., Montmarquette, C. (2003). "Demand". In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham, UK, pp. 201–213.
- Lewis, G., Seaman, B.A. (2004). "Sexual orientation and demand for the arts". *Social Science Quarterly* 85, 523–538.
- Lopéz Sintas, J., García Álvarez, E.G. (2004). "Omnivore versus univore consumption and its symbolic properties: Evidence from Spaniards' performing arts attendance". *Poetics* 32, 471–491.
- Luksetich, W.A., Lange, M.D. (1995). "A simultaneous model of nonprofit symphony orchestra behavior". *Journal of Cultural Economics* 19, 49–68.
- Marburger, D.R. (1997). "Optimal ticket pricing for performance goods". *Managerial and Decision Economics* 18, 375–381.
- McCain, R.A. (1979). "Reflections on the cultivation of taste". *Journal of Cultural Economics* 3, 30–52.
- McCain, R.A. (1981). "Cultivation of taste, catastrophe theory, and the demand for works of art". *American Economic Review* 71, 332–334.
- McCain, R.A. (1995). "Cultivation of taste and bounded rationality: Some computer simulations". *Journal of Cultural Economics* 19, 1–15.

- McCain, R.A. (2003). "Taste formation". In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham, pp. 445–450.
- McCaughy, C. (1989). "Arts education and demand for the arts in Canada". In: Chartrand, H.H., et al. (Eds.), *Cultural Economics '88: A Canadian Perspective*. Association for Cultural Economics, Akron, pp. 43–51.
- Michael, R.T. (1972). "The effect of education on efficiency in consumption", N.B.E.R. Occasional Paper 116. Columbia University Press, New York.
- Michael, R.T., Becker, G.S. (1973). "On the new theory of consumer behavior". *Swedish Journal of Economics* 75, 378–396.
- Montgomery, S.S., Robinson, M.D. (2005). "Take me out to the opera: Are sports and arts complements? Evidence from the performing arts research coalition data". Presented at the 8th International Conference on Arts and Cultural Management (AIMAC), Montreal, July 3–6.
- Moore, T.G. (1966). The demand for Broadway theater tickets. *Review of Economics and Statistics* 48, 79–87.
- Moore, T.G. (1968). "The Economics of the American Theater". Duke University Press, Durham.
- Morrison, W.G., West, E.G. (1986). "Child exposure to the performing arts: The implications for adult demand". *Journal of Cultural Economics* 10, 17–24.
- National Endowment for the Arts (2004). "2002 survey of public participation in the arts". Research Division Report #45. National Endowment for the Arts, Washington, DC.
- National Research Center of the Arts, Inc. (1976). *Americans and the Arts: A Survey of the Attitude toward and Participation in the Arts and Culture of the United States Public*. National Research Center of the Arts, Inc., New York.
- O'Hagan, J.W. (1996). "Access to and participation in the arts: The case of those with low income/educational attainment". *Journal of Cultural Economics* 20, 269–282.
- Orend, R.J., Keegan, C. (1996). *Education and Arts Participation: A Study of Arts Socialization and Current Arts-related Activities using 1982 and 1992 SPPA Data*. National Endowment for the Arts, Washington, DC.
- Oteri, M., Trimarchi, M. (1990). "Public subsidies and cultural habits: An empirical test of drama attendance". *Rivista di Diritto Finanziario e Scienza delle Finanze* 49, 524–537.
- Owen, J.D. (1969). *The Price of Leisure*. Rotterdam University Press, Rotterdam.
- Paton, D., Cooke, A. (2005). "Attendance at county cricket: An economic analysis". *Journal of Sports Economics* 6, 24–45.
- Peterson, R.A. (1992). "Understanding audience segmentation: From elite and mass to omnivore and univore". *Poetics* 21, 243–258.
- Peterson, R.A., Hull, P.C., Kern, R.M. (2000). "Age and arts participation: 1982–1997". Research Division Report # 42. National Endowment for the Arts, Washington, DC.
- Pollak, R.A. (1970). "Habit formation and dynamic demand functions". *Journal of Political Economy* 78, 745–763.
- Pommerehne, W., Kirchgassner, G. (1987). "The impact of television on the demand for cinema and theatre performances". In: Grant, N.K., et al. (Eds.), *Economic Efficiency and the Performing Arts*. Association for Cultural Economics, Akron, pp. 44–61.
- Prieto-Rodríguez, J., Fernández-Blanco, V. (2000). "Are popular and classical music listeners the same people?". *Journal of Cultural Economics* 24, 147–164.
- Putler, D.S., Lele, S. (2003). "An easily implemented framework for forecasting ticket sales to performing arts events". *Marketing Letters* 14, 307–320.
- Relish, M. (1997). "It's not all education: Network measures as sources of cultural competency". *Poetics* 25, 121–139.
- Ryans, A.B., Weinberg, C.B. (1978). "Consumer dynamics in nonprofit organizations". *Journal of Consumer Research* 5, 89–95.
- Scheff, J. (1999). "Factors influencing subscription and single-ticket purchases of performing arts organizations". *International Journal of Arts Management* 1, 113–125.
- Schimmelpfennig, J. (1997). "Demand for ballet: A non-parametric analysis of the 1995 Royal Ballet summer season". *Journal of Cultural Economics* 21, 119–127.

- Seaman, B.A. (2004). "Competition and the non-profit arts: The lost industrial organization agenda". *Journal of Cultural Economics* 28, 167–193.
- Seaman, B.A. (2005). "Empirical studies of demand for the performing arts". Nonprofit Studies Program Working Paper NP05-03. Andrew Young School of Policy Studies, Georgia State University.
- Smith, T.M. (1998). "Two essays on the economics of the arts: The demand for culture and the occupational mobility of artists". Unpublished Ph.D. Thesis. University of Illinois at Chicago.
- Spinnewyn, F. (1981). "Rational habit formation". *European Economic Review* 15, 91–109.
- Stavins, J. (1997). "Estimating demand elasticities in a differentiated product industry: The personal computer market". *Journal of Economics and Business* 49, 347–367.
- Stigler, G.J., Becker, G.S. (1977). "De gustibus non est disputandum". *American Economic Review* 67, 76–90.
- Throsby, D. (1990). "Perception of quality in demand for the theatre". *Journal of Cultural Economics* 14, 65–82.
- Throsby, D. (1994). "The production and consumption of the arts: A view of cultural economics". *Journal of Economic Literature* 32, 1–29.
- Throsby, C.D., Withers, G.A. (1979). *The Economics of the Performing Arts*. St. Martins, New York.
- Tobias, S. (2004). "Quality in the performing arts: Aggregating and rationalizing public opinion". *Journal of Cultural Economics* 28, 109–124.
- Touchstone, S.K. (1980). "The effects of contributions on price and attendance in the lively arts". *Journal of Cultural Economics* 4, 33–46.
- Trienekens, S. (2002). "'Colourful' distinction: The role of ethnicity and ethnic orientation in cultural consumption". *Poetics* 30, 281–298.
- Ulibarri, C.A. (2005). "Bayesian learning from arts goods? – A comment". *Journal of Cultural Economics* 29, 137–141.
- Upright, C.B. (2004). "Social capital and cultural participation: Spousal influences on attendance at arts events". *Poetics* 32, 129–143.
- Urrutiaguer, D. (2002). "Quality judgements and demand for French public theatre". *Journal of Cultural Economics* 26, 185–202.
- Van Eijck, K. (1997). "The impact of family background and educational attainment on cultural consumption: A sibling analysis". *Poetics* 25, 195–224.
- Van Eijck, K. (2000). "Richard A. Peterson and the culture of consumption". *Poetics* 28, 207–224.
- Verhoeff, R. (1992). "Explaining differences in the geographical reach of performances". *Journal of Cultural Economics* 16, 73–82.
- Villani, A. (1992). "Rational addiction in the arts". *Recherche Economique XLVI*, 41–54.
- Waterman, D., Schechter, R., Contractor, N.S. (1991). "Overcoming barriers to the live arts: Can the media compensate?" *Journal of Cultural Economics* 15, 19–40.
- Weinberg, C.B. (1986). "ARTS PLAN: Implementation, evolution, and usage". *Marketing Science* 5, 143–159.
- Weinberg, C.B., Shachmut, K.M. (1978). "ARTS PLAN: A model based system for use in planning a performing arts series". *Management Science* 24, 654–664.
- West, E.G. (1985). *Subsidizing the Performing Arts*. Ontario Economic Council.
- West, E.G., McKee, M. (1983). "De gustibus non est disputandum: The phenomenon of 'merit wants' revisited". *American Economic Review* 73, 1110–1121.
- Wiesand, A. (1995). "Musiktheater und Konzerte: mehr Rückhalt in der Bevölkerung. Aus den Ergebnissen des 5. Kulturbarometers". *Das Orchester* 6/95, 2–14.
- Withers, G. (1980). "Unbalanced growth and the demand for the performing arts: An econometric analysis". *Southern Economic Journal* 46, 735–742.
- Zuzanek, J., Lee, M. (1985). "Social ecology of arts audiences". *Journal of Cultural Economics* 9, 65–84.

## NONPROFIT FIRMS IN THE PERFORMING ARTS\*

ARTHUR C. BROOKS

*Maxwell School of Citizenship and Public Affairs, Syracuse University, USA*

### Contents

Abstract	474
Keywords	474
1. Introduction	475
2. Nonprofit performing arts firms: Basic facts	477
3. Why are there performing arts nonprofits?	480
3.1. The demand for nonprofit activity	480
3.2. The objectives of nonprofit firms	482
4. Performing arts production and cost	484
4.1. Inputs to production	484
4.2. Production functions	485
4.3. Output decisions	487
5. The cost disease	489
6. Revenues of performing arts firms	493
6.1. Earned revenues	494
6.2. Donated revenues	495
6.3. Fundraising innovations	497
6.4. Government funding	498
7. Summary and future research directions	500
References	502

\* For their helpful suggestions on this chapter, I am grateful to Victor Ginsburgh, David Throsby, Sir Alan Peacock, and Bruce Seaman.

**Abstract**

The nonprofit performing arts have received substantial attention in the cultural economics literature, and represent an interesting application for many areas of economic inquiry. This chapter surveys the relevant theory and the most prominent empirical studies on performing arts nonprofits. The chapter begins with a description of the nonprofit sector – and the role of the performing arts in this sector – around the world. I then ask why performing arts nonprofits exist, taking into account the objectives of both consumers and suppliers of performing arts services. Next, I study the production and cost conditions that these firms face, paying particular attention to issues such as product quality, product cross-subsidization, and the so-called “cost disease”. The issue of revenue sources and their generation follows, with a special emphasis on earned revenues, donations, and government subsidies. This discussion includes topics such as ticket pricing strategies, fundraising innovations, and the relationship between private giving and public funding. The chapter closes with suggestions for future research on the nonprofit performing arts.

**Keywords**

performing arts firms, nonprofit organizations, cultural economics

*JEL classification:* Z11, L31

## 1. Introduction

The nonprofit sector encompasses an important portion of the world economy. In 1995, for example, it made up about 4.6 percent of the GDP of 22 industrialized and developing nations, 5 percent of total nonagricultural employment, and 10 percent of all service employment [Salamon et al. (1999)]. In the United States, the nonprofit sector had revenues in 1997 of \$665 billion, or about 6.7 percent of GDP, and represented 7.1 percent of all paid employment [Independent Sector (2001)].<sup>1</sup>

There is no uniform international legal definition of “nonprofit organization” (NPO) or “nongovernmental organization” (NGO). Organizations are subjected to different laws in different countries. In the United States, for example, there are currently 29 different types of legally-defined nonprofits in the tax code, from charities, to credit unions, to farmers’ collectives. The largest category in the tax code includes what most other countries also define as nonprofits: organizations that exist for charitable, religious, or educational purposes. The American tax code opaquely defines these organizations as

Corporations, and any community chest, fund, or foundation, organized and operated exclusively for religious, charitable, scientific, testing for public safety, literary, or educational purposes, or to foster national or international amateur sports competition (but only if no part of its activities involve the provision of athletic facilities or equipment), or for the prevention of cruelty to children or animals, no part of the net earnings of which inures to the benefit of any private shareholder or individual, no substantial part of the activities of which is carrying on propaganda, or otherwise attempting, to influence legislation . . . and which does not participate in, or intervene in (including the publishing or distributing of statements), any political campaign on behalf of (or in opposition to) any candidate for public office.<sup>2</sup>

Most nonprofit activity is dedicated to areas in which we typically find goods and services with public-goods characteristics: religion, education, health, social services, international development, the environment, and arts and culture. Nonprofit designation generally entitles organizations to public or private donations. In many countries, the revenues raised by these organizations are free from corporate taxation; in some countries, private donations are tax-deductible for the donors as well.

Functional definitions of nonprofits are more illuminating than legal definitions. Hall (1987) defines a nonprofit organization as a private group that associates in order to

- (i) undertake public tasks on behalf of the government,
- (ii) provide public goods and services for which there is a demand but no supply from either the public or for-profit sectors, or
- (iii) influence public policy.

<sup>1</sup> Arguably, these figures overestimate the true “contribution” of nonprofits to the economy, because they don’t measure the value added of these firms to GDP.

<sup>2</sup> Source: US Internal Revenue Code Title 26, Subtitle A, Chapter 1, Subchapter F, Part I, Section 501 (c) 3.

Arguably less descriptive, but more precise, is the functional definition of the nonprofit firm usually used by economists: A firm  $i$  may be defined as a nonprofit if, as an organization, it seeks to maximize some utility  $u_i(\cdot)$ , subject to the “nondistribution constraint”  $ER_i + UR_i - TC_i = 0$ , where  $ER_i$  = earned revenues,  $UR_i$  = “unearned revenues”, i.e. those donated by government and the private sector, and  $TC_i$  = the organization’s total costs. The nondistribution constraint can also be modified to the weaker assumption that economic profits sum to zero over a multi-period time horizon, to allow for such phenomena as borrowing and endowments.

The nonprofit sector in almost every country, however defined, contains firms dedicated to the arts in general, and the performing arts in particular. In the US, the National Center for Charitable Statistics (NCCS) at the Urban Institute defines 64 different categories of nonprofit arts activity, 10 of which are undertaken by performing arts organizations.<sup>3</sup> These include performing arts centers, dance ensembles, ballet ensembles, theater companies, music organizations, symphony orchestras, opera companies, choruses, and other music groups.<sup>4</sup>

Although a multi-billion dollar industry, the nonprofit performing arts still comprise only a small part of the nonprofit economy. Figure 1 illustrates this for the United States, showing that the arts and culture are about 2 percent of all nonprofit activity, and the performing arts are about a quarter of that amount. Further breakdown shows that most nonprofit performing arts activity involves performance promotion and facilities (33 percent), theater companies (24 percent), and classical music organizations (21 percent), followed by opera and dance companies (12 and 7 percent, respectively).<sup>5</sup>

The arts are a larger portion of the nonprofit sector in other parts of the world than they are in the US. For example, they comprise 7 percent of the sector in Germany, 9 percent in Italy, 21 percent in the United Kingdom, and 57 percent in Hungary [Rose-Ackerman (1996)]. In general, the arts are especially heavily represented in the nonprofit sectors of Eastern Europe: on average, 35 percent [Salamon et al. (1999)]. Part of the difference between the United States and much of the rest of the world has to do with the fact that many of the health, education, and social services that make up the largest component of the American nonprofit sector reside in the public sector in many social democracies and transition economies.

Whether a small or large part of a country’s nonprofit sector, the performing arts have played an interesting role in the development of the literature on nonprofit economics.

<sup>3</sup> See [www.urban.org](http://www.urban.org), for details on the National Taxonomy of Exempt Entities (NTEE) classification system.

<sup>4</sup> The forms of nonprofit performing arts firms are highly variegated. As DiMaggio (1987) suggests, their common legal status indicates homogeneity that does not exist. Jeffri (1980, pp. 48–49) makes this point in surveying the landscape of “dance companies”, which can range from single performers to large nonprofits managed by for-profit companies.

<sup>5</sup> The categories in Figure 1 do not correspond precisely with the NCCS taxonomy because these data were derived from the *US Census of Service Industries* (USCSI) [US Department of Commerce, Bureau of the Census (1997)], which categorizes nonprofits slightly differently.

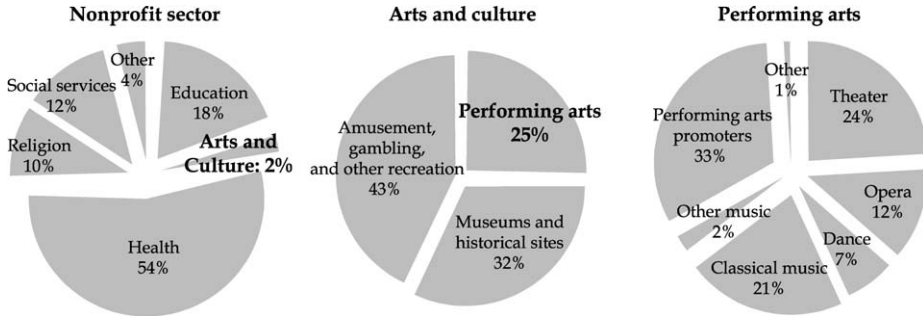


Figure 1. The American nonprofit sector. Source: Independent Sector (2001), US Department of Commerce, Bureau of the Census (1997).

In this chapter, I follow the development of nonprofit economic theory over the past 35 years, discussing some of the most salient findings and relating them to the literature on the economics of nonprofit performing arts organizations. In so doing, I aim both to show how research on the performing arts has enriched the nonprofit field in general, and also to expose the areas of work on the performing arts that are relatively underdeveloped.

The rest of this chapter is organized in six parts. I begin with a brief survey of the facts surrounding the nonprofit performing arts sector, and then move on to ask why nonprofit performing arts firms exist (both from the perspective of consumers as well as the firms themselves). Following this, I look at production and cost, and then focus on a topic of particular interest in the performing arts, the so-called “cost disease”. I then look at the financing of performing arts nonprofits at the earned, government, and philanthropic levels. Finally, I discuss the trends and neglected areas in this literature, and suggest future work.

## 2. Nonprofit performing arts firms: Basic facts

The best data available suggest that the nonprofit arts represent about one-half of one percent of the workforces in most countries [Salamon et al. (1999)]. Figure 2 summarizes these percentages in 22 countries around the world. It is reasonable to assume that not more than half of this amount is dedicated to the performing arts. To attain much higher detail about the dimensions of the nonprofit performing arts requires looking more closely at a specific country. The United States is particularly convenient in this regard because statistics on the sector are relatively abundant.

We start by asking how many performing arts firms there are, their average size, how they break down across discipline, and whether they are organized for-profit or nonprofit. Table 1 shows that in 1997, there were more than 8000 organizations dedicated to the performance of theater, opera, dance, and music in the US. This does not include



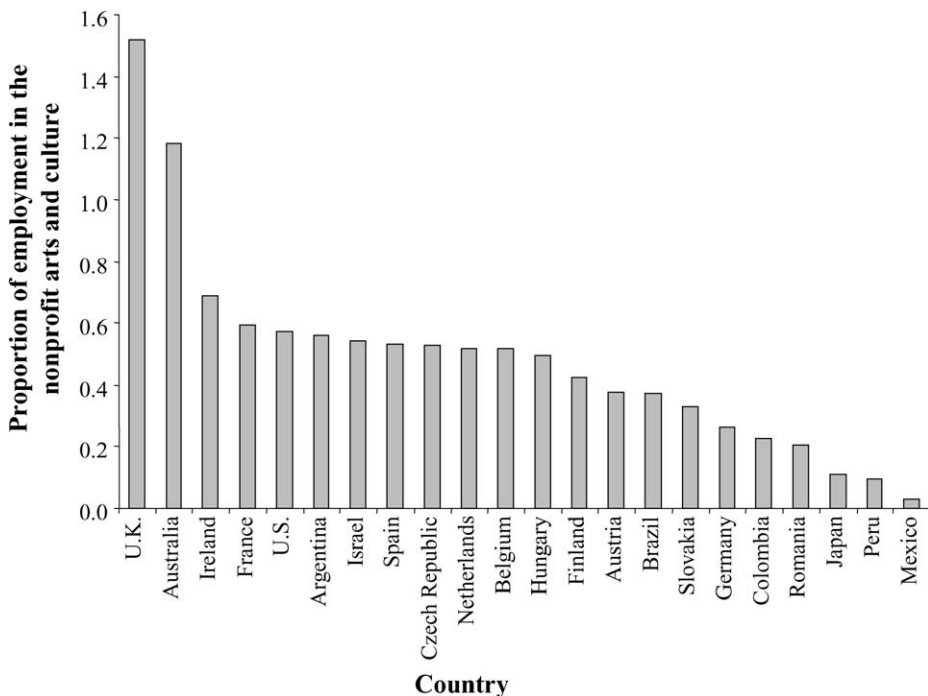


Figure 2. Proportion of employment in the nonprofit arts, 22 countries. Source: Salamon et al. (1999).

performing arts agents, promoters, or facilities; it also excludes many companies such as circuses that are commonly listed alongside the performing arts in the *US Census of Service Industries (USCSI)*.

Theater in America is fairly evenly split between the commercial and nonprofit sectors, while opera, symphony orchestras, and chamber music are dominated by nonprofits. The average annual revenues of nonprofit opera companies are much larger than the other disciplines, at over \$4 million. The average nonprofit theater company, on the other hand, had annual revenues under \$800,000. Note that these figures almost certainly understate the number of organizations and overstate the average size, because US Census data only consider nonprofits over \$25,000 (which are required to file tax forms with the US Internal Revenue Service). Smith (1997) estimates that small, non-filing organizations outnumber filers by as many as eight to one.

The average revenues of American performing arts have been falling. The USCSI indicates, for example, that the average nonprofit performing arts firm had seven percent lower revenues (in real terms) in 1997 than in 1982. Rather than indicating a moribund performing arts market, however, this is a result of the explosion in the number of these firms, which increased 81 percent over the same period, an average annual growth rate of about 4 percent. In general, earned revenues make up the largest portion of total

Table 1  
Dimensions of the performing arts in the United States

	Number of organizations	Percentage nonprofit	Average annual revenues per nonprofit firm
Theater	3077	49%	\$788,237
Opera	170	86%	\$4,038,150
Dance	530	70%	\$871,445
Classical music	805	87%	\$1,487,401
Other music	3775	14%	\$225,743

Source: US Department of Commerce, Bureau of the Census (1997).

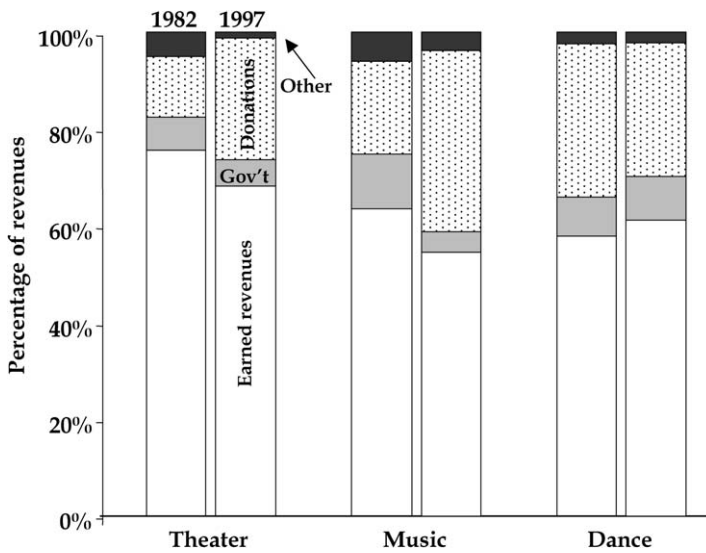


Figure 3. Funding sources to US nonprofit performing arts firms, 1982–1997. Source: US Department of Commerce, Bureau of the Census (1997).

revenues for performing arts firms in the US, followed by donations, and then government contributions. In 1997, 59 percent of the sector’s income was earned, 29 percent was donated, 5 percent came from government, and 7 percent came from other sources [US Department of Commerce, Bureau of the Census (1997)]. We might suppose that the US is anomalous in this regard internationally, assuming that private giving and market activity are disproportionately high in cultural affairs, relative to government support. I will turn to this subject in greater detail later in this chapter.

Figure 3 illustrates that the proportions of funding changed in the US over the 15-year period ending in 1997. For firms in theater and music, earned revenues fell as

a percentage of the total, while contributions rose to fill the gap. Dance is somewhat different, in that the increase in the budget of dance companies from 1982 to 1997 came disproportionately in the form of earned revenues.

In most industries, high fixed costs lead to high firm concentration; the performing arts are no exception. In the disciplines in [Table 1](#), the firms with the highest fixed costs tend to be opera and classical music (mainly symphony orchestras), which require facilities for hundreds of employees and thousands of audience members. Measuring size in terms of revenues, the largest 20 nonprofit firms in each of these disciplines control 80 and 50 percent of their industries, respectively. Theater and dance companies, on the other hand, have more structural flexibility regarding the scale of operations, meaning that fixed costs are not necessarily so high. The market shares controlled by the top 20 nonprofit firms in these disciplines are respectively 25 and 20 percent [[McCarthy et al. \(2001\)](#)].

### 3. Why are there performing arts nonprofits?

A discussion of the economics of the nonprofit performing arts does well to include the most basic question: Why do these firms exist? This question really has two sides to it. First, we would like to know why these firms have a place in free market economies – what is the source of demand for a specifically *nonprofit* performing arts product? Second, what is the motivation of performing arts firms that are organized not-for-profit – does the nonprofit form facilitate attainment of these firms’ objectives? The literature on the theory of the nonprofit firm provides guidance on these issues, and authors on the nonprofit performing arts have enriched the discussion.

#### 3.1. The demand for nonprofit activity

The most frequent explanation for the demand for nonprofit provision of goods and services comes from the “public goods model”, developed by [Weisbrod \(1978\)](#), [Warr \(1982\)](#), [Bergstrom, Blume and Varian \(1986\)](#), [Andreoni \(1988\)](#), and others.<sup>6</sup> This model starts with the assumption that an agent  $i$  exhausts his budget  $m_i$  between purchases of a private good  $x_i$  and gifts to a public good  $d_i$ . Setting the price of a unit of each to unity,

$$x_i + d_i = m_i. \tag{1}$$

Assuming that the public good is truly nonrivalrous and nonexcludable in consumption, the total amount of resources to the good donated across the population,  $\sum_i d_i = D$ , can be enjoyed by each member of the population. Hence, the agent’s utility can be

<sup>6</sup> The particular treatment here loosely follows [Andreoni \(1998\)](#).

represented by the equation

$$u_i = u_i(x_i, D). \quad (2)$$

We assume that  $u_i$  has the standard properties. Notice that  $D$  can be redefined as  $d_i + D_{-i}$ , where  $D_{-i}$  is the contribution of the rest of the population. At the Nash equilibrium, this contribution is assumed constant by each agent, and thus can be added to each side of the budget constraint. We define the agent's problem as  $\max_{x_i, D} \{u_i(x_i, D)\}$ , subject to  $x_i + D = m_i + D_{-i}$ ,  $D \geq D_{-i}$ .

The first-order conditions for this problem will produce a private demand function for  $D$ , which can be written as<sup>7</sup>

$$D = f_i(m_i + D_{-i}). \quad (3)$$

Assuming that  $f_i(0) = 0$  and that both  $x$  and  $d$  are normal goods, we know that  $f'_i \in (0, 1)$ , and thus that if at least one agent  $i$  has  $m_i > 0$ ,  $D$  must be positive. This establishes the possible existence of private donations for a public good.<sup>8</sup> This demand can manifest itself in private donations or public subsidies for the public good.<sup>9</sup>

The nonprofit organizational form equips a firm to garner donated funds. This is often called the “subsidy theory” of why nonprofits exist [Hansmann (1987)]. West (1987, p. 39) summarizes this theory when he says that “a policy of aiding the nonprofit [performing arts] organizations . . . can easily be predicted to encourage their growth”. Nonprofits are positioned to collect donated funds for legal reasons, as well as because nonprofit status assuages fears that donated funds will simply be converted into profits [Bilodeau and Slivinski (1998)]. This constitutes the (complementary) “contract failure” explanation for nonprofits [Hansmann (1980)]: Donors cannot observe how funds are used, and insist on an institutional arrangement that prohibits the distribution of revenues for activities not involving operations. Relatively few studies on the nonprofit performing arts employ these explanations. One exception is O’Hagan and Purdy (1993), who implicitly use both subsidy theory and contract failure in study of the history and management strategies of an Irish opera festival.

Another explanation for the nonprofit form is that of “club goods” [Cornes and Sandler (1984)]. Club goods are generally exclusionary goods and services from which people derive mutual benefits from shared characteristics, or the actual exclusion of nonmembers. Nonprofit status allows member control over these types of organizations. The club goods model has received some application in the performing arts literature.

<sup>7</sup> Authors [e.g., Bergstrom, Blume and Varian (1986)] have shown that a unique Nash equilibrium of donations exists in the vector  $d = (d_1^*, \dots, d_n^*)$ .

<sup>8</sup> This model assumes that agents recognize the public value they are paying for; a common argument regarding the public good in the arts is that ordinary citizens don't understand this value and hence the choice to fund it, privately or publicly, must come from those who are better-informed. These are so-called “merit goods” [Musgrave (1959)].

<sup>9</sup> In the latter case, we would predict that public support would occur if those with above-average demand for the public good have the political influence necessary to generate the subsidy.

DiMaggio (1982) uses the basic concept to explain how performing arts organizations in the United States were protected by social elites from the aesthetic contamination of commercial markets. Kuan (2001) treats the subject theoretically, creating a model of the performing arts world in which people are either intensive or casual in their consumption, and then shows straightforwardly that the former group has an incentive to “produce” the performing arts service itself (that is, set up production as a mutually-beneficial club), as opposed to purchasing it from a for-profit. Kushner and King (1994) empirically confirm the proposition that an American choral society possesses club good characteristics by looking at the proportion of admission tickets preferentially purchased by society “members”.

Models that begin by assuming the performing arts are pure public goods may be problematic, in that the assumption of complete nonexcludability is unrealistic with respect to the private benefit that audience members derive from performing arts events. However, some authors have argued that  $D$  in Equation (2) contains a number of nonexcludable, nonuse public arts benefits as well, although very few writers have tested empirically for the existence of such benefits. For example, Frey (1997) lists the following nonuse, public values from the arts in general, which could also be argued to exist for the performing arts in particular.

- *Existence value.* Even if they don’t directly consume a particular cultural good, some people appreciate its existence.
- *Option value.* Nonusers may place a positive value on the option to become users of it in the future, and hence favor its preservation.
- *Education value.* Cultural goods might create intellectual and cultural spillovers among users and nonusers.
- *Prestige value.* Cultural goods might produce prestige for their region of origin.
- *Bequest value.* Users and nonusers may derive utility from the expected enjoyment of a cultural good by future generations.

Some authors would add the economic impact of performing arts activities to this list [Brooks and Kushner (2001)]. Others, particularly those that look at the performing arts among indigenous peoples, might add *cultural value* to this list, where “cultural” has an anthropological connotation.<sup>10</sup> For example, Mackerras (1999) studies how practice of the Tibetan performing arts helps to maintain the integrity of traditional Tibetan culture in the face of modernization.

### 3.2. *The objectives of nonprofit firms*

Given a demand for nonprofit performing arts services, what do we know about the objectives of the firms themselves? In other words, what do we know about the nature

<sup>10</sup> This anthropological sense of “culture” is in general contrast to the way the term is almost always used in the field of cultural economics. Cultural economics usually employs Matthew Arnold’s (1869, p. viii) definition: “a pursuit of our total perfection by means of getting to know, on all the matters which most concern us, the best which has been thought and said in the world”.

of  $u_i(\cdot)$  for a firm  $i$ ? The nonprofit literature has produced a limited set of theories on the topic.<sup>11</sup>

One of the most influential works on the economic objective of nonprofit firms in general came from Steinberg (1986), who suggested that, given zero economic profits, nonprofit “utility” would tend to fall along some continuum between spending on core operations and budget size. The key in his formulation was fundraising expenditures; as such, he called organizations that maximized revenues to be spent on operations (net of fundraising expenditures) *service maximizers*, while those that ignored them *budget maximizers*. Utility was defined as a linear combination of these two possibilities; that is, for an organization  $i$ ,

$$u_i = \alpha S_i + (1 - \alpha)[ER_i + D(F_i)],$$

subject to  $ER_i + D(F_i) - S_i - F_i = 0,$  (4)

where  $0 \leq \alpha \leq 1$ ,  $S_i$  = service, or expenditures on core operations;  $ER_i$  = earned revenues;  $F_i$  = fundraising expenditures; and  $D(F_i)$  = the donations that come from fundraising. Clearly, if  $\alpha = 0$ , a firm is a budget maximizer; if  $\alpha = 1$ , it is a service maximizer; if  $0 < \alpha < 1$ , the objective is mixed. We assume that  $D' \geq 0$ ,  $D'' < 0$ , and  $0 \leq F \leq ER + D$ .

Solving the firm’s problem, we see that when  $\partial u_i / \partial F_i = 0$ ,  $\partial D_i / \partial F_i = \alpha$ . That is, the value of  $\alpha$  is equivalent to the marginal impact of fundraising on donations (in the language of nonprofit managers, the “fundraising yield”). Obviously, the value of  $\alpha$  is an empirical question. Steinberg found that health nonprofits tended to be budget maximizers, while those in social welfare, education, and (most notably) the arts were service maximizers.

In the case of the performing arts, budget is easy to understand, while service is less obvious: As noted by Throsby (1994) and others, it could refer to the quantity of service, quality, or both. Hansmann (1981) distinguishes these as different maximands in his treatment of performing arts firms. He assumes that performing arts firms maximize utility with respect to audience size, quality, or budget. He then contrasts socially-optimal firm behavior (in which consumer surplus is maximized) with that predicted for audience maximizers, quality maximizers, and budget maximizers. He shows that if the firm is an audience maximizer, quality will tend to be suboptimally low; if the firm is a quality maximizer, quality will generally be suboptimally high relative to audience size. If the firm is a budget maximizer, the social optimum may be obtained, but only by coincidence.

In one empirical paper designed specifically to test Hansmann’s model, Luksetich and Lange (1995) study the objectives of American orchestras of different sizes in the 1970s

<sup>11</sup> Some sociologists have challenged the notion underlying the literature on performing arts objective functions; namely, that objectives are stationary and well-defined. DiMaggio (1987) points out that the concept of pursuing a focused strategy may be unrealistic for many arts firms. In addition, strategies may change according to the firm’s external circumstances [Salem (1984)]. Objectives may even change incoherently due to contradictory directives from funders [DiMaggio (1984)].

and 80s. They find that in general, large orchestras (in terms of budget) are primarily quality maximizers and secondarily budget maximizers. In contrast, medium-sized and small orchestras tend to be audience maximizers. Several other empirical papers, while not specifically designed to test models such as Hansmann's or Steinberg's, nevertheless provide evidence that nonprofit performing arts firms behave in one or more of the ways these models predict. In his analysis of England's Royal Shakespeare Company, for example, [Gapinski \(1985\)](#) begins with the question of whether performing arts nonprofits optimize with respect to anything at all. He concludes that theater firms do indeed optimize, maximizing output subject to the nondistribution constraint. This is consistent with Steinberg's service maximization.

Other papers have focused more specifically on quality as the maximand, generally finding that it is in decline due to adverse financial constraints faced by the organizations. For example, [Heilbrun \(2001\)](#) sees falling programmatic diversity among US opera companies as a way to lower rehearsal costs. [Throsby \(1990\)](#) finds that theater audiences in Australia don't value quality in terms of new repertoire, leading firms to sacrifice new works for higher-revenue repertoire "chestnuts". And [Peacock \(1976\)](#) finds that symphony orchestras in London have increasingly sacrificed rehearsal time for performances. While a more rigorous discussion of quality (as well as quantity) is deferred to a later section of this chapter, we might note that the literature just cited suggests a difference between the way patrons and audiences define quality, and the way *researchers* do.

#### 4. Performing arts production and cost

The foregoing discussion of performing arts nonprofits' objectives omitted the answers to several important questions: What are the inputs to production of the performing arts? How are they most appropriately defined and measured? How are they combined to produce the service? How do nonprofits compare (and compete) with each other and with for-profit performing arts organizations?

##### 4.1. *Inputs to production*

Most treatments of nonprofit production inputs focus on labor, because this is the area in which most authors feel that nonprofits differ most dramatically from the for-profit and public sectors. Specifically, just as the organizations themselves face a multidimensional utility function, so do nonprofit workers, who presumably value an organization's mission – something they believe in. The result might be the presence of volunteer labor in the sector, or persistently sub-market wages [[Handy and Katz \(1998\)](#)]. Indeed, volunteers make up between 25–40 percent of the nonprofit labor forces in the US, Western Europe, Latin America, and Central Europe [[Salamon et al. \(1999\)](#)]. And nonprofit employees in the US are paid 35–40 percent lower wages than their professional equivalents in the for-profit sector [[Weisbrod \(1983\)](#)];

Frank (1996)]. Note that volunteers are disproportionately highly represented in the arts: While arts firms are only 2 percent of the US third sector, they use 5.2 percent of American volunteers [Independent Sector (2001)].

Relative to nonprofit labor, the subjects of capital, land, and entrepreneurship have received little treatment in the nonprofit economic literature. Authors have noted that the exemption of nonprofits from corporate taxation in most countries makes the use of plant, equipment, and land relatively cheaper, which should encourage higher utilization than in the for-profit world [Brown and Slivinski (2002)], but research has not gone past this observation. The only rigorous work on nonprofit entrepreneurship is Bilodeau and Slivinski (1998), which models the decision of how and why someone would set up a nonprofit firm. They show circumstances under which an entrepreneur voluntarily adopts a budget constraint in order to collect the private gifts on offer in the public-goods model introduced earlier.

In the performing arts economics literature, no specific treatments of capital or entrepreneurship have appeared, to my knowledge. In contrast, quite a bit has been written about labor, which receives several chapters in this volume as well as a discussion later in the present chapter under the heading of the “cost disease”. Suffice it to say at this point that labor conditions for firms have been found to vary significantly from country to country in the performing arts. Specifically, it seems that rising labor cost problems plague American firms more than others. Schwarz (1992) finds that among orchestras, labor costs rose and fell in Canada more than in the US, following the financial circumstances of the firms. Rubinstein, Baumol and Baumol (1992) finds that labor costs were astronomical in the US compared to the former Soviet Union (60 percent versus 20 percent of total performing arts production costs). Benhamou (2000) finds substantial growth in performing arts employment in France and Great Britain in the 1980s.

#### 4.2. Production functions

In combining resources, there is no especially intuitive reason to believe that a nonprofit performing arts production function would vary dramatically from some traditional form in production theory. So we might begin with

$$Q = f(L, Z), \tag{5}$$

where  $L$  = labor and  $Z$  is a vector of capital, land, and entrepreneurship. We assume that  $f$  is defined only over nonnegative input values, and is a regular, strictly quasi-concave function. Without a statement of the firm’s problem, Equation (5) makes no assumptions about any organizational objective.

Equation (5) raises several questions. First of all, what exactly is  $Q$ ? A number of authors have defined it in different ways; Throsby and Withers (1979, pp. 11–12) summarize the definitions nicely along four dimensions over a given time period:

- number of works produced by a firm;
- number of performances;



- number of potential performance attendees (generally, the number of seats in a performance venue); or
- number of actual attendees.

The first two measures can give ideas of economies of scale and scope – whether, say, a theater company should perform more works less frequently, or rather have longer runs of fewer works. The third and fourth measures distinguish between quantity supplied and quantity demanded, and hence allow study of disequilibrium prices and policies.

However defined, Equation (5) will probably have to be refined to reflect the multiple output definitions. However, this still leaves out the important issue of *quality*, which enters into practically every management discussion of the performing arts. Shanahan (1978, p. 13) points out the importance of quality when he states what is probably obvious: “To properly study the markets for artistic goods, it is necessary to deal with the aesthetic nature of art”. Abbé-Decarroux (1994) underscores the importance of quality by noting that demand for the performing arts is affected by quality expectations; as such, it functions much like risk in the consumption decision.

Throsby (1990) takes on the measurement of quality, adopting (as was the case with quantity) a multidimensional approach. His taxonomy breaks performing arts quality into three parts:

- source material characteristics (such as the quality of a script or score);
- technical factors (such as performance execution); and
- benefits to audiences, society and to the art form itself.

This approach allows research to put a finer point on empirical estimations of, for example, Hansmann’s (1981) treatment of quality presented earlier.<sup>12</sup>

Given these considerations, we might redefine Equation (5) as

$$Q = f(L, m, v, Z), \quad (6)$$

where  $Q$  = performance attendees,  $m$  = the number of performances, and  $v$  = quality.

Given these inputs, modeling production is not simplified by assuming the nondistribution constraint; it is potentially complicated by the fact that nondistribution, as well as multidimensional firm objectives, may lead to output incentives that are not compatible with traditional firm theory. In other words, nonprofit firms may operate in “un-economic” regions of production. The performing arts pose the further complication that they are characterized by inputs that diverge wildly in their importance to production. For example, while some orchestras can operate at least in the short run without formal administrators, none can without musicians.

One way to examine these characteristics is by employing the transcendental production function [Halter, Carter and Hocking (1957)]. Suppose we have a performing arts firm with two inputs; for example, imagine a theater company  $i$  at time  $t$  that employs actors  $L_{it}$  and equipment  $K_{it}$ . The output  $Q_{it}$  of the organization – productions,

<sup>12</sup> To date, little work on the performing arts has tried to put such a multi-dimensional approach into practice. Rather, most papers have tended to construct one proxy for quality (such as programming diversity), and then treat the issue tangentially to quantity.

or perhaps services – is characterized by the function

$$Q = \Omega e^{\gamma t} L^\alpha e^{\beta L} e^{\theta K + \delta K^2} e^\varepsilon, \quad (7)$$

where the firm and time subscripts are dropped for ease of exposition. In Equation (7),  $\Omega$  represents a composite environment parameter to capture the firm's unique exogenous circumstances, including quality;  $\gamma$  represents the effect of time (the marginal impact of time on the log of  $Q$ ) and measures technological change; the production parameters are  $\alpha$ ,  $\beta$ ,  $\theta$ , and  $\delta$ ; and  $\varepsilon$  is a random error.

Note that  $f(0, K) = 0$ , while  $f(L, 0) > 0$  (for positive values of  $L$ ), reflecting the relative importance of the inputs. The time and production parameters may be either positive or negative, reflecting the fact that nonprofit performing arts firms may have a marginal products that are decreasing, increasing, or even negative. The output elasticities of  $L$  and  $K$  also depend on the values of the production parameters. [Gapinski \(1980\)](#) uses American data on theater, opera, orchestra, and ballet companies from the 1960s and 70s to estimate these elasticities, and found all principal inputs (artists and capital) had positive elasticities less than unity; that secondary inputs (administrators) for ballet had a negative elasticity; and that the elasticity was higher for artists than for any other input. In a later paper [[Gapinski \(1984\)](#)], he uses data on English theater to estimate labor and capital output elasticities, finding similar values (including the finding that artist elasticity is twice that of capital). [Goudriaan and Pommer \(1987\)](#) perform a similar estimation (although using a simpler Cobb–Douglas production function) on 1984 data on orchestras and theater companies in the Netherlands. While measured elasticities were positive and less than unity, it is notable that in the case of theater, the elasticity of capital (0.65) exceeded that of labor (0.50).

### 4.3. Output decisions

Obviously, performing arts firms have considerable latitude in deciding what to produce and how much. Should an orchestra perform more “pops”, contemporary music, or standard repertoire? More matinees, or night concerts? Should it go on tour? Should it put on major concerts on the same nights as performances by other entertainment organizations? Understanding of nonprofit firms' output decisions – including those of performing arts firms – was greatly enhanced by [James' \(1983\)](#) influential model of nonprofit firm behavior. The model begins by describing a nonprofit firm that produces two outputs,  $Q_1$  and  $Q_2$ . The firm's profit equation is

$$(P_1 + D_1 + G_1)Q_1 + (P_2 + D_2 + G_2)Q_2 + D + G - C(Q_1) - C(Q_2) = 0, \quad (8)$$

where  $P_i$  is the output price for  $Q_i$ ,  $D_i$  and  $G_i$  the private donations and government grants associated with each unit of  $Q_i$ ,  $D$  and  $G$  are fixed donations and grants, and  $C(Q_i)$  is the cost function for  $Q_i$ . Both cost functions are assumed identical for ease, although no important results depend on this. Cost is positive and twice-continuously

differentiable. Assume that  $dC/dQ_i > 0$  and  $d^2C/d^2Q_i > 0$ . All of the variables in Equation (8) are nonnegative.

The firm maximizes some utility function  $u = u(Q_1, Q_2)$ , subject to Equation (8). In contrast to a typical utility model, we make no assumptions on the sign of  $\partial u/\partial Q_i$ . Depending on the firm and the product, the marginal utility of  $Q_i$  may be positive, negative, or zero. However, we assume that  $\partial^2 u/\partial^2 Q_i \leq 0$ , and  $\partial^2 u/\partial Q_i \partial Q_j = 0$ .

The first-order conditions of the firm's Lagrangian function (with respect to each  $Q_i$ ) are

$$\frac{\partial L}{\partial Q_i} = \frac{\partial u}{\partial Q_i} + \lambda \left( P_i + D_i + G_i - \frac{dC}{dQ_i} \right) = 0, \quad i = 1, 2. \quad (9)$$

Solving the two equations in Equation (9) simultaneously with the zero-profit constraint yields the optimal values  $Q_1^*$  and  $Q_2^*$ . Now imagine a profit-maximizing, for-profit firm that produces the same outputs, although not necessarily at the same levels. This firm's first-order conditions are given by the two corresponding equations

$$\frac{\partial \pi}{\partial Q_i} = P_i + D_i + G_i - \frac{dC}{dQ_i} = 0, \quad i = 1, 2. \quad (10)$$

The two equations in Equation (10) yield  $\tilde{Q}_1$  and  $\tilde{Q}_2$ . Comparing Equations (9) and (10), we come to the following important conclusions:

- (i) If  $\partial u/\partial Q_i > 0$ ,  $Q_i^* > \tilde{Q}_i$ . In other words, nonprofits tend to produce more of a particular good than for-profits if they receive positive organizational utility from it. If this is a public good that is naturally underproduced by for-profits, the nonprofit may thus ameliorate a market failure. On the other hand, if there is no market failure, the nonprofit produces the good at inefficiently high levels.
- (ii) If  $\partial u/\partial Q_i < 0$ ,  $Q_i^* < \tilde{Q}_i$ . Furthermore,  $Q_i^* > 0$  only if at this point average revenues exceed average costs (that is,  $P_i + D_i + G_i > C(Q_i^*)/Q_i^*$ ). In other words, nonprofits cross-subsidize utility-enhancing activities with utility-lowering ones, as long as the latter turns a profit for the firm.

Several obvious empirical questions take shape from James' model. First of all, do the arts constitute a public good, and hence might nonprofit organizations enhance efficiency? Second, do nonprofit performing arts firms compete with for-profits? Third, what activities do performing arts firms engage in, such that cross-subsidization takes place?

The first question is somewhat outside the scope of this chapter; however, the earlier discussion of the purported nonuse, nonmarket values of the performing arts might be useful in deciding whether there is a true public-goods component to the performing arts. Throsby and Withers (1986) provide empirical evidence that people tend to believe the performing arts are public goods. Still, Netzer (1992) notes that most performing arts consumption is private and spills over relatively little onto nonpayers.

The efficiency of performing arts nonprofits was been treated earlier, in the context of production and costs. West (1987) argues that in general we will not see performing arts

nonprofits operate at minimum cost owing to the disincentives created by the nondistribution constraint, a prediction that James' model supports. And indeed, most studies suggest that performing arts firms in many countries tend to produce at suboptimal levels. For example, [Fazioli and Filippini \(1997\)](#) estimate that Italian theater companies operate at a scale that is too low to minimize cost. [Gray \(1997\)](#) uses theater data from Norway, and estimates that while large companies tend to operate beyond the point of minimum cost, small companies operate at a scale below this minimum. [Taalas \(1997\)](#) finds that Finnish theater companies tend to operate above minimum average cost due to overutilization of both labor and capital.<sup>13</sup>

Not surprisingly, the main addition to the theory of competition between nonprofits (of all types) and for-profits came from an extension to James' model. [Schiff and Weisbrod \(1991\)](#) note that (assuming for-profits rely on earned revenues), in the absence of corporate tax exemptions and donations, nonprofits would naturally be at a disadvantage compared with for-profits and thus unable to compete with them. This disadvantage would be due to inefficiently-high production levels for favored activities ( $Q_1$ ), and the disutility from unfavored activities ( $Q_2$ ). Empirical research has not looked at competition between nonprofit and for-profit performing arts firms, however. The closest treatment is probably [Gapinski \(1986\)](#), who shows that different performing disciplines (theater, opera, symphony, and dance) substitute for one another in consumers' demand decisions, and hence implicitly compete.

The idea of cross-subsidization of performing arts activities is easy to grasp intuitively. For example, symphony orchestras may perform obscure works at a loss, but make profits on "pops" concerts. The James model suggests that the former programming provides positive utility, the latter negative utility. This cross-subsidization relates closely to economies of scope: Can performing arts firms find multiple outputs that improve their financial position by more than they can by increasing the scale of their core activity? Two studies suggest that the performing arts do indeed enjoy economies of scope. [Lange and Luksetich \(1993\)](#) find that large American orchestras benefit from diverse types of performance activity. Similarly, [Fazioli and Filippini \(1997\)](#) find that Italian theaters can more efficiently provide multiple types of performances than they can just one type.

## 5. The cost disease

The discussion of performing arts production functions pointed out that the main input to the production of the performing arts is labor. The literature shows that, at least for American performing arts firms, labor costs have been rising rapidly.<sup>14</sup> This fact

<sup>13</sup> [Globerman and Book \(1974\)](#) estimate the optimal scale of services per year – where cost per performance is minimized – for Canadian orchestras and theater companies to be 115 and 210 performances, respectively. However, they do not report whether this falls below or above the actual scale of operations.

<sup>14</sup> [Faine \(1972\)](#) explains this by noting that American performing artists are heavily unionized.

broadly supports the so-called “cost disease” theory, which has been the subject of much research in cultural economics, and which deserves some special attention here. While this theory does not predict anything that is uniquely “nonprofit” – indeed, if it is valid, it should be so irrespective of sector – it is conventionally invoked in discussions of the nonprofit performing arts.

In a 1965 paper (and 1966 book), William Baumol and William Bowen (1965, 1966) argue that labor costs will plague the performing arts, for a simple reason: Real wages rise all over the economy with the increasing average levels of labor productivity; labor productivity (Baumol and Bowen assert) is largely stagnant in performing arts firms, which produce a service technologically unchanged over the past 100 years or so and thus have unchanged labor input requirements; therefore, costs will constantly rise over time to the performing arts. The implications of this argument are that these cost increases will necessitate more donated revenues, borrowing, production cuts, or all three.

Baumol and Baumol (1967) model the problem very simply as follows.<sup>15</sup> Imagine an orchestra that produces output  $Q_{1t}$  in period  $t$ , and a representative nonarts firm that produces  $Q_{2t}$ . Assume the production of each is a linear function of labor  $L$ , and the orchestra sees no labor productivity growth, while the other firm has constantly-compounded productivity growth at the rate of  $r$ . That is,

$$Q_{1t} = aL_{1t}, \quad Q_{2t} = bL_{2t}e^{rt}, \quad (11)$$

where  $a$  and  $b$  are constants. Assume that sector 2 dominates the economy, and hence the wage  $w_t$  – to which both firms are subject – also rises with productivity. Suppose, for example, that<sup>16</sup>

$$w_t = we^{rt}. \quad (12)$$

We can derive four main conclusions from this model:<sup>17</sup>

- (i) Costs for the orchestra rise over time, but do not for the other firm. Defining costs per unit of output as  $c_i = w_t L_{it} / Q_{it}$ , we see that  $c_1 = we^{rt} / a$  and  $c_2 = w / b$ .
- (ii) If labor proportions are constant between the sectors, the orchestra’s output falls over time, as a proportion of total output. Defining  $L_{1t} / L_{2t} = A$  where  $A$  is a constant, we see that  $Q_{1t} / Q_{2t} = aA / be^{rt}$ .
- (iii) On the other hand, if the output proportions are constant between the firms, labor will be systematically transferred to the orchestra. Assume that  $(b/a)(Q_{1t} / Q_{2t}) = K$ , where  $K$  is constant, and that  $L_{1t} + L_{2t} = L_t$ . Then,  $L_{1t} = L_t K e^{rt} / (1 + K e^{rt})$  and  $L_{2t} = L_t / (1 + K e^{rt})$ . As  $t$  increases,  $L_{1t}$  goes to  $L_t$  and  $L_{2t}$  goes to zero.

<sup>15</sup> For an alternative modeling approach, see Brooks (1998).

<sup>16</sup> The assumption that wages rise at the same rate as productivity is made only for modeling simplicity. It is not a necessary assumption for the conclusions that follow, only that wages rise monotonically with productivity.

<sup>17</sup> These conclusions are modified by Keren (1972) from Baumol’s original article.

(iv) Economic growth will suffer over time if the labor force does not grow, output proportions do not change, and unproductivity persists in the orchestra. Define  $q_t$  as a weighted sum of the outputs of the two firms:  $\beta_1 Q_{1t} + \beta_2 Q_{2t} = q_t$ .

From Equations (11) and the derivations above, we obtain

$$q_t = \frac{L_t(K\beta_1 a + \beta_2 b)e^{rt}}{1 + Ke^{rt}}. \quad (13)$$

Note that

$$q'_t = \frac{\partial q_t}{\partial t} = \frac{rL_t(K\beta_1 a + \beta_2 b)e^{rt}}{(1 + Ke^{rt})^2} \quad (14)$$

and thus that  $\lim_{t \rightarrow \infty} (q'_t/q_t) = 0$ .

There are two main testable hypotheses about performing arts firms that proceed from the model. The first is that the operating costs of performing arts firms rise over time because of wage increases. The second is that the performing arts sector is shrinking relative to the rest of the economy.<sup>18</sup> If these hypotheses are not supported, we may then conclude that

- (i) a force outside the model – such as rising earned or donated revenues – must be ameliorating the cost problem, or
- (ii) the cost disease theory is somehow flawed.

These hypotheses are the themes on which most of the empirical studies of cost disease have varied.

Raw data provide mixed support for these hypotheses. For example, looking at data on American performing arts nonprofits, McCarthy et al. (2001) show that over the ten-year period ending in 1997, real costs to opera and dance companies increased, but costs to theater companies and musical ensembles fell. Figure 2 above shows that from 1982–1997 donations rose as a portion of total revenues for theater and music firms, but fell for dance companies. And while Figure 4 shows that unearned revenues have indeed generally increased for performing arts firms over the 1990s, it also shows that these increases have been more-or-less proportional to increases in total revenues. Not surprisingly, then, economists have come to mixed conclusions about the degree to which cost disease has affected the performing arts, or indeed exists at all [Throsby (1994)]. Some empirical work supports the proposition, while some does not.

The work of Baumol himself consistently points to cost disease as a real and ongoing problem in the performing arts. For example, Baumol and Baumol (1980) show that costs in the performing arts have increased disproportionately to rises in the price level, except in periods of high inflation. Baumol and Oates (1972) hold that the principle also applies historically, arguing that low real wages for actors in the past are responsible for the success of, among other things, English Renaissance theater. Baumol has also extended the concept beyond the nonprofit performing arts to other labor intensive

<sup>18</sup> This hypothesis assumes the constancy of labor proportions between sectors, as opposed to the constancy of output proportions.

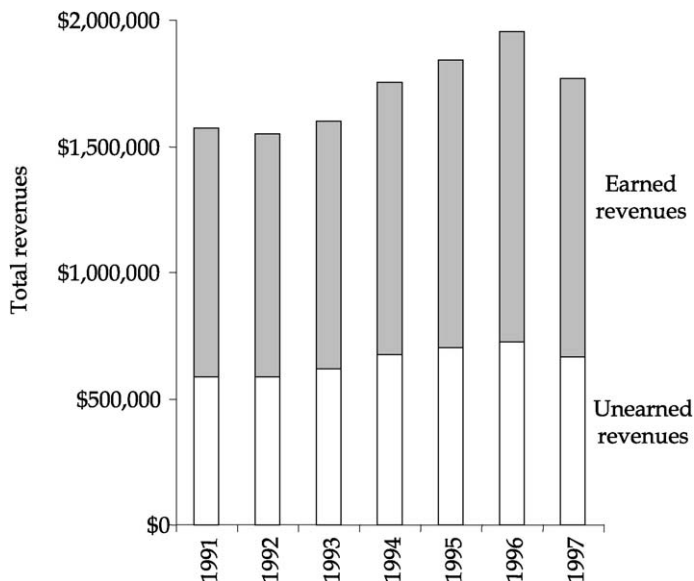


Figure 4. Total and earned average revenues of US performing arts nonprofits, 1991–1997. Source: McCarthy et al. (2001).

industries, presenting evidence of the cost disease in areas such as health care [Baumol and Baumol (1993)], education [Baumol and Baumol (1996)], and libraries [Baumol and Blackman (1983)].

Other writers have encountered some evidence of the cost disease as well. Gapinski (1984) and Fazioli and Filippini (1997) find that technological progress in English and Italian theater (respectively) is very weak, compared to other industries, validating this modeling assumption. Felton (1994) finds that rising labor costs to US orchestras have led to cheaper programming. She notes, however, that such production adjustments belie the inevitability of cost disease problems. Throsby (1996) agrees with Felton on both of these points in his study of labor costs in the Australian performing arts in the 1980s and 90s.

However, a number of writers have argued either that the cost disease does not exist, or that it does not pose a significant problem for the performing arts. For example, Cowen (1996) believes that performing arts firms should be able to greatly enhance labor productivity vis-à-vis recording technologies (that is, a single musician can reach millions of customers at once with a compact disk, as opposed to a live performance).<sup>19</sup> Baumol would not agree with the technology argument, however; they use data on

<sup>19</sup> The opposite side of this coin is that technology simply lowers the demand for labor by firms. For example, Colonna, Kerns and Anderson (1993) argue that synthesized technologies have greatly lowered the employment of certain classes of performing musicians.

American television and film to show that mass media can put off cost disease pressures, but only temporarily. Labor productivity might also be enhanced if performers and other workers can be combined in ways that lower their cost per unit of performance. For example, theater companies can employ tactics such as programming works with fewer actors, or actors in multiple roles.<sup>20</sup>

The demand side is absent from the model presented above, but some authors [e.g., Cowen (1996)] would say that this is not appropriate: Performing arts firms can generally change their revenue positions by changing programming in ways that stimulate demand, thus driving up prices and quantities. Other demand-side cost-disease strategies may exist as well. Peacock (1996) notes that income effects from general economic growth could totally swamp cost disease pressures. This would be especially likely, if, as seems probable, the performing arts have a high income elasticity. Kesenne (1994), however, notes that this would be mitigated by the so-called Linder effect [Linder (1970)]: As incomes rise, so does the opportunity cost of time-intensive leisure activities, such as attending live performing arts events. To exploit income increases optimally – even in spite of Linder effects – Brooks (1997a) suggests that performing arts firms market the luxury aspects of their products so as to maximize income elasticity.<sup>21</sup>

On balance, it seems fair to say that the cost disease is a threat to performing arts firms, but that there exist enough mitigating circumstances that these firms don't generally face any clear and present danger of extinction. Principal among these circumstances is the role of demand increases accompanying rising rates of prosperity and education. Strategies for demand stimulation represent a promising avenue for economists, managers, and policy researchers on the performing arts.

## 6. Revenues of performing arts firms

As discussed earlier, nonprofit organizations of all types receive their revenues from three sources: earned income, private donations, and government subsidies. Surveying the sector in 22 countries worldwide, Salamon et al. (1999) find that earned income makes up about half of revenues, governments underwrite another 40 percent, and private giving constitutes about 10 percent. This masks considerable variance, however. For example, earned income represents 85 percent of nonprofit revenues in Mexico, but just 16 percent in Ireland. This discrepancy is mirrored by public-sector funding of nonprofits: 77 percent of revenues in Ireland, but just 9 percent in Mexico. Private

<sup>20</sup> Certain types of firms (for example, symphony orchestras) have less recourse than others to this kind of labor flexibility, and consequently would be expected to turn to other ways of fighting off cost disease pressures, such as lobbying the government for subsidies.

<sup>21</sup> We should not overlook the fact that technology itself might impact demand for the live performing arts. Heilbrun (1993), for example, believes that technology has hurt the demand for symphony orchestra concerts by improving electronically reproduced substitutes.



philanthropy sees far more uniformity worldwide, where in most countries it makes up between 5 and 15 percent of all revenues.

For the average American nonprofit performing arts firm, 59 percent of income is earned, about 36 percent is donated, and about 5 percent comes directly from government [US Department of Commerce, Bureau of the Census (1997)]. Each of these revenue sources warrants discussion.

### 6.1. Earned revenues

Americans in 1997 spent more than \$10 billion on performing arts events, substantially more than on tickets to movies or sporting events [National Endowment for the Arts (1998)]. As already discussed, this amounts to between 50 and 60 percent of the total revenues of nonprofit performing arts firms. In some countries, the performing arts have an even higher proportion of earned revenues. In the late Soviet era, for example, musical organizations in the USSR (“nonprofits” in function, if not by legal definition) apparently earned upwards of 90 percent of their income at the box office [Rubinstein, Baumol and Baumol (1992)].

As such, discussions of earned revenues might begin by looking at demand for the performing arts. Many authors have treated this subject empirically, and it is the focus of another chapter in this volume.<sup>22</sup> Here, I simply note that equations approximating demand often take some variant of the form

$$\ln Q_{it} = \alpha + \beta_P \ln P_{it} + (\ln \bar{P}_{it})\beta_{\bar{P}} + \beta_Y \ln I_{it} + X_{it}\beta_X + \varepsilon_{it}, \quad (15)$$

where  $Q_{it}$  is the quantity demanded of performing arts firm  $i$  at time  $t$ ,  $P_{it}$ ,  $\bar{P}_{it}$  and  $I_{it}$  are own price, substitutes’ prices, and income levels, respectively; and  $X_{it}$  is a vector of other relevant variables. The coefficients are interpreted as the relevant elasticities. Unfortunately, pricing among nonprofits is general highly nonlinear, reflecting firm strategy in the face of client heterogeneity. Equation (15) estimates some notion of “average” elasticity, but such a concept may not be meaningful. Hence inquiry into earned revenues from the firm side does well to study actual pricing schemes. A number of authors have done so in the case of the performing arts. For example, Huntington (1993) shows that in the mid-1980s about half of theaters in Great Britain charged differential ticket prices based on seating and that this was generally a revenue-enhancing strategy. Seaman (1985) finds that in the US, differential pricing is most common for opera, and least for theater. Blaug (1978), looking specifically at Covent Garden, concludes that “superstar” programming explains much of the difference in prices between venues and events.

Many performing arts firms appear to bundle donations with their ticket prices, asking for voluntary payments that are higher than the minimum for event admission.<sup>23</sup>

<sup>22</sup> See Chapter 14 by Seaman in this volume.

<sup>23</sup> Some authors refer to this phenomenon as “voluntary price discrimination” [e.g., Hansmann (1981)].

O'Hagan and Purdy (1993) and Seaman (1985) provide evidence of this in Europe and the US, respectively. Luksetich and Lange (1995) come to the interesting conclusion that while smaller American orchestras frequently underprice tickets in order to solicit donations, on average this is a money-losing strategy. Kushner and Brooks (2000) extend the idea to street performance, noting that voluntary price discrimination in its purest form (i.e. where all payment is totally voluntary) characterizes support for "buskers".

Another common strategy is one of offering free admission to certain events in order to build audience bases, but this has received little attention by economists. However, given the general finding that performing arts demand is inelastic [Brooks (1997a)], we might question whether lowering prices – even to zero – would really have an effect on attendance.<sup>24</sup> Indeed, Kolb (1997) reports on a survey of students, showing that price is not the issue in performing arts attendance; the real problem seems to be that many young people simply find the performing arts boring. Globerman (1978) finds that demand elasticity can be increased – and thus pricing schemes can be more effective – through the dissemination of better price information. On the other hand, demand inelasticity may be a blessing for some firms losing donations: Touchstone (1980), for example, shows that for American performing arts organizations, a complete loss of donations would require more than doubling ticket prices. Given low demand elasticity, this would decrease audiences by less than 10 percent for theater, less than 15 percent for ballet and opera, and less than 20 percent for orchestras.

## 6.2. Donated revenues

Nonprofit economic theory has occasionally focused on philanthropic behavior. Andreoni (1989, 1990), the most prominent theorist in this area, has developed models that separate out purely altruistic giving from that which is motivated instead by a "warm glow". Other motives have been examined as well. Particularly in the case of religious giving, authors [for example, Asheim (1991)] have modeled deontological, or duty-based, altruism in which a certain giving threshold must be reached before an agent can derive utility from private consumption. And consistent with much of the sociological literature on altruism, others [for example, Rose-Ackerman (1996)] have modeled altruism related in some measure to social status. The data presented earlier suggested that private charitable giving is substantial internationally, but especially large in the United States, where it amounted to \$132 billion in 1997 [Independent Sector (2001)]. 70.1 percent of American households gave in 1998; 11.5 percent made contributions to the arts. The arts consistently represent less than 4 percent of household donations (\$4.4 billion), but a much greater part of corporate and foundation philanthropy. In 1997, private donations from all sources to the arts in the US were about \$10.6 billion [AAFRC (1998)].<sup>25</sup>

<sup>24</sup> It may be that demand is too inelastic and weak to intersect with supply at any nonnegative prices for many performing arts firms [Brooks (1997b)].

<sup>25</sup> See further Schuster's Chapter 36 in this volume.

Between the theoretical interest in philanthropy and the high stakes for nonprofits, it is not surprising that several nonprofit scholars have studied nonprofit fundraising. Among the most influential – and useful for studying performing arts firms – is [Rose-Ackerman's \(1982\)](#) model. She begins with a measure  $\bar{D}_{it}$ , which represents the average expected donation to nonprofit firm  $i$  in period  $t$ . This measure is a function of the proportion of the population (where population is  $n$  individuals) solicited for donations  $k_{it}$  (where  $0 \leq k_{it} \leq 1$ ) and the number of similar organizations  $z_t$  with which firm  $i$  competes for funding in period  $t$ . That is,

$$\bar{D}_{it} = \bar{D}_{it}(k_{it}, z_t). \quad (16)$$

Assume that those people most likely to donate are known and are solicited first, and that more competition drives down the average donation. Then,  $\partial \bar{D}_{it} / \partial k_{it} < 0$ , because wider fundraising appeals have lower average returns, and  $\partial \bar{D}_{it} / \partial z_t < 0$ . Defining the (constant) cost per fundraising contact as  $c$ , dropping subscripts, and assuming firm  $i$  has a one-year time-horizon, its problem is  $\max_k \{kn[\bar{D}(k, z) - c]\}$ . The first-order conditions yield  $k(\partial \bar{D} / \partial k) + \bar{D} = c$ . Dividing this last expression by  $\bar{D}$  yields

$$w = \mu + 1, \quad (17)$$

where  $\mu$  represents the elasticity of the average donation with respect to fundraising penetration, and  $w = c/\bar{D}$  is the budget proportion spent on fundraising.

Equation (17) leads to two main conclusions about nonprofit spending on fundraising, both of which are sensible. First, differentiating the first-order condition with respect to the elasticity gives  $\partial w / \partial \mu > 0$ . That is, as the sensitivity of the average gift to fundraising penetration decreases – for example, if more people become potential strong donors for some reason – fundraising expenditures will increase as a proportion of the nonprofit's budget. Second, differentiating  $w = c/\bar{D}$  with respect to  $z$  yields  $\partial w / \partial z = (-c/\bar{D}^2)(\partial \bar{D} / \partial z) > 0$ . That is, more competition leads to higher budget shares devoted to fundraising.<sup>26</sup>

Apart from [Baumol and Bowen's \(1966\)](#) study of the performing arts and the papers on the importance of private philanthropy in combating the cost disease, there has been relatively little empirical and theoretical economic research on fundraising specifically for the nonprofit performing arts. What little work that has appeared has generally focused on fundraising yields and funding diversification. [McCarthy et al. \(2001\)](#) show that, while arts philanthropy continues to rise, yields to fundraising are falling to performing arts firms, probably due to the explosion in the number of firms – in other words  $z$  is increasing which, consistent with Rose-Ackerman's model, should be driving up fundraising budgets. [Brooks \(1997b\)](#) looks at American symphony orchestra budgets in the 1980s and 90s, and finds that there is a significant difference between large and

<sup>26</sup> These conclusions assume that donors' utility is not affected by  $w$ . [Rose-Ackerman \(1982\)](#) also looks at cases in which donors dislike high  $w$ , and alternatively where they see a high  $w$  as a valuable investment in donor cultivation.

small orchestras: While small (in budget size) orchestras generally find fundraising to be a good investment, many large budget orchestras don't even earn back their fundraising budgets. He attributes this to the fact that large orchestras have a higher  $k$  – they have saturated a larger proportion of their potential donor pools than small orchestras. Nelson (1983) looks at ballet companies, and charts the way that funding has moved over time from box office revenues, to private giving, to greater reliance on corporate philanthropy.

### 6.3. Fundraising innovations

Fundraising and philanthropy in the performing arts are at an interesting frontier with the advent of innovations such as “venture philanthropy” and “e-philanthropy”. Both of these phenomena have received popular attention and deserve mention in the current context.

Venture philanthropy refers to foundation giving in search of large, immediate payoff outcomes across a wide variety of potential activities [Goodale (2000)]: the maximum philanthropic “bang for the buck”. The idea sounds consistent with the rising popularity of venture capital funding in the for-profit economy. In practice, however, there seems to be less here than meets the eye. In 2001, for example, there were just 42 foundations engaged in this type of giving, and they represented just 0.2 percent of all foundation giving [Venture Philanthropy Partners (2002)]. 25 percent of venture philanthropists made no grants at all, and only half made grants to nonprofit organizations, let alone performing arts nonprofits [Jana (2001)].

Broadly speaking, e-philanthropy refers to donations given over the Internet, and takes three basic forms. First, some charities set up their own systems to collect credit card donations on their websites. Second, there are *donation portals*, or Internet clearinghouses for multiple charities which administer funds and donor information for a commission. Third, there are *charity malls*, which are retail establishments that advertise on nonprofit websites and dedicate a portion (usually 5 percent) of resulting Internet sales to the charity. Hopes in the e-commerce world have been high for the future of e-philanthropy. Harvard Business School's Initiative on Social Enterprise, for example, projected that by 2010, one third of all philanthropy would take place over the Internet [Hart (2001)]. This may be a heroic projection, however, given that 1999 online giving amounted to \$10 million, or less than \$1 for every \$13,000 in traditional giving. And indeed, Feller (2001) argues that e-philanthropy firms are seeing far less growth than expected, with some e-philanthropy firms (such as former leader Charitableway) going out of business along with their for-profit e-commerce counterparts.

Economists have not yet studied these fundraising innovations in the context of the performing arts. Any work that is forthcoming would do well to begin by putting hard numbers to these mechanisms and hence to give an idea as to whether firms should invest time and resources in these areas. It may be that venture philanthropy and e-philanthropy are simply not suited to the performing arts, perhaps because donors tend also to be consumers, to whom performing arts firms already have ample exposure.

It is also possible that these organizations, like many for-profits, require closer consumer contact and information than the Internet allows. Potentially instructive for the case at hand is the failure of commercial Internet art auctions [Brooks (2003)].

#### 6.4. Government funding

The final source of funding for performing arts nonprofits is the public sector. Public sector funding comes in two main varieties: direct and indirect. Direct subsidies are payments by governments to arts organizations. Indirect subsidies are taxes forgone on private contributions to these organizations. They result from tax laws that allow charitable donations to be deducted from taxable personal or corporate income before income taxes are calculated.

Direct government subsidies are ubiquitous internationally whereas indirect subsidies are less common, for the simple reason that most countries do not have provisions in their tax laws that allow tax deductions for arts contributions [Zimmer and Toepler (1999)]. Schuster (1999) notes several exceptions, including Australia and Romania (both of which have arts deductions similar to the American system), and Chile (which has an even more generous system of tax credits – not deductions – for corporate arts support). In the United States, however, indirect funding is the primary source of government support. Indeed, indirect federal subsidies to the arts in the US in 1999 (about \$1.7 billion) outweighed direct federal subsidies by about 16-to-1.<sup>27</sup> And indirect subsidies from state and local taxes would augment this figure substantially.

The political economy of government arts funding is the subject of other chapters in this book, and represents a substantial literature in and of itself.<sup>28</sup> Two points bear making here, however, because they explicitly tie government funds to the way nonprofits function in the performing arts. First, several authors have looked at the way government funding affects management decisions. For example, Austen-Smith (1980) shows that government support for British theater companies pushes output down, but variety of repertoire up; further, he predicts that the level of government (central versus local) making grants to performing arts firms will influence output decisions [Austen-Smith (1984)]. In one of the rare looks at the privatization of cultural organizations, Klaic (1998) finds that the delinking of European performing arts organizations from the public sector has been more managerial than financial.

The second issue regards the extent to which public sector funds to performing arts nonprofits displace, or “crowd out”, private donations. One of the most provocative predictions from all of nonprofit economic theory is based on a variant of the public-goods

<sup>27</sup> These data are based on National Endowment for the Arts allocations (see [www.arts.endow.gov](http://www.arts.endow.gov)), total giving to the arts in 1999 as calculated by AAFRC in *Giving USA*, and assuming the average marginal tax rate of 14.8 percent estimated by the US Congressional Budget Office (see [www.cbo.gov](http://www.cbo.gov)). If, as seems likely, the average marginal tax rate for givers to the arts is higher than the national average, this ratio would actually be higher than 16:1.

<sup>28</sup> See further Chapter 34 by van der Ploeg and Chapter 35 by Netzer in this volume.

model outlined earlier, and can be summarized as follows. Suppose a representative agent  $i$  has utility  $u_i = u_i(x_i, D_i, G)$ , where  $D_i$  represents private donations to a public charitable good,  $G$  is the total amount of the public good provided, and  $x_i$  is a composite private good. Assume that utility is concave in all arguments, is twice-continuously differentiable in each, and that at least one argument has a nonzero second derivative. Note that  $D_i$  enters the utility function twice: once through the public good (altruism), and once for the act of giving itself (the “warm glow” from giving). Assume the agent spends all his income  $m_i$  on  $x_i$  and  $D_i$ .  $G$  is financed through government spending  $\bar{G}$  and the agent’s private donations, and is financed through a lump-sum tax  $T$ . Thus, dropping the subscripts, the agent’s problem is  $\max_{x, D} \{u(x, D, G)\}$ , subject to  $m = x + D + T$ ;  $G = \bar{G} + D$ ;  $\bar{G} = T$ ;  $x, D, G \in \Re_+$ .

Folding the constraints into the objective function, the agent’s problem can be restated as a function of  $D$ . The first-order condition is:

$$\frac{du}{dD} = u' = -u_x + u_D + u_G = 0, \quad (18)$$

where  $u_s$  is the partial derivative of  $u$  with respect to argument  $s$ . The second-order condition indicates that  $u(D^*)$  is a unique global maximum. The Implicit Function Theorem tells us that

$$\frac{\partial D}{\partial \bar{G}} = -\frac{\partial u' / \partial \bar{G}}{\partial u' / \partial D} = -\frac{u_{xx} + u_{GG}}{u_{xx} + u_{DD} + u_{GG}}. \quad (19)$$

Equation (19) makes three points:

- (i)  $-1 \leq \partial D / \partial \bar{G} \leq 0$ . That is, government grants will have a crowding-out impact on giving, if any at all.
- (ii) If  $u_{xx} \neq 0$  and/or  $u_{GG} \neq 0$ , and giving provides no “warm glow” (so  $u_D = u_{DD} = 0$ ), grants will crowd out donations dollar-for-dollar.
- (iii) If  $u_{xx} \neq 0$  and/or  $u_{GG} \neq 0$ , and  $u_{DD} < 0$ , grants will crowd out donations fractionally. Bergstrom, Blume and Varian (1986) also note that in a multi-agent model, in which not all agents give voluntarily, crowding-out will also be fractional.

In general, these predictions square with the empirical literature on nonprofits, which finds that a dollar in government grants displaces between 10 and 50 cents in private giving on average [Steinberg (1993)]. However, the few articles on performing arts nonprofits present slightly more complicated results. While Brooks (1999) finds that American orchestras tended to see no significant crowding out, he later (2000) estimated that the true effect was nonlinear; low levels of government funding crowded in donations, while high levels crowded them out. This is consistent with work on non-profit “public” radio stations [e.g., Kingma (1989)], and may be due to the prevalence of “matching grants” in the arts and culture, in which governments match private donations with government money. Note that adding the matching-grant constraint  $\bar{G} = a + bD$ , where  $a$  is a constant lump-sum grant and  $bD$  is a grant for which \$1 in  $D$  is matched

by  $b$  dollars in government money, changes Equation (19) to

$$\frac{\partial D}{\partial a} = \frac{-(1+b)(u_{xx} + u_{GG})}{(1+b)^2(u_{xx} + u_{GG}) + u_{DD}}, \quad (20)$$

and

$$\frac{\partial D}{\partial b} = \frac{-(1+b)D(u_{xx} + u_{GG}) + u_x - u_G}{(1+b)^2(u_{xx} + u_{GG}) + u_{DD}}. \quad (21)$$

Not surprisingly, Equation (20) must be negative. Equation (21), however, may be either positive or negative, suggesting that the matching mechanism may be effective for reversing the crowding out problem. Clearly, much interesting theoretical and empirical work remains to be done on how such institutional arrangements might affect the efficacy of government funding to the performing arts (as well as other parts of the nonprofit economy).

## 7. Summary and future research directions

This chapter has traced the development of the economics literature on nonprofit performing arts firms through some of the most important milestones in nonprofit economic theory. In covering more than 100 papers to date on the subject, we observe contributions to the nonprofit literature as well as an incomplete research agenda. The main areas I see for potential future work on the nonprofit performing arts – to complement and expand general knowledge of the economics of nonprofits – are in updating and expanding older studies, theory building, new empirical applications, and the development of policy and management implications.

Much of the work on performing arts cost and production is based on data from the 1970s and 80s; there appears to be a significant drop-off in empirical work focusing on the 1990s and beyond. Given the rapid changes in production conditions in other parts of the economy, simply re-estimating many of the cost and production measures in the earlier studies surveyed here would be illuminating. Indeed, given that much of the cost disease argument was predicated on resistance to change in performing arts production conditions, new studies could potentially add to the cost disease debate in addition to providing current values of measures such as input elasticities and efficient scales of production. Second, the overwhelming majority of performing arts studies have focused on the United States, Western Europe and Australia, with far less attention to the performing arts in other parts of the world. Naturally, much of this has to do with the availability of data. Nevertheless, data collection efforts and comparative studies featuring developing and transition economies would add significant value to the literature. Third, studies have focused far more on the professional performing arts than activity among amateurs, once again due almost certainly to data availability. Recent work makes it clear that amateur activity affects many more people than previously thought [Brooks (2002)]. Hence, any policy work on the importance of performing arts participation should be accompanied by inquiry into grassroots performing arts organizations.



Several areas of work stand out in the development of the theory of the nonprofit performing arts firm. First, relatively little work rigorously connects the public-goods theory of philanthropic resources to the actual emergence of performing arts nonprofits. Do high fixed costs explain why these firms require donations to survive, and do tax-deductibility and contract failure explain why these donations only flow to firms organized on a not-for-profit basis? Second, it seems reasonable on balance to conclude that the cost disease has not presented increasingly serious problems for the performing arts. Theoretical work on the cost disease could work to understand this phenomenon in more of an intertemporal general equilibrium context, such that the arts interface with education and tastes, and the supply and demand for the performing arts interact over time. Finally, more theory-building is needed on the relationship between revenue sources. Earlier, I suggested the way that just one institution – government matching grants – might defray the “crowding out” relationship between subsidies and gifts. What about other institutional mechanisms? What about the relationships between other sources of revenues? One large unanswered question, for example, involves the degree to which crowding out exists between donors themselves.

Many interesting empirical questions remain unanswered in the nonprofit performing art literature. Examples include the following:

- Tests of the public-goods theory of nonprofits might try to establish the existence of the purported “nonuse” benefits of the performing arts (existence value, bequest value, etc.). At present, these values are empirically unsubstantiated.
- Production functions in the performing arts have not generally included unpriced resources. What is the production value of, say, volunteer time?
- Quality could receive a comprehensive, empirical treatment. Taxonomies of quality such as [Throsby’s \(1990\)](#) could be subjected to scrutiny, by, for example, comparing how alternative definitions of quality relate to one another and predict firm or audience behavior.
- Competition between nonprofits and for-profits is not well understood in the performing arts. To what extent are their products substitutes? In what disciplines does the most competition exist?
- It appears that fundraising yields may vary dramatically by discipline, firm size, and other characteristics. What does this tell us about different firm objectives, within the performing arts, as per [Steinberg \(1986\)](#)?

Finally, the work on nonprofit performing arts economics could be better connected to the literature on arts management and cultural policy. For example, the work on pricing policies and demand suggests that consumers are extremely insensitive to price changes for performing arts events. This fact, as well as strategies to change it, would be invaluable to arts administrators. Second, the innovations in philanthropy reviewed earlier are the subject of much discussion among managers; they deserve some theoretical and empirical treatment by economists. Finally, specifically regarding non-US performing arts markets facing privatization movements, what can economists tell cultural policymakers about what performing arts firms can expect when public subsidies are ended, or American-style tax incentives for private charity are implemented? In all of these cases



and others, economists could take their analysis several steps forward in practicality and accessibility for more general audiences.

## References

- AAFRC (1998). Giving USA. Center on Philanthropy, Indianapolis, IN. [www.aafc.org](http://www.aafc.org).
- Abbé-Decarroux, F. (1994). "The perception of quality and the demand for services". *Journal of Economic Behavior and Organization* 23, 99–107.
- Andreoni, J. (1988). "Why free ride? Strategies and learning in public goods". *Journal of Public Economics* 37, 291–304.
- Andreoni, J. (1989). "Giving with impure altruism: Applications to charity and Ricardian equivalence". *Journal of Political Economy* 97 (6), 1447–1458.
- Andreoni, J. (1990). "Impure altruism and donations to public goods: A theory of warm-glow giving". *Economic Journal* 100 (401), 464–477.
- Andreoni, J. (1998). "Toward a theory of charitable fundraising". *Journal of Political Economy* 106 (6), 1186–1213.
- Arnold, M. (1869). *Culture and Anarchy*. Cambridge University Press, Cambridge.
- Asheim, G.B. (1991). "Unjust intergenerational allocation". *Journal of Economic Theory* 54, 350–371.
- Austen-Smith, D. (1984). "Subsidies to the arts with multiple public donors". *Economic Record* 60 (171), 381–389.
- Austen-Smith, J. (1980). "On the impact of revenue subsidies on repertory theater policy". *Journal of Cultural Economics* 4 (1), 9–18.
- Baumol, H., Baumol, W.J. (1967). "Macroeconomics of unbalanced growth: The anatomy of urban crisis". *American Economic Review* 57 (3), 495–502.
- Baumol, H., Baumol, W.J. (1980). "On finances of the performing arts during stagflation: Some recent data". *Journal of Cultural Economics* 4 (2), 1–14.
- Baumol, H., Baumol, W.J. (1993). "Health care, education and the cost disease: A looming crisis for public choice". *Public Choice* 77, 17–28.
- Baumol, H., Baumol, W.J. (1996). "Children of performing arts: The economic dilemma". *Journal of Cultural Economics* 20, 183–206.
- Baumol, W.J., Blackman, S.A.B. (1983). "Electronics, the cost disease, and the operation of libraries". *Journal of the American Society for Information Science* 34 (3), 181–191.
- Baumol, W.J., Bowen, W.G. (1965). "On the performing arts: The anatomy of their economic problems". *American Economic Review* 50 (2), 495–502.
- Baumol, W.J., Bowen, W.G. (1966). *Performing Arts: The Economic Dilemma*. Twentieth Century Fund, New York.
- Baumol, W.J., Oates, W.E. (1972). "The cost disease of the personal services and the quality of life". *Skandinaviska Enskilda Banken Quarterly Review* 2, 44–54.
- Benhamou, F. (2000). "The opposition between two models of labour market adjustment: The case of audiovisual and performing arts activities in France and Great Britain over a ten year period". *Journal of Cultural Economics* 24, 301–319.
- Bergstrom, T., Blume, L., Varian, H. (1986). "On the private provision of public goods". *Journal of Public Economics* 29, 25–50.
- Bilodeau, M., Slivinski, A. (1998). "Rational nonprofit entrepreneurship". *Journal of Economics and Management Strategy* 7 (4), 551–571.
- Blaug, M. (1978). "Why are Covent Garden seat prices so high?" *Journal of Cultural Economics* 2 (1), 1–20.
- Brooks, A.C. (1997a). "Toward a demand-side cure for cost disease in the performing arts". *Journal of Economic Issues* 31 (1), 197–207.
- Brooks, A.C. (1997b). *Economic Strategies for Orchestras*. Symphony Orchestra Institute, Evanston, IL.

- Brooks, A.C. (1998). *Arts, Markets, and Governments: A Study in Cultural Policy Analysis*. RAND, Santa Monica, CA.
- Brooks, A.C. (1999). "Do public subsidies leverage private philanthropy for the arts? Empirical evidence on symphony orchestras". *Nonprofit and Voluntary Sector Quarterly* 28 (1), 32–45.
- Brooks, A.C. (2000). "Public subsidies and private giving: Crowding out, crowding in, or both?" *Journal of Policy Analysis and Management* 19 (3), 451–464.
- Brooks, A.C. (2002). "Artists as amateurs and volunteers". *Nonprofit Management and Leadership* 13 (1), 5–16.
- Brooks, A.C. (2003). "Contingent valuation and the winner's curse in internet art auctions". In: Mossetto, G., Vecco, M. (Eds.), *Economics of Art Auctions*. FrancoAngeli, Milan, pp. 153–170.
- Brooks, A.C., Kushner, R.J. (2001). "Cultural districts and urban development". *International Journal of Arts Management* 3 (2), 4–15.
- Brown, E., Slivinski, A. (2002). "Nonprofit organizations and the market". In: Powell, W., Steinberg, R. (Eds.), *The Nonprofit Sector: A Research Handbook*, second ed. Yale University Press, New Haven, CT.
- Colonna, C.M., Kerns, P.M., Anderson, J.E. (1993). "Electronically produced music and its economic effect on the performing musician and music industry". *Journal of Cultural Economics* 17 (2), 69–76.
- Cornes, R., Sandler, T. (1984). "Easy riders, joint production, and public goods". *Economic Journal* 94, 580–598.
- Cowen, T. (1996). "Why I do not believe in the cost disease: Comment on Baumol". *Journal of Cultural Economics* 20, 207–214.
- DiMaggio, P. (1982). "Cultural entrepreneurship in nineteenth century Boston, part I: The creation of an organizational base for high culture in America". *Media, Culture and Society* 4, 33–50.
- DiMaggio, P. (1984). "The nonprofit instrument and the influence of the marketplace on policies in the arts". In: McNeil Lowry, W. (Ed.), *Public Policies and the Arts in the United States*. Prentice-Hall, Englewood Cliffs, NJ, pp. 57–99.
- DiMaggio, P. (1987). "Nonprofit organizations in the production and distribution of culture". In: Powell, W. (Ed.), *The Nonprofit Sector: A Research Handbook*. Yale University Press, New Haven, CT, pp. 341–361.
- Faine, H.R. (1972). "Unions and the arts". *The American Economic Review* 62 (2), 70–77.
- Fazioli, R., Filippini, M. (1997). "Cost structure and product mix of local public theatres". *Journal of Cultural Economics* 21, 77–86.
- Feller, G. (2001). "Charitableway gives up on online giving". *Nonprofit World* 19 (4), 35–37.
- Felton, M.V. (1994). "Evidence of the existence of the cost disease in the performing arts". *Journal of Cultural Economics* 18, 301–312.
- Frank, R.H. (1996). "What price the high moral ground?". *Southern Economic Journal* 63, 1–17.
- Frey, B.S. (1997). "Evaluating cultural property: The economic approach". *International Journal of Cultural Property* 6 (2), 231–246.
- Gapinski, J.H. (1980). "The production of culture". *Review of Economics and Statistics* LXII (4), 578–586.
- Gapinski, J.H. (1984). "The economics of performing Shakespeare". *American Economic Review* 74 (3), 458–466.
- Gapinski, J.H. (1985). "Do the nonprofit performing arts optimize? The moral from Shakespeare". *Quarterly Review of Economics and Business* 25 (2), 27–37.
- Gapinski, J.H. (1986). "The lively arts as substitutes for the lively arts". *American Economic Review* 76 (2), 20–25.
- Globerman, S. (1978). "Price awareness in the performing arts". *Journal of Cultural Economics* 2 (2), 27–41.
- Globerman, S., Book, S.H. (1974). "Statistical cost functions for performing arts organizations". *Southern Economic Journal* 40, 668–671.
- Goodale, G. (2000). "Selfless in Seattle: High-tech philanthropists are remaking the city's arts scene". *Christian Science Monitor* (Boston, MA), August 18, p. 13.
- Goudriaan, R., Pommer, E. (1987). "Productivity trends in the subsidized performing arts". In: Grant, N., Hendon, W., Owen, V. (Eds.), *Economic Efficiency in the Performing Arts*. Association for Cultural Economics, Akron, OH, pp. 16–30.

- Gray, C. (1997). "Art costs and subsidies: The case of Norwegian performing arts". In: Towse, R. (Ed.), *Cultural Economics: The Arts, the Heritage and the Media Industries*, vol. 1. Edward Elgar, Cheltenham, pp. 337–343.
- Hall, P.D. (1987). "A historical overview of the private nonprofit sector". In: Powell, W. (Ed.), *The Nonprofit Sector: A Research Handbook*. Yale University Press, New Haven, CT, pp. 341–361.
- Halter, A.N., Carter, H.O., Hocking, J.G. (1957). "A note on the transcendental production function". *Journal of Farm Economics* 39, 966–974.
- Handy, F., Katz, E. (1998). "The wage differential between nonprofit institutions and corporations: Getting more by paying less". *Journal of Comparative Economics* 26, 246–261.
- Hansmann, H. (1980). "The role of nonprofit enterprise". *Yale Law Journal* 89 (5), 835–898.
- Hansmann, H. (1981). "Nonprofit enterprise in the performing arts". *Bell Journal of Economics* 12 (2), 341–361.
- Hansmann, H. (1987). "The effect of tax exemption and other factors on the market share of nonprofit versus for-profit firms". *National Tax Journal* 40 (1), 71–82.
- Hart, T. (2001). "The philanthropy revolution". *Fund Raising Management* 32 (3), 22–27.
- Heilbrun, J. (1993). "Innovation in art, innovation in technology and the future of the high arts". *Journal of Cultural Economics* 17 (1), 89–98.
- Heilbrun, J. (2001). "Empirical evidence of a decline in repertory diversity among American opera companies 1991/92 to 1997/98". *Journal of Cultural Economics* 25, 63–72.
- Huntington, P. (1993). "Ticket pricing policy and box office revenue". *Journal of Cultural Economics* 17 (1), 71–88.
- Independent Sector (2001). "The new nonprofit almanac IN BRIEF: Facts and figures on the independent sector". Independent Sector, Washington, DC. <http://www.IndependentSector.org>.
- James, E. (1983). "How nonprofits grow: A model". *Journal of Policy Analysis and Management* 2 (3), 350–366.
- Jana, R. (2001). "Funding art for business' sake". <http://www.wired.com/news/culture> Last accessed: March 8, 2002.
- Jeffri, J. (1980). *The Emerging Arts: Management, Survival and Growth*. Praeger Publishers, New York.
- Keren, M. (1972). "Macroeconomics of unbalanced growth: Comment". *American Economic Review* 62 (1), 149.
- Kesenne, S. (1994). "Can a basic income cure Baumol's disease?" *Journal of Cultural Economics* 18, 93–100.
- Kingma, B. (1989). "An accurate measurement of the crowd-out effect, income effect, and price effect for charitable contributions". *Journal of Political Economy* 97, 1197–1207.
- Klaic, D. (1998). "Unlike airline companies and phone companies: Performing arts in Europe". In: Boorsma, P., van Hemel, A., van der Wielen, N. (Eds.), *Privatization and Culture: Experiences in the Arts, Heritage and Cultural Industries in Europe*. Kluwer Publishing, Dordrecht.
- Kolb, B.M. (1997). "Pricing as a key to attracting students to the performing arts". *Journal of Cultural Economics* 21, 139–146.
- Kuan, J. (2001). "The phantom profits of the opera: Nonprofit ownership in the arts as a make-buy decision". *Journal of Law, Economics, and Organisation* 17 (2), 507–520.
- Kushner, R.J., Brooks, A.C. (2000). "The one-man band by the quick lunch stand: Modeling audience response to street performance". *Journal of Cultural Economics* 24, 65–77.
- Kushner, R.J., King, A.E. (1994). "Performing arts as a club good: Evidence from a nonprofit organization". *Journal of Cultural Economics* 18, 15–28.
- Lange, M.D., Luksetich, W.A. (1993). "The cost of producing symphony orchestra services". *Journal of Cultural Economics* 17 (2), 1–15.
- Linder, S. (1970). *The Harried Leisure Class*. Columbia University Press, New York.
- Luksetich, W.A., Lange, M.D. (1995). "A simultaneous model of nonprofit symphony". *Journal of Cultural Economics* 19, 49–68.
- Mackerras, C. (1999). "Tradition and modernity in the performing arts of the Tibetans". *International Journal of Social Economics* 26 (1/2/3), 58–78.

- McCarthy, K., Brooks, A.C., Lowell, J., Zakaras, L. (2001). *The Performing Arts in a New Era*. RAND, Santa Monica, CA.
- Musgrave, R.A. (1959). *The Theory of Public Finance*. McGraw-Hill, New York.
- National Endowment for the Arts (1998). *The Arts in the GDP: Consumers Spent More than \$10 Billion on Admission Receipts for Performing Arts Events in 1997*, 69. Washington, DC.
- Nelson, K. (1983). "The evolution of the financing of ballet companies in the United States". *Journal of Cultural Economics* 7 (1), 43–62.
- Netzer, D. (1992). "Arts and culture". In: Clotfelter, C. (Ed.), *Who Benefits from the Nonprofit Sector?* University of Chicago Press, Chicago, IL, pp. 174–206.
- O'Hagan, J., Purdy, M. (1993). "The theory of non-profit organisations: An application to a performing arts enterprise". *Economic and Social Review* 24 (2), 155–167.
- Peacock, A.T. (1976). "The problems of the performing arts and economic analysis". In: Chen, Y.-P. (Ed.), *Understanding Economics: Essays on Public Policy*. Little, Brown and Company, Boston, MA, pp. 99–110.
- Peacock, A.T. (1996). "The 'Manifest Destiny' of the performing arts". *Journal of Cultural Economics* 20, 215–224.
- Rose-Ackerman, S. (1982). "Charitable giving and 'excessive' fundraising". *Quarterly Journal of Economics* 97 (2), 193–212.
- Rose-Ackerman, S. (1996). "Altruism, nonprofits and economic theory". *Journal of Economic Literature* 34, 701–728.
- Rubinstein, A.J., Baumol, W.J., Baumol, H. (1992). "On the economics of the performing arts in the Soviet Union and the USA". *Journal of Cultural Economics* 16 (2), 1–23.
- Salamon, L.M., et al. (Eds.) (1999). *Global Civil Society*, vol. 1. Johns Hopkins Center for Civil Society Studies, Baltimore, MD.
- Salem, M. (1984). *Organizational Survival in the Performing Arts: The Making of the Seattle Opera*. Praeger, New York.
- Schiff, J., Weisbrod, B. (1991). "Competition between for-profit and non-profit organizations in commercial activities". *Annals of Public and Cooperative Economics* 62, 619–640.
- Schuster, M.J. (1999). "The other side of the subsidized muse: Indirect aid revisited". *Journal of Cultural Economics* 23, 51–70.
- Schwarz, S. (1992). "Economic behavior of the symphony orchestras in the decade of the 1970s". *Journal of Cultural Economics* 15 (2), 95–101.
- Seaman, B.A. (1985). "Price discrimination in the arts". In: Owen, V., Hendon, W. (Eds.), *Managerial Economics for the Arts*. Association for Cultural Economics, Akron, OH, pp. 47–60.
- Shanahan, J.L. (1978). "The consumption of music: Integrating aesthetics and economics". *Journal of Cultural Economics* 2 (2), 13–26.
- Smith, H.D. (1997). "The rest of the nonprofit sector: Grassroots associations as the dark matter ignored in prevailing 'Flat Earth' maps of the sector". *Nonprofit and Voluntary Sector Quarterly* 26 (2), 114–131.
- Steinberg, R. (1986). "The revealed objective functions of nonprofit firms". *Rand Journal of Economics* 17 (4), 508–526.
- Steinberg, R. (1993). "Does government spending crowd out donations? Interpreting the evidence". In: Ben-Ner, A., Gu, B. (Eds.), *The Nonprofit Sector in the Mixed Economy*. The University of Michigan Press, Ann Arbor, MI, pp. 99–125.
- Taalas, M. (1997). "Generalised cost functions for producers of performing arts – allocative inefficiencies and scale economies in theatres". *Journal of Cultural Economics* 21, 335–353.
- Throsby, D. (1990). "Perception of quality in the demand for theatre". *Journal of Cultural Economics* 14 (1), 65–82.
- Throsby, D. (1994). "The production and consumption of the arts: A view of cultural economics". *Journal of Economic Literature* 32, 1–29.
- Throsby, D. (1996). "Economic circumstances of the performing artist: Baumol and Bowen thirty years on". *Journal of Cultural Economics* 20, 225–240.

- Throsby, D., Withers, G.A. (1979). *The Economics of the Performing Arts*. St. Martin's Press, New York.
- Throsby, D., Withers, G.A. (1986). "Strategic bias and demand for public goods: Theory and application to the arts". *Journal of Public Economics* 31, 307–327.
- Touchstone, K.S. (1980). "The effects of contributions on price and attendance in the lively arts". *Journal of Cultural Economics* 4 (1), 33–46.
- US Department of Commerce, Bureau of the Census (1997). *US Census of Service Industries (Various Years)*. US GPO, Washington, DC.
- US Internal Revenue Code Title 26, Subtitle A, Chapter 1, Subchapter F, Part I, Section 501 (c) 3.
- Venture Philanthropy Partners, Venture Philanthropy (2002). *Advancing Nonprofit Performance through High-Engagement Grantmaking*. Venture Philanthropy Partners, Washington, DC.
- Warr, P.G. (1982). "Pareto optimal redistribution and private charity". *Journal of Public Economics* 19, 131–138.
- Weisbrod, B.A. (1978). *The Voluntary Nonprofit Sector: An Economic Analysis*. Lexington Books, Lexington, MA.
- Weisbrod, B.A. (1983). "Nonprofit and proprietary sector behavior: Wage differentials among lawyers". *Journal of Labor Economics* 1, 246–263.
- West, E.G. (1987). "Nonprofit versus profit firms in the performing arts". *Journal of Cultural Economics* 11 (1), 37–48.
- Zimmer, A., Toepler, S. (1999). "The subsidized muse: Government and the arts in Western Europe and the United States". *Journal of Cultural Economics* 23, 33–49.

## CREATIVITY AND THE BEHAVIOR OF ARTISTS\*

WILLIAM D.A. BRYANT and DAVID THROSBY

*Macquarie University, Sydney, Australia*

### Contents

Abstract	508
Keywords	508
1. Introduction	509
2. Definitions of creativity	510
3. Theories and models of creativity	511
4. Creativity in economic behavior	513
5. Modeling the creative process in the arts	514
6. Game theory, optimization and artistic creativity	516
6.1. A game against financiers	516
6.2. A creative optimization decision	517
7. Creativity, talent and artistic output: Some empirical evidence	521
8. Conclusions	527
References	527

\* With the usual *caveat* we express our gratitude to Moshe Adler, Roberto Cellini, Tiziana Cuccia and Victor Ginsburgh for comments on an earlier draft of this chapter.

**Abstract**

Creativity is a basic human trait that comes into play in a variety of contexts, including the production of art. It relates to the capacity of individuals to think inventively and imaginatively and to go beyond traditional ways of solving problems. In this chapter we consider various definitions of creativity and proceed to consider theories and models of creativity that endeavor to characterize both the creative individual and creative modes of thought. Next, we examine some of the ways in which creativity has been brought into economic analysis. We then turn to the central concern of the chapter, namely modeling the creative process in the arts. Our consideration of the issue leads us to propose an approach in which the creative choice by artists is viewed as an optimization decision with respect to 'creative effort', given the artists' perceptions of what 'the market' and what 'the artworld' care about. In the penultimate section we discuss the relationship between creativity and talent in an empirical context, using recent data on artists' attitudes and behavior. The chapter ends with some conclusions and suggestions for further research.

**Keywords**

creativity, talent, artists, innovation

*JEL classification:* Z11

## 1. Introduction

The concept of creativity as a dynamic force in human behavior has long been a subject for study among philosophers, psychologists, sociologists and mathematicians. In more recent times, neurophysiologists have become interested in creativity as their understandings of the complexities of the functioning of the human brain have developed. In another sphere, the revolution in communications technology has stimulated interest in creativity as a means to finding solutions to problems in computer science. In economics, however, creativity has been of little concern, beyond some attention to its role as a precursor to innovation, following Schumpeter. By and large creativity has not been seen *per se* as a significant variable in economic models of individuals, firms, industries or economies, although the potential role of creativity as a key resource in the so-called new economy is awakening some interest amongst economists at the present time.

In the arts, of course, creativity has always been central to discussions about how art is produced, and countless studies of creativity in music, the visual arts, literature etc. have been undertaken from a variety of disciplinary standpoints. Investigations into the creative processes of the great artists, composers and writers of the past have tried to identify how they found their inspiration and how this was transformed into finished artistic product. A particular purpose of creativity research in the arts carried out by psychologists and educationists has been to determine how best to frame educational strategies to unlock the creative potential of children and young people.

In this chapter we examine the notion of creativity in the context of the economics of the arts and culture. We are motivated particularly by the following question: are there parallel processes where the economic and cultural worth of artistic goods and services are generated separately, or is the process of value creation one that could be understood only as a unitary phenomenon? Partly because of its elusiveness, but mostly because of its potential for profound and unpredictable impacts on modes of production and sources of value, the phenomenon of creativity and the behavior of artists is a matter that deserves to be given serious attention by economists.

We begin our account with definitional questions, demonstrating perhaps the validity of Shea's (1990, p. xiii) proposition that "Only a very rash person would attempt to define creativity in either the arts or science". We proceed to consider theories and models of creativity that endeavor to characterize both the creative individual and creative modes of thought. Next, we examine some of the ways in which creativity has been brought into economic analysis. We then turn to the central concern of the chapter, namely modeling the creative process in the arts. Our consideration of the issue leads us to propose an approach in which the creative choice by artists is viewed as an optimization decision with respect to 'creative effort', given the artists' perceptions of what 'the market' and what 'the artworld' care about. In the penultimate section we discuss the relationship between creativity and talent in an empirical context, using recent data on artists' attitudes and behavior. The chapter ends with some conclusions and suggestions for further research.



## 2. Definitions of creativity

Standard definitions of creativity that can be found in any dictionary emphasize the ability to produce new products or new solutions to problems, where the individual involved uses inventive means and shows imagination as well as routine skill. In general usage, creativity is seen as a capacity of individuals to go beyond traditional ways of thinking, doing, knowing and making. Thus the end result of creativity is something new and original. In this vein, an intriguing and concise definition is provided by Sanchez-Capdequi (2000) who describes creativity as a human activity “which generates a non-pre-existent human order”. Creativity can be contrasted with intelligence insofar as it is characterized by heuristic, open-ended or divergent thinking rather than by algorithmic or convergent thought processes. These characteristics of creativity imply an instability or unpredictability, suggesting that whereas intelligence is measurable (at least in principle), creativity is likely to be less amenable to standardized evaluative tests.

Apart from relying on traditional disciplinary definitions of creativity, it is useful to ask creative people themselves how they would define it. For example, Glück, Ernst and Unger (2002) surveyed artists from different domains of the pictorial arts and elicited from them their definitions of creativity. A control group of forty-seven psychology students were also asked the same question. Both groups of respondents were asked to rate the salience of product and personal attributes in their definitions of creativity. The responses showed that definitions and salience ratings differed systematically between ‘free’ artists such as painters and sculptors, artists in more constrained professions such as architects and designers, and the students of psychology. The only commonality among the groups was that it was agreed that ‘a creative person should have many ideas’. Also, the psychology students tended to emphasize positive feelings evoked by creative activities while both groups of artists often referred to creativity as ‘hard work’. Indeed the latter characteristic reflects a particular difficulty in pinning down a precise definition of creativity, i.e. the fact that it is unclear to what extent creativity comes from random processes of inspiration and to what extent it results simply from determined and purposeful thinking.<sup>1</sup>

Overall, it seems that the prospects for finding a clear-cut definition of creativity are dim. Hayes and Stratton (2003, p. 70) sum up the present state of the field from a psychological viewpoint as being one in which “we have no plausible theory of how creativity happens, no reliable way of measuring the creativity of a person, and no real idea of whether creativity happens because of characteristics of the individual, or because of particular kinds of circumstances”. Despite these problems, we turn in the next

<sup>1</sup> The poet A.E. Housman is one artist who saw his own creativity as arising from *both* sources, describing his composition of a particular poem thus: “Two of the stanzas . . . came into my head . . . while I was crossing the corner of Hampstead Heath between the Spaniard’s Inn and the footpath to Temple Fortune. A third stanza came with a little coaxing after tea. One more was needed, but it did not come: I had to turn to and compose it myself, and that was a laborious business” [Housman (1961, p. 195)].

section to a brief account of some theories and models of creativity that are currently under discussion.

### 3. Theories and models of creativity

Studies of creativity in the behavioral sciences distinguish between creativity as a *trait* of individuals and creativity as a *process* by which problems are solved or new ideas generated. In the former context, a further distinction is made between those aspects of an individual's personality and character that are intrinsic, arising, for example, as a result of genetic inheritance, and those that are acquired through the influence of external circumstances, such as education and experience.

Looking more closely at the concept of the creative individual, we can observe three characteristics that have been extensively discussed. First, there is family background, which has the capacity to influence creative development both positively and negatively. In particular there appears to be a strong tendency in people of exceptional creative ability to have come from family backgrounds in which parents and other role models have provided a climate conducive to the fullest development of creative talent [Ochse (1991); Albert (1994); Runco (1999a)]. Nevertheless there are contrary examples where other factors have had to assert themselves in order to overcome unfavorable or discouraging family circumstances.

Important amongst such other factors is the second characteristic or group of characteristics of the creative individual, namely personality traits. A large number of studies have looked at the sorts of personal qualities that facilitate creative thinking and creative behavior. Generally these qualities have been found to be distinct from high general intelligence, and more important than IQ as predictors of creative achievement. They include such traits as flexibility, freedom of thought, independent mindedness, a willingness to take risks and a tolerance of ambiguity (i.e. an ability to remain open-minded in the face of uncertainty). Other traits such as perseverance and courage have also been discussed [Dacey and Lennon (1998, Chapter 5)].

The third characteristic of creative individuals is the acquisition of cognitive skills, such as the capacity for divergent thinking, i.e. thinking 'beyond the square', where the exercise of the imagination leads to multiple possible solutions to a problem. This has proved to be a fruitful area for empirical research, and tests for divergent thinking have been devised which have been widely used in creativity research and in practical applications, for example, as indicators of potential for creative thought in areas such as management [Runco (1999b)].

Turning to creativity as a process, we note that the representation of creativity in this context has generally dissected the process into a series of stages [Gilhooly (1988); Lubart (2001)]. Typically the stages as described involve preparation or observation, followed by definition of the problem, perhaps a stage of 'incubation', and then illumination leading to a solution. A final stage may also be identified involving verification. Whilst the characterization of the creative process as one of rational decision-making

according to a logical procedure has much appeal, it seems to imply that anyone could become creative simply by following the required steps. In fact, it may well be that creativity, especially in art, is the very antithesis of a rational process. Indeed revolutionary movements in art such as dada have been explicitly aimed at subversion of the established order.

Alternatively, the staged model of rational choice could be replaced by one that recognizes creativity as a stochastic process. For example, [Simonton \(2004a, p. 84\)](#) argues that

to claim that creativity is stochastic is to assert that it entails much more uncertainty and unpredictability than would be expected from a forthright, rational process. At the same time, to hold that creativity is stochastic is not tantamount to the assertion that it is totally random, and therefore capricious and illogical. On the contrary, . . . creativity has the characteristics of *constrained* stochastic behavior. Creativity is to a certain degree predictable, but far from deterministic.

Finally, we turn to creativity specifically in the arts. During the 17th and 18th centuries there was much speculation about the nature of genius – the indefinable quality that touched certain individuals and turned them into great artists.<sup>2</sup> More recently, the existence of an absolute standard of aesthetic judgment in defining genius has been challenged. Furthermore, it is argued that the social and political context within which artists work has a profound effect on the exercise of their creativity. Indeed [Simonton \(2004b, pp. 3–12\)](#) identifies both genius and the social context amongst four explanations of creative behavior in science, the other two being logic (simply following an orderly experimental path) and chance (serendipity or the chance juxtaposition of events producing unexpected outcomes). Somewhat similar propositions could be made about the origins of creativity in the arts.<sup>3</sup>

Despite all the difficulties, efforts have continued towards understanding the workings of the mind in the making of art, with attempts to model the cognitive, affective, behavioral and contextual factors associated with the process of artistic creation. One particular line of investigation has been to analyze the evidence from artists themselves in describing how they work [[Mace \(1997\)](#); [Mace and Ward \(2002\)](#)]. Along these lines, a number of researchers have scoured the biographical records and writings of great artists of the past in an effort to identify how artists of outstanding creativity have found their inspiration [[Gedo and Gedo \(1992\)](#)]. We return to the question of modeling the creative process in the arts in later sections of this chapter.

<sup>2</sup> See [Etlin \(1996, p. 38ff\)](#); from a psychological viewpoint [Weisberg \(1993\)](#) argues that the idea of genius as deriving from some unfathomable or mystical source is unfounded, but rather that creative thinking is simply an extension of normal mental functions.

<sup>3</sup> For another comparison between creativity in art and in science, see [Rothenberg \(1979\)](#).

#### 4. Creativity in economic behavior

Although this chapter deals primarily with creativity in the economics of art and culture, it is appropriate as a prelude to our discussion of the behavior of artists to look briefly at creativity as it occurs in economics more generally, notwithstanding the fact that, as we remarked earlier, the concept of creativity has not stirred a great deal of interest amongst economists. Three approaches to this issue can be identified.

First, the phenomenon of the creative individual can be interpreted in economic terms by reference to the theory of human capital. Although [Becker \(1964\)](#) does not mention creativity in his seminal book on this subject, it is not hard to see that, in the portfolio of human characteristics that go to make up a given individual's human capital, creativity could be seen as one element.<sup>4</sup> Even if creativity as such cannot be separately identified, at least those attributes that appear to contribute to a person's creative ability – flexibility, personality traits, cognitive skills, etc. as noted above – can be interpreted as components of the human capital resource. Carrying the proposition further opens up the possibility in a given individual of improving the creative productivity of one or more of these components through investment, via the usual avenues of education, training, on-the-job experience, etc. Following this line, [Sternberg and Lubart \(1991\)](#) suggest that creative performance results from a confluence of six different individual characteristics – intellectual processes, knowledge, intellectual style, personality, motivation and the person's environmental context; investment in improving any of these could improve creativity. However, the relationships involved are not necessarily simple or linear; for instance, it is likely that creativity increases with knowledge at a decreasing rate, and that the marginal product will turn negative if, as seems plausible, “too much” knowledge actually inhibits creativity [[Lubart and Runco \(1999, p. 624\)](#)]. Moreover it may be that traditional education, including arts training, may not be the most effective human capital investment for unlocking creativity. History is replete with examples of brilliant artists who did poorly at school; D.H. Lawrence, for instance, was ranked only 13th out of 21 in his secondary school composition class, yet became a major novelist and poet [[Simonton \(1999, p. 191\)](#)]. Many artists regard experience as a better teacher.

Second, creativity in economics is implicated in the analysis of innovation, market structures and technological change, especially following the seminal contributions to thinking in this area made by Joseph Schumpeter, who challenged the proposition that competitive markets provide the strongest incentive towards innovation [[Schumpeter \(1943, Chapter 8\)](#)]. He argued on the contrary that large firms and firms with significant market power would be more likely than small ones to encourage innovative effort and that this would lead to more rapid technological progress than would occur under other structural conditions. The monopoly-competition-innovation issue has been discussed at length by a number of subsequent writers, notably [Arrow \(1962\)](#) and [Demsetz \(1969\)](#), both of whom examined the supposed underinvestment in inventive effort in a free-market competitive economy. All of this work on innovation as an element in industrial

<sup>4</sup> See also [Chapter 24](#) by Towse in this volume.

organization and as an object of public policy has made little or no reference to creativity as such, although there has been some attention paid to the invention process and the sources of scientific ideas in studies of the role of R&D in the economy [MacKinnon (1962); Hope (1988)].

There has also been some allusion to the nature of creative thought *per se* in the third line of development to which we can refer, namely the role of creativity within organizations. Indeed this is an area that has developed significantly in recent years. It extends a long way beyond simply the invention of new products to gain a competitive advantage for the firm. Certainly it has a lot to say about this matter, especially in the context of contemporary production technologies, but it also looks at creativity as a more ubiquitous resource within the firm (and other types of organization such as bureaucracies), informing processes of strategic planning, personnel management, marketing, community relations and so on. Factors that contribute to creativity in these respects may arise from personal qualities of managers and employees, or from institutional structures that deliberately encourage creative thought [Heunks (1998)].

In the end it may be that in economic terms creativity both of individuals and of organizations may spring from similar sources and hence may be able to be modeled in similar ways. Thus, for example, McCain (1992, Chapters 18–19) develops a theory of individual creativity based on the cognitive filtering of impulses which he goes on to show can be readily extended to business enterprises. Without wishing to become anthropomorphic, we might likewise extend other theories of the creative individual, enabling us to give some substance to otherwise imprecise concepts such as ‘the creative firm’, ‘the creative city’, etc.

## 5. Modeling the creative process in the arts

Given the centrality of creativity to the act of artistic production, whether by an individual artist or by groups of artists, it is difficult to imagine analyzing production decisions in the arts without at least some acknowledgment of the creative origins of artistic work. Nevertheless, much of the modeling of the production processes of arts firms – theater groups, music ensembles, opera and dance companies, etc. – has proceeded without specific reference to creativity. It should be noted however that decisions about output quality have certainly played a prominent part in such models and hence the role of creativity is at least implied if not explicitly specified as a variable. Similarly, research into the behavior of individual artists has generally adopted a labor market framework within which to analyze their working patterns, concentrating on estimating labor supply functions, earnings functions, etc. rather than looking for the creative origins of artists’ decision-making.<sup>5</sup> Nevertheless, again it must be acknowledged that in some studies the role of talent as a determinant of success in artistic labor markets has been

<sup>5</sup> See Chapter 22 by Menger and Chapter 23 by Alper and Wassall in this volume.

brought into account, and hence the possibility of a link to the 'pure' concept of creativity is established [Rosen (1981)].

One area of considerable interest in the economics of artists' behavior is the matter of motivation, interpreted in the context of a staged creative process as the stimulus towards taking the first step. Artists may be motivated by a variety of influences and desires, ranging from the pursuit of a grand artistic vision to more mundane concerns for pecuniary gain. Frey (1997; 2003, pp. 141–144) has distinguished between motivations arising from within the artist's being and those imposing themselves on the artist from outside; he identifies these as 'intrinsic' and 'extrinsic' motivations respectively. These motivations might be transformed into a desire on the part of the artist to create value, realized in either an economic or a cultural payoff [Throsby (2001, Chapter 6)]. Thus for the intrinsically motivated artist seeking purely artistic reward, a link between creativity and cultural value is established; for the commercially minded artist the extrinsic profit motive consigns creativity to a lesser role, since more routine skills of market awareness, entrepreneurial flair, etc. are likely to be implicated in the generation of economic value.

A further aspect of creativity in the arts with particular relevance to the question of motivation is the pattern of creative activity over time. Some artists are able to produce a steady stream of output over long periods; for example, Dickens and Ibsen wrote regularly for most of their working lives. Others reach a point in their career at which creativity seems to dry up – an example is the Finnish composer, Jan Sibelius. A particularly intriguing case is the artist who produces work in cycles; the output patterns of Mozart and van Gogh, for example, show clear cyclical phases of intense productivity followed by inactivity. Rinaldi, Cordone and Casagrandi (2000) model the dynamics of production in creative professions by proposing two differential equations to specify the interaction between two state variables, satisfaction (self-esteem felt on the basis of past achievements) and creativity (fluency in identifying new ideas and directions). The flow of new achievements in this model is made a function of creativity, under an assumption of diminishing returns in the output of new work. The interactions between the two state variables are specified such that satisfaction increases with the flow of achievements, and creativity is stimulated by the ups and downs of satisfaction. Rinaldi et al. show that regular output over time is likely to be produced by artists who quickly forget their past achievements and are not affected by variations in their levels of satisfaction. By contrast oscillatory behavior will be demonstrated by individuals who do not forget their past too quickly, and who are sensitive to variations in satisfaction.

The Rinaldi et al. model relies on intrinsic factors as conditioning the creative process. An alternative approach is to look to external circumstances as affecting the incentives to artistic work. There is no doubt that historical, political and social factors influence creative activity in the arts. Although some artists may be characterized as living and working entirely in an imaginative realm derived only from within themselves, it is more likely that external influences will play some role in initiating or mediating the creative process. Generally it might be thought that the role of such factors – whether they be the prevailing artistic fashion, political or social conditions, or the

demands of the marketplace<sup>6</sup> – will be positive in their influence on creativity. Alternatively, Seitz (2003) argues that political and religious censorship, corporate control and influence, copyright restrictions and other cultural and economic considerations deeply constrain creative self-expression. He suggests that creative product is most likely to emerge when individual creative abilities are exercised within a communal structure with strong associative networks and a shared sense of collective values. An example of such circumstances might be the emergence of ‘genres’ within particular artforms, such as in the music community of Nashville where country musicians live and interact with one another in the production of their music [Negus and Pickering (2004, p. 73)].

Further evidence for the effects of exogenous circumstances on artists’ creativity comes from Simonton’s (1997, pp. 29–53) analysis of the socio-cultural determinants of artistic distinction. Using data for 772 visual artists and sculptors over a 900 year period up to the beginning of the 20th century, he shows that high levels of creative achievement among these artists were positively associated with contacts with associates, challenges from rivals, guidance from an eminent predecessor (a “paragon”) and relationships with successors (apprentices, admirers, etc.).

## 6. Game theory, optimization and artistic creativity

### 6.1. *A game against financiers*

As John Nash saw it, the aim of game theory is to provide an understanding of the ‘underlying dynamics’ of human behavior, particularly in a social context. It would be surprising, therefore, if game theory could not provide a fruitful means for analyzing the creative process in the arts. Yet so far little attention has been paid to this possibility. One notable exception is the model developed by Cellini and Cuccia (2003) to represent a two-person game between an ‘artist’ and a ‘financier’, i.e. a profit-seeking individual who has to decide whether or not to finance the artist’s work. The artist’s strategies are to experiment, producing creative or innovative output, or not to experiment, producing conservative commercial work. The artist population contains an unknown proportion of innovative artists. Solving the model as a one-shot game shows that a certain number of conservative artists in the population is necessary (and sufficient in the case of a risk-neutral financier) for private financing of the arts.

If the model is analyzed as a repeated game, the opportunity arises for both players to gain information on critical parameters as the game proceeds. Cellini and Cuccia solve the model as a two-stage process, identifying the options open to both sides at each stage. The higher the probability of finding innovative artists in the second stage

<sup>6</sup> Cowen (1998), for example, argues that market incentives lead directly to the production of great works of art; since such works undoubtedly require creativity in their making, the link between financial incentives and the exercise of creativity is established, if Cowen’s argument is correct.



of the game, the lower the probability that the financier will finance an art project at this stage, and the more likely it will be that an artist, in anticipation of this outcome, will choose to experiment in the first stage. For our purposes, the interest of this model lies particularly in the way it depicts the exercise of creativity as a choice the artist makes, conditioned by the expectation of receiving funding. One of the results of this analysis is to show that the probability of creative experimentation is a continuous, non-monotonic and concave function of  $q$ , the *ex ante* probability of a financier facing an innovative artist drawn from the population; the probability of experimentation increases for low values of  $q$  and then declines. The authors interpret this result as indicating that “the presence of conservative artists is a necessary condition to guarantee the possibility of a private financing of artists who choose to experiment” (p. 32).

What this analysis reveals is that the act of being ‘truly creative’ (i.e. ‘experimental’ to use the language of the above analysis) can be understood as a strategic response to the environment in which an artist finds himself or herself, and is not something that needs to be perpetually shrouded in mystery. To carry this point further, we suggest in the following paragraphs an alternative model of the artist’s creative choice.

## 6.2. A creative optimization decision

Consider the following artistic decision,<sup>7</sup> in which a creative artist is thought of as choosing the amount of ‘creative effort’ to invest in a particular work such as a painting, a novel or a musical composition. Suppose the artist’s decision is conditioned by her beliefs about the likely reaction to her work in the domains where the work is received. The domains in which artistic work is received are assumed to be comprised of two parts: ‘the market’ and ‘the world of ideas’ (here referred to simply as ‘the artworld’).

Let  $A$  be a non-empty set of artists and let  $i \in A$  be a typical artist. Assume that there is a trait  $T$  called ‘creativity’ which is exogenously given in amount  $T_i$  to artist  $i$  and in general  $T_i \neq T_j$  if  $i$  and  $j$  are different artists. For each work of art  $w_i$  produced by  $i$  there is a production function. One of the inputs of that production function is the proportion of his or her total creative talent that  $i$  chooses to use in the production of  $w_i$ . Let  $\alpha_i$  denote the proportion of  $T_i$  that  $i$  chooses to use in the production of  $w_i$ . Clearly  $0 \leq \alpha_i \leq 1$ , and the amount of  $i$ ’s productive talent that is actually used in the production of  $w_i$  is  $\alpha_i T_i$ . So  $\alpha_i$  is the basic choice variable for artist  $i$  in the decision that she has to make. If  $i$  chooses  $\alpha_i = 1$  (‘full and free creativity’),  $w_i$  is called a ‘creative work’. If  $i$  chooses  $\alpha_i = 0$ ,  $w_i$  is called a ‘commercial work’. Works for which  $0 < \alpha_i < 1$  contain varying degrees of creative and commercial content, according to the size of  $\alpha_i$ .

<sup>7</sup> The distinction between a *game* and a *decision* may be captured as follows: “When a person . . . decides how to act in dealings with other people . . . there must be some cross-effect of their actions; what one does must affect the outcome for the other . . . For the interaction to become a strategic game, however, we need something more, namely the participants’ mutual awareness of this cross-effect” [Dixit and Skeath (1999, p. 16)].



Suppose that a work of art  $w_i$  can have ‘economic value’  $v_e$  usually conferred by ‘the market’, and ‘cultural value’  $v_c$ , usually conferred by ‘the artworld’. Then the decision problem faced by artist  $i$  can be summarized as follows:

<i>Action</i>		<i>Goods</i>	<i>Utility</i>
<i>space</i>		<i>space</i>	<i>space</i>
$\{\alpha_i: 0 \leq \alpha_i \leq 1\}$	(Production function) $\longrightarrow$	$w_i$	(World) $\longrightarrow U_i(v_e, v_c)$

Artist  $i$  aims to choose  $\alpha_i$  so as to maximize  $U_i(v_e, v_c)$  given her understanding of how  $w_i$  will be transformed into  $(v_e, v_c)$  by the market M and the artworld W.

ASSUMPTION 1. The market M cares only about the economic value of the work  $v_e$ . The artworld W cares only about the cultural value of the good  $v_c$ . The artist cares about both the economic and cultural value of the work, the relative weights assigned to each depending on her preference pattern between the two types of value (see below).

REMARK. Assumption 1 means that  $U^M(w_i) = v_e$ ,  $U^W(w_i) = v_c$  and  $U_i(w_i) = (1 - \lambda_i)v_e + \lambda_i v_c$ , with  $0 \leq \lambda_i \leq 1$ , where  $\lambda_i$  can be thought of as a ‘creative preference parameter’ of artist  $i$ . Given her preferences (i.e. the weighting  $\lambda_i$ ), artist  $i$  has to choose the type of work to produce. This decision will depend on the beliefs she has about the connection between the choice of  $\alpha_i$  (i.e. the *type* of good that  $w_i$  is), and the payoffs  $v_e$  and  $v_c$ . These beliefs may be many and varied. We suppose here that the beliefs of  $i$  are captured by the following assumption.

ASSUMPTION 2.  $v_e = v_e(\alpha_i)$  and  $v_c = v_c(\alpha_i)$  are both concave in  $\alpha_i$ . Assume also that  $v_e$  is positive for  $\alpha_i = 0$ , increasing for  $0 < \alpha_i < k$  and decreasing for  $1 > \alpha_i > k$ , where  $k$  is a “critical value” of  $\alpha$  (see below). Assume that  $v_c$  is zero for  $\alpha_i = 0$  and increasing in  $\alpha_i$  for any  $0 < \alpha_i < 1$ .

REMARK. Assumption 2 means that a “fully commercial” artwork ( $\alpha_i = 0$ ) has a moderate  $v_e$  and a zero  $v_c$ . A “fully creative” artwork ( $\alpha_i = 1$ ) has zero  $v_e$  and relatively high  $v_c$ . The assumption also holds the idea that  $v_e$  is an increasing function of  $\alpha_i$  provided that  $\alpha_i$  is in the ‘safe’ range  $0 < \alpha_i < k$ , i.e. for  $\alpha_i$  less than some critical value  $k$  (e.g.,  $k = 0.5$ ). In other words *some* creative content will contribute to raising the economic value of the work beyond that of a purely routine work with little or no creative appeal. However, the economic value of the work is then a decreasing function for a choice of  $\alpha_i$  greater than  $k$ , because such values of  $\alpha_i$  mean more ‘adventurousness’ and so less market acceptability. As far as the artworld is concerned, the artist conjectures that  $v_c$  is an increasing function of  $\alpha_i$  over the range  $0 \leq \alpha_i \leq 1$  (i.e. the more adventurousness the better in purely artistic terms). A picture of what Assumption 2 is trying to say is in Figure 1.

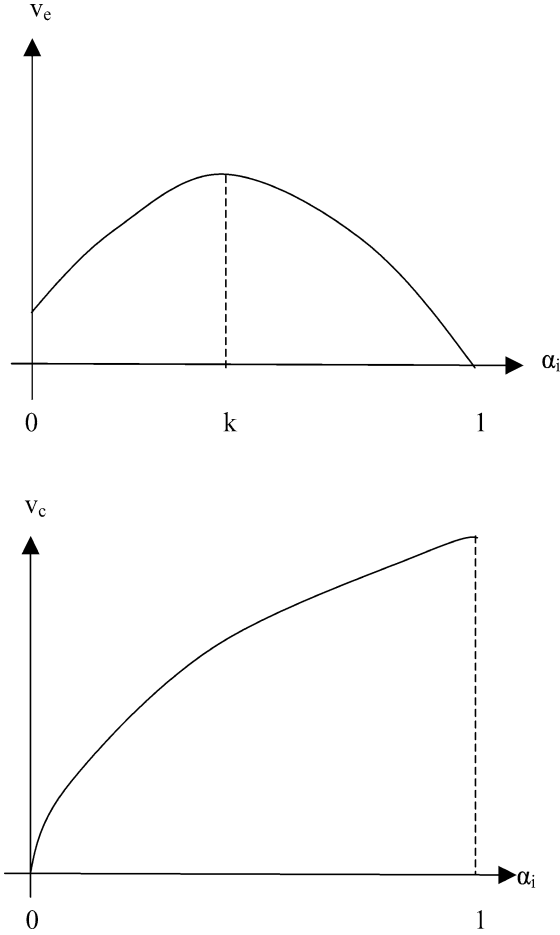


Figure 1.

The artist’s problem is then to choose the proportion of her talent to apply in the production of a work of art, given her preferences for economic and artistic rewards, and her conjecture about the way the market and the artworld will value the work which she produces. Specifically, the artist’s problem is to choose  $\alpha_i$  to maximize utility, i.e. to choose a level of creative effort  $\alpha_i$  that solves the following constrained optimization problem

$$\text{Max}_{\alpha_i} U_i(w(\alpha_i)) = (1 - \lambda_i)v_e(\alpha_i) + \lambda_i v_c(\alpha_i) \tag{1}$$

subject to

$$0 \leq \alpha_i \leq 1. \tag{2}$$

To see what the solution to this problem looks like and how it responds to variations in the artist's preferences, rewrite the problem as

$$\text{Max}_{\alpha_i} (1 - \lambda_i)v_e + \lambda_i v_c \tag{3}$$

subject to

$$v_e, v_c \in V \tag{4}$$

where

$$V = \{v_e, v_c \wedge v_e(\alpha_i) \geq v_e, v_c(\alpha_i) \geq v_c, 0 \leq \alpha_i \leq 1\}.$$

By virtue of **Assumption 2**,  $V$  is a convex set so the problem in Equation (1) is a convex maximization problem with a linear objective function. Since this function is maximized, it is easy to see that at the optimum the first two inequalities in  $V$  will hold as equalities.

A sketch of what is going on here is provided in **Figure 2**. In the context of this figure we can explore the consequences for the choice of creative effort in the following cases (where a star [\*] indicates optimality):

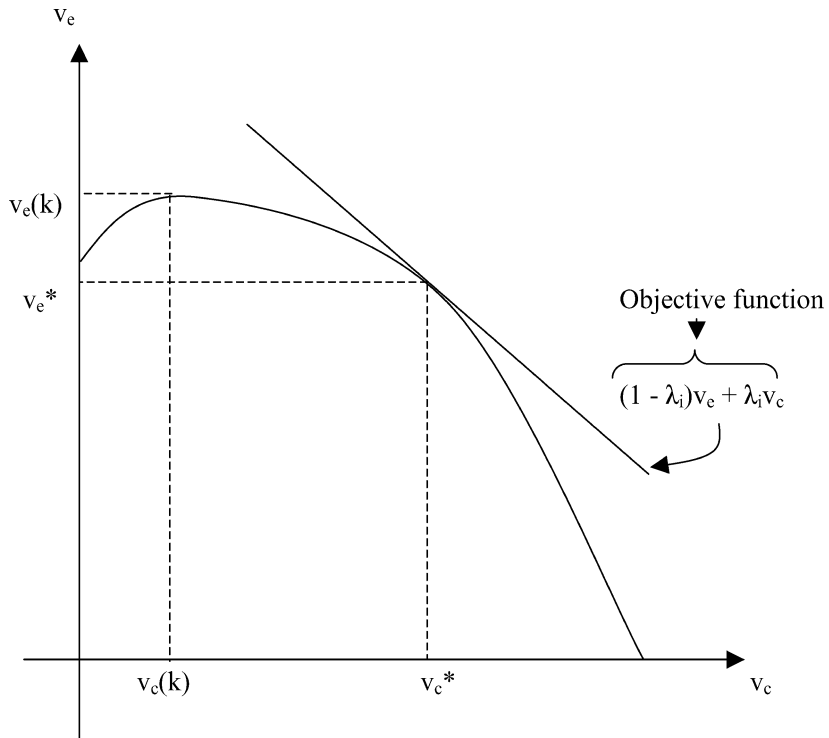


Figure 2.

*Case 1:* If the artist doesn't care at all about money, in other words is uninterested in the evaluation of the market, then  $\lambda_i = 1$  and the objective function in Figure 2 is vertical, with the implied  $\alpha_i^* = 1$ .

*Case 2:* If the artist doesn't care at all about the evaluation of the artworld, but is interested only in making money, we have  $\lambda_i = 0$ ; the objective function becomes horizontal and the implied level of creative effort is then  $\alpha_i^* = k$ .

*Case 3:* If the artist cares about both money and cultural contribution so that  $0 < \lambda_i < 1$ , the implied choice of  $\alpha_i^*$  yields values of  $v_c^*$  and  $v_c^*$  as in the example shown in Figure 2.

Consideration of these three cases and contemplation of the artist's optimization problem leads to the following summary proposition:

**PROPOSITION.** *If an artist imagines that they are in a decision context in which either or both the market and the artworld have a role in the evaluation of their work and if Assumptions 1 and 2 capture the beliefs of the artist, then  $\alpha_i$ , the level of creativity used in the production of an artwork  $w_i$ , is an increasing function of  $\lambda_i$ .*

We note that this proposition is in effect simply a logical consequence deriving from the assumptions that go to make up our model.<sup>8</sup> Nevertheless it is still useful to state it in these terms, given that our purpose in constructing the model and in investigating what it implies has been to illuminate the underlying mechanisms governing the creative choices of the artists.

There are a number of fruitful ways in which this model could be extended. One possibility would be to endogenize the creative preference parameter  $\lambda_i$ . Although it is beyond our present scope to do so here, we can at least suggest one or two comparative static results based on informal speculations about an artist's preference formation. For instance, it might be supposed that an artist who had an unexpectedly gratifying experience in the way her work was received by the artworld would subsequently change her weighting of  $v_c$  (so that  $\lambda_i$  would increase); if so, the model predicts that  $v_c^*$  would shift to the right and  $v_c^*$  would shift down in the new equilibrium. On the other hand commercial success could have differential effects on preferences depending on the initial starting position. These and other conjectures could be translated into statements about how  $\lambda_i$  changes as a result of feedback effects, and the consequent impacts on creative choices in equilibrium could be worked out.

## 7. Creativity, talent and artistic output: Some empirical evidence

We turn now to the question of incorporating creativity into empirical models of artist behavior. In doing so we note that the term 'talent' is more often used than the word

<sup>8</sup> As Suppes (1987) points out, any proposition derived in a formal axiomatic system is necessarily an expression of the axioms or assumptions that comprise the system.

'creativity' in describing an individual's capability of achieving artistic success. Is there a difference between talent and creativity? At one level a differentiation could be made between the possession of a capacity for creative work (talent) and how much of that talent is actually expended on a given work, i.e. how truly creative the final work actually is. Thus, for example, a highly talented pianist could on one occasion perform Rachmaninov's Third Piano Concerto in a technically competent way without playing any wrong notes, whereas on another occasion might offer an entirely new and creative interpretation of this work. This was how creativity was modeled in our previous theoretical section.

We also referred above to speculation about whether creativity or talent are inherited or acquired characteristics, observing that if creativity is inherited it is likely to be enhanced in an individual by a supportive family background. If it is acquired, the process of acquisition could occur in many different ways, including via an artist's education and training.<sup>9</sup> However it arises and however it might be measured, we have noted that artists are likely to possess differing levels of creativity, ranging from highly creative individuals to routine practitioners. Our model proposes that, whatever the level of creative talent an artist possesses, she faces a choice: whether to use it to produce innovative work requiring a high creative input or commercial work requiring a lower creative input.

Can these theoretical propositions be put to the empirical test? To do so requires a means of measuring creativity and a basis for distinguishing creative from commercial output. To examine these questions, we take cross-section data from a recent sample survey of Australian professional artists [Throsby and Hollister (2003)]. These data can be used to identify creativity amongst artists in different artforms by making certain assumptions as to the origins and manifestations of creative ability in the individual practitioner. The data also allow us to propose a distinction between the creative and commercial output of respondents to this survey. In the following paragraphs we construct a *creativity index* for a group of 'initial creative' artists (writers, visual artists, craftspeople, and composers),<sup>10</sup> and use the index to evaluate the importance of creativity in their production of creative and commercial artistic output.

We turn first to the derivation of the creativity index. Assessment of a person's creativity may come from their own opinions about their talent (self-assessment) or from the opinions of observers of their behavior or of their work (external assessment). The survey data do not provide any direct assessment of creativity, but they do contain some information about respondents that can be taken as indicators of or proxies for creativity which can then be aggregated into a combined index. In the survey, participating artists were asked to identify factors that they believed had contributed to their professional development as an artist. They could indicate multiple factors as having had an effect

<sup>9</sup> The growth in 'creative writing' courses in recent years would suggest that in this artform creativity can be acquired, or at least enhanced, through training.

<sup>10</sup> As distinct from performing artists (actors, dancers, musicians, etc.)

and were then asked to indicate which one of the nominated factors they regarded as having been *most* important. A range of 14 possible influences was suggested, including the artist's training, a 'lucky break', financial assistance at a critical time, etc. One of the alternatives was 'my talent' and another was 'support from family and friends'. It might be presumed that those indicating their talent as having been important would see themselves as being talented or creative, on a self-evaluative basis. Furthermore, since as discussed above it is known that creative individuals are more likely to arise from supportive family circumstances, those artists in the survey who indicated this factor as having been important might also be seen as more creative.

In addition to these self-assessments, there is also an external assessment of the artists' work in the survey data. Artists were asked about their receipt of a grant from a Federal or State government arts funding agency. These grants are subject to strict peer-assessment procedures. Hence award of such a grant might be seen as an external indication of the artist's creative achievement or potential, and might be thought of, in terms of the theoretical model in the previous section, as 'the evaluation of the artworld'.

Putting these characteristics together enables us to assemble five separate indicators that might be thought of as contributing towards an individual's creativity. They are:

- Talent considered a factor in advancing own artistic development;
- Talent considered the most important factor;
- Family support considered a factor in advancing own artistic development;
- Family support considered the most important factor;
- Peer-assessed grant received.

If for a given artist a zero-one value is attached to each factor, an unweighted sum of all five factors can be obtained, which could be called a *creativity index* for that artist<sup>11</sup> ranging from zero to 4. Of course calculating the index in this way implies a cardinal scale for measuring creativity, where each factor contributes equally in numerical terms to the aggregate assessment. While these are strong assumptions, they are not entirely implausible in the limited context in which we are applying them, especially if the index is interpreted not as a measure of intensity of creativity but simply as a count of relevant contributing factors when a higher score is a *prima facie* indication of greater creative potential.

Table 1 shows the proportions of artists in this survey sample with a positive score on each criterion, and Table 2 gives the frequency distribution of the creativity index across the whole sample. It can be seen from Table 2 that there is indeed a wide spread of artists according to the creativity measure we have derived. At one extreme these results suggest that almost 15 percent of artists have relatively low creative potential – these artists do not have confidence in their own talent, they do not appear to have come from supportive family backgrounds, and the artworld has not recognized their creative

<sup>11</sup> Note the maximum possible score is 4, not 5; only one (or none) of the second and fourth characteristics listed above could hold for any individual.

Table 1  
Proportion of artists<sup>a</sup> with various creativity attributes: Australia, 2002 (percent)

	Writers (%)	Visual artists (%)	Crafts- people (%)	Composers (%)	Total (%)
Talent: one factor advancing my professional development	74.9	69.9	71.5	92.4	74.2
Talent: most important factor advancing development	33.8	25.1	22.0	42.4	29.2
Family support: one factor advancing my professional development	48.4	48.1	53.7	59.1	50.4
Family support: most important factor advancing development	5.9	7.9	14.6	1.5	7.9
Received peer-assessed grant in last five years	13.7	23.9	18.7	33.3	20.4
<i>n</i>	219	239	123	66	647

<sup>a</sup>Writers, visual artists, craftspeople and composers.

Table 2  
Frequency distribution of creativity index for artists<sup>a</sup>: Australia, 2002

Creativity index <sup>b</sup>	Frequencies	
	no.	%
0	91	14.1
1	150	23.2
2	219	33.8
3	158	24.4
4	29	4.5
Total	647	100.00

<sup>a</sup>Writers, visual artists, craftspeople and composers.

<sup>b</sup>See text for method.

work by award of a peer-assessed grant. At the other end of the spectrum, somewhat less than 5 percent of artists achieve the maximum score.<sup>12</sup>

Using this creativity index we can proceed now to a simple indirect test of the model proposed in the previous section. Following [Throsby \(2006\)](#) we can suggest that the

<sup>12</sup> These numbers would seem to confirm the oft-expressed belief held by skeptical observers of the contemporary arts that there are more bad artists than good artists in the world.

output of artists can be represented by a production function in which labor time, working capital and human capital are the inputs. In this specific context the vector of human capital characteristics includes creativity,<sup>13</sup> in addition to the usual variables of education, training and on-the-job experience. Furthermore, the artist's total output can be divided into "creative" and "commercial" output; the former comprises the novels, the poetry, the paintings and sculptures, the original craft works, the musical compositions etc. that the artist produces, while the commercial output includes art-related activities such as teaching, arts administration etc., where production uses some artistic skills but at a more routine and commercially-oriented level.

Formally, we can define for a given time period for the  $j$ th artist ( $j = 1, \dots, n$ ):

$$y_j^{\text{cr}} = f_1(L_j^{\text{cr}}, PK_j^{\text{cr}}, HK_{ij}), \quad (5a)$$

$$y_j^{\text{co}} = f_2(L_j^{\text{co}}, PK_j^{\text{co}}, HK_{ij}), \quad (5b)$$

where

$y$  = quantity of output;

$L$  = labor input;

$PK$  = input of physical capital;

$HK_i$  = vector of human capital characteristics ( $i = 1, \dots, m$ )

and where the superscripts cr and co denote creative and commercial artistic production, respectively.

In the empirical estimation of Equations (5a) and (5b), we measure creative and commercial output in value terms as the gross revenue from the sale of goods and services of a creative or commercial (arts-related) nature respectively in the survey year (\$A'000). The labor inputs are the average hours per week worked by the artist at creative and commercial artwork, respectively. Physical capital input is measured as art-related expenses incurred in the year of the survey; note that this figure was available only for creative output, so the expenses variable is omitted from the commercial output equation. The components of  $HK_i$  ( $i = 1, \dots, 4$ ) are the artist's years of arts training; age (as a proxy for experience); level of general education (measured as a dummy variable where post-secondary education completed = 1, zero otherwise); and the creativity index as derived above.

Equations (5a) and (5b) were first estimated by OLS, but because of non-normality in residuals and the presence of heteroscedasticity they were re-estimated by generalized least squares, a satisfactory procedure given that we are interested in the relative size and significance of coefficients rather than in using the fitted equations for prediction. Table 3 gives the results. We note in passing that labor and operating capital are significant inputs in producing creative output, and that labor and to a lesser (non-significant) extent arts training have a positive effect on commercial output. However, our interest

<sup>13</sup> Note that the creativity measure used here differs slightly from that used in Throsby (2006).



Table 3  
Determinants of the creative and commercial output of Australian artists,  
2000–2001

Explanatory variable	Creative output <sup>a</sup>	Commercial output <sup>b</sup>
Constant	−0.616 (−0.53)	7.669** (5.08)
Labor time <sup>c</sup>	0.311** (3.81)	0.682** (7.56)
Operating capital <sup>c</sup>	0.814** (14.73)	–
Human capital 1: general education	−0.097 (−0.71)	0.136 (0.68)
Human capital 2: arts training <sup>c</sup>	−0.326** (−3.79)	0.194 (1.49)
Human capital 3: experience <sup>c</sup>	0.400 (1.49)	−0.225 (−0.60)
Human capital 4: creative talent	0.196** (3.06)	0.063 (0.73)
<i>n</i>	424	192
<i>F</i>	57.058	13.155
Adjusted <i>R</i> <sup>2</sup>	0.443	0.241

Artists included are writers, visual artists, craftspeople and composers; all logs are natural logs; t-statistics shown in parentheses; coefficients significantly different from zero at 1 percent (\*\*) level; both equations estimated by generalized least squares.

<sup>a</sup>Dependent variable is log of value of creative artwork produced.

<sup>b</sup>Dependent variable is log of value of commercial (art-related) work produced.

<sup>c</sup>Variable measured in logs.

here is in creativity. The creativity variable shows up as a positive and very significant influence on creative output, and is indeed the only one of the four human capital variables specified in the model to have such a positive effect.<sup>14</sup> On the other hand, creativity has a negligible and insignificant effect on the output of commercial artistic work. These results are consistent with the primary hypothesis put forward above.

Although this analysis is very crude, it does at least suggest, as we remarked earlier, that creativity in the arts, for all its elusive and protean quality, can be systematically studied and that results can be obtained which provide plausible insights into the contribution that creative talent makes to the production of artistic output.

<sup>14</sup> The negative influence of arts training on creative output is difficult to explain, even though it is more than offset in quantitative terms by the positive (though non-significant) influence of experience.

## 8. Conclusions

Creativity remains an elusive phenomenon. This chapter has pointed to some of the ways in which social scientists have attempted to pin down that elusiveness and propose testable theories about how creativity arises and how it asserts itself in decision-making. We have noted that creativity as such has not received much attention in economics; even economists working in fields such as R & D and technological change, where creativity would seem to be especially important, have tended to regard it as a 'black box' whose contents are best left undisturbed. Yet in the economics of arts and culture, the prominence of creativity as an element in artistic production would seem to suggest that ignoring it is likely to compromise the explanatory power of models purporting to represent artistic behavior.

We have suggested in this chapter that some progress can be made, both theoretically and empirically, if creativity can be brought more directly into analysis of artistic decision-making. Clearly work in this area is still at a formative stage, and many avenues exist for carrying the study of the economics of creativity in the arts forward. In this process a number of opportunities would seem to present themselves for economists to cooperate with creativity specialists in other disciplines, notably in psychology. Creativity is indeed a field which would seem to be an ideal one for cross-disciplinary research in the future.

## References

- Albert, R.S. (1994). "The contribution of early family history to the achievement of eminence". In: Colangelo, N., Assouline, S. (Eds.), *Talent Development*. Ohio Psychology Press, Dayton, pp. 311–360.
- Arrow, K.J. (1962). "Economic welfare and the allocation of resources for invention". In: National Bureau of Economic Research, *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton University Press, Princeton, pp. 609–625.
- Becker, G.S. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. National Bureau of Economic Research, New York.
- Cellini, R., Cuccia, T. (2003). "Incomplete information and experimentation in the arts: A game theory approach". *Economia Politica* 20 (1), 21–34.
- Cowen, T. (1998). *In Praise of Commercial Culture*. Harvard University Press, Cambridge, MA.
- Dacey, J.S., Lennon, K.H. (1998). *Understanding Creativity: The Interplay of Biological, Psychological and Social Factors*. Jossey-Bass, San Francisco.
- Demsetz, H. (1969). "Information and efficiency: Another viewpoint". *Journal of Law and Economics* 12 (1), 1–22.
- Dixit, A., Skeath, S. (1999). *Games of Strategy*. W.W. Norton and Co., New York.
- Etlin, R.A. (1996). *In Defense of Humanism: Value in the Arts and Letters*. Cambridge University Press, Cambridge.
- Frey, B. (1997). *Not just for the Money: An Economic Theory of Personal Motivation*. Edward Elgar, Cheltenham.
- Frey, B. (2003). *Arts and Economics: Analysis and Cultural Policy*, second ed. Springer, Berlin.
- Gedo, J.E., Gedo, M.M. (1992). *Perspectives on Creativity: The Biographical Method*. Ablex Publishing Corp., Norwood, NJ.

- Gilhooly, K.J. (1988). *Thinking: Directed, Undirected and Creative*, second ed. Academic Press, San Diego, CA.
- Glück, J., Ernst, R., Unger, F. (2002). "How creatives define creativity: Definitions reflect different types of creativity". *Creativity Research Journal* 14 (1), 55–67.
- Hayes, N., Stratton, P. (2003). *A Student's Dictionary of Psychology*. Arnold, London.
- Heunks, F.J. (1998). "Innovation, creativity and success". *Small Business Economics* 10 (3), 263–272.
- Hope, E. (1988). "Market structure and innovation". In: Grønhaug, K., Kaufmann, G. (Eds.), *Innovation: A Cross-Disciplinary Perspective*. Norwegian University Press, Oslo, pp. 475–490.
- Housman, A.E. (1961). *Selected Prose*. Cambridge University Press, Cambridge. Edited by Carter, J.
- Lubart, T.I. (2001). "Models of the creative process: Past, present and future". *Creativity Research Journal* 13 (3–4), 295–308.
- Lubart, T.I., Runco, M. (1999). "Economic perspective on creativity". In: Runco, M.A., Pritzker, S.R. (Eds.), *Encyclopedia of Creativity*. Academic Press, San Diego, CA, pp. 623–627.
- Mace, M.-A. (1997). "Towards an understanding of creativity through a qualitative appraisal of contemporary art making". *Creativity Research Journal* 10 (2–3), 265–278.
- Mace, M.-A., Ward, T. (2002). "Modeling the creative process: A grounded theory analysis of creativity in the domain of art making". *Creativity Research Journal* 14 (2), 179–192.
- MacKinnon, D.W. (1962). "Intellect and motive in scientific inventors: Implications for supply". In: National Bureau of Economic Research, *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton University Press, Princeton, pp. 361–378. And comment by Kuhn, T.S., pp. 379–384.
- McCain, R.A. (1992). *A Framework for Cognitive Economics*. Praeger, Westport, CT.
- Negus, K., Pickering, M. (2004). *Creativity, Communication and Cultural Value*. Sage, London.
- Ochse, A. (1991). "Why there were relatively few eminent women creators". *Journal of Creative Behaviour* 25 (4), 334–343.
- Rinaldi, S., Cordone, R., Casagrandi, R. (2000). "Instabilities in creative professions: A minimal model". *Nonlinear Dynamics, Psychology and Life Sciences* 4 (3), 255–273.
- Rosen, S. (1981). "The economics of superstars". *American Economic Review* 71 (5), 845–858.
- Rothenberg, A. (1979). *The Emerging Goddess: The Creative Process in Art, Science, and Other Fields*. University of Chicago Press, Chicago.
- Runco, M.A. (1999a). "A longitudinal study of exceptional giftedness and creativity". *Creativity Research Journal* 12 (2), 161–164.
- Runco, M.A. (1999b). "Divergent thinking". In: Runco, M.A., Pritzker, S.R. (Eds.), *Encyclopedia of Creativity*, vol. I. Academic Press, San Diego, CA, pp. 577–582.
- Sanchez-Capdequi, C. (2000). "Social creativity: The unexplored dimension". *Revista de Filosofia (Venezuela)* 34 (1), 77–88.
- Schumpeter, J.A. (1943). *Capitalism, Socialism and Democracy*. George Allen & Unwin, London.
- Seitz, J.A. (2003). "The political economy of creativity". *Creativity Research Journal* 15 (4), 385–392.
- Shea, W.R. (1990). "Introduction: The mainspring of the arts and science". In: Shea, W.R., Spadafora, A. (Eds.), *Creativity in the Arts and Sciences*. Science History Publications, Canton, MA, pp. xiii–xviii.
- Simonton, D.K. (1997). *Genius and Creativity: Selected Papers*. Ablex Publishing, Greenwich, CT.
- Simonton, D.K. (1999). *Origins of Genius: Darwinian Perspectives on Creativity*. Oxford University Press, New York.
- Simonton, D.K. (2004a). "Creativity as a constrained stochastic process". In: Sternberg, R.J., et al. (Eds.), *Creativity: From Potential to Realization*. American Psychological Association, Washington, DC, pp. 83–101.
- Simonton, D.K. (2004b). *Creativity in Science: Chance, Logic, Genius and Zeitgeist*. Cambridge University Press, Cambridge.
- Sternberg, R.J., Lubart, T.I. (1991). "An investment theory of creativity and its development". *Human Development* 34 (1), 1–31.
- Suppes, P. (1987). "Axiomatic theories". In: Eatwell, J., et al. (Eds.), *The New Palgrave Dictionary of Economics*, vol. 1. Macmillan, London, pp. 163–165.

- Throsby, D. (2001). *Economics and Culture*. Cambridge University Press, Cambridge.
- Throsby, D. (2006). "An artistic production function: Theory and an application to Australian visual artists". *Journal of Cultural Economics* 30 (1), 1–14.
- Throsby, D., Hollister, V. (2003). *Don't Give up Your Day Job: An Economic Study of Australian Professional Artists*. Australia Council, Sydney.
- Weisberg, R.W. (1993). *Creativity: Beyond the Myth of Genius*. W.H. Freeman, New York.

## ORGANIZATION OF ARTS AND ENTERTAINMENT INDUSTRIES

RICHARD E. CAVES

*Harvard University, USA*

### Contents

Abstract	534
Keywords	534
1. Introduction: Organization and contracts in creative industries	535
2. Simple creative goods	537
2.1. Supply and returns on investment	538
2.2. Organizing supply of creative goods	539
2.3. Prevalence of option contracts	541
2.4. Agents and matchmakers	543
2.5. Agency and intermediation	545
2.6. Internal organization of gatekeeping firms	547
3. Structures and contracts in complex creative industries	549
3.1. Integration and disintegration	549
3.2. Contracts for complex creative goods	552
3.3. Vertically differentiated talents	553
3.4. Distributing complex creative goods	554
3.5. Structures of creative industries	555
3.6. Multi-activity firms in creative industries	558
4. Role of non-profit enterprise	560
4.1. Non-profits and governance in creative industries	560
4.2. Two-part pricing in practice: The donor-supported non-profit organization	561
5. Conclusions	562
References	564

## Abstract

The organization of the creative (arts and entertainment) industries rests on many types of contracts. These contracts govern collaborations between artists and other parties – at arm’s length, or within an enterprise. These contracts’ structures devolve from a few bedrock properties of creative work and creative products. Artists invest in developing their talents, presenting themselves before ‘gatekeepers’ who seek talents that can profitably be developed and marketed. Gatekeepers commonly function as agents for selecting artists and as match-makers between artists and complementary inputs. As an extension of the gatekeeping function, the participants in creative industries take part in a continuous ranking process that sets and revises the ranks of vertically differentiated talents. Real option contracts pervasively govern the sequential steps of developing a creative product. These can leave the artist an autonomous creative agent (pop musicians and record labels) or enclose artists’ talents in an employment relationship (classic Hollywood studios). The transformation of the movie industry to ‘flexible specialization’ illustrates how changing basic conditions can transform the dominant form of organization. The scales of enterprise in the creative industries tend to be driven by the efficient scales with which creative goods are distributed (very large for record labels and movie studios, small for art galleries), and they tend to assort themselves into those focused on the distribution of creative goods (‘promoters’) and those concerned with identifying and nurturing creative talents (‘pickers’). Large enterprises also include the ‘entertainment conglomerates’ which seek synergistic gains that depend theoretically on quite special conditions; foreclosure and its avoidance may be principal motives. Non-profit enterprises dominate a number of arts activities, apparently for two interrelated reasons. These activities incur high fixed but low marginal costs, pressing them to employ two-part prices and club arrangements to ensure fixed costs’ coverage. When product quality is endogenous, however, non-profit status may be necessary for the manager credibly to forewear degrading quality once the fixed payment is in hand. Non-profits supported by donation streams thus enjoy functional advantages.

## Keywords

agglomeration, creative industries, gatekeepers, job-matching, joint ventures, motion picture industry, non-profit organizations, option contracts, publishing industry, recording industry, royalties, theaters, toys and games industry, vertical differentiation, visual arts

*JEL classification:* L11, Z11

## 1. Introduction: Organization and contracts in creative industries

The field of industrial organization is divided into two branches. The more traditional branch is concerned with how the structures of markets and the behavior conditioned by those structures affect their allocative efficiency. The second branch addresses the question of why markets are organized the way they are. This latter pathway proves highly inviting for study of the arts and entertainment industries, because they pose a richer array of questions about the logic of organizations than do most other sectors. Consider the task of explaining an industry's organization. Any unitary transaction can be carried on between independent firms, with competition among buyers and sellers determining the market's price and quantity. Or transactions can be bundled inside of firms, with quantities determined by administrative decisions. Theory identifies the strengths and weaknesses associated with each mode of organization. Empirical investigators commonly assume that the most effective mode of organization prevails by means of Darwinian survival, then test the match between theoretically predicted and empirically observed organizational choices.

This approach to the organization of industry, originated by [Ronald Coase \(1937\)](#) and [Oliver Williamson \(1985\)](#), has lately been much enriched by research on the theory of contracts. The alternative to internalizing decisions within the firm is to govern them by means of arm's-length contracts between independent agents. Williamson emphasized the hazards to which arm's-length contracts are subject as the key to understanding why decisions are internalized within the firm. Many arm's-length contracts, formal and informal, nonetheless persist. Moreover allocative decisions made within the firm do not automatically escape the shortcomings of contracts. Indeed they are governed by incentive contracts that ply various carrots and sticks in order to influence employees' actions. The firm is a "nexus of contracts", and the success of internalization depends on the performance of arm's-length contracts relative not to "administrative decisions" but rather to the efficacy of contracts drawn and implemented within the firm. In Darwinian fashion we expect the whole set of prevailing arrangements (extent of internalization, organization of firms, structures of arm's-length contracts) to reflect the relative efficiency of arm's-length and internal contractual dealings.

The arts and entertainment industries (hereafter "creative industries") provide an attractive site for applying this approach because they employ distinctive types of deals and intra-firm governance arrangements. While the volume of quantitative research on these organizational arrangements is small, a great deal of casual evidence exists, especially for the United States. When casual observation confirms the prevalence of a practice with a clear and apposite theoretical rationale, the news is worth reporting even where the niceties of controlled experiments and statistical inference remain out of reach.<sup>1</sup> Deciding when a theoretical model pertains to an empirical situation is, of

<sup>1</sup> Much of the theoretical analysis in this paper was set forth in [Caves \(2000\)](#), though some points get fuller development here. That study also assembled a good deal of diffuse empirical evidence that will not be cited or repeated here.

course, a tricky step on which judgments may differ, and researchers ought to leave their tracks uncovered. We take what can pretentiously be labeled an axiomatic approach, laying out some properties that seem common to all creative industries, or to some substantial and specific group of them. These properties were established inductively from a broad body of mostly descriptive evidence, but also with the guidance of the structure-conduct-performance paradigm that informs much empirical research on industrial organization. Thus these properties are hypothesized to be the “bedrock” elements of market structure based in tastes and technologies that determine important but endogenous aspects of structure – the organization of contracts and less formal deals, and the number, sizes, and activity sets of firms.

Creative industries combine inputs from various types of artists with other inputs to turn out some creative good or service intended for consumers’ enjoyment. This production process may involve nothing more than the marketing of the artist’s creation (the visual arts, for example), or it may entail substantial further manufacture (book publishing, music recording). Whether simple or complex, this activity is conditioned by two axiomatic properties. The first, *art for art’s sake*, holds that artists’ utility functions commonly contain two features that strongly affect their participation in contractual economic relationships. The first is a taste for undertaking artistic work for its own sake, which depresses the supply price for the artist’s services below the pecuniary compensation expected from the artist’s best alternative (non-creative) job. The second is preferences as to how the artistic task should be executed. The widely accepted nineteenth-century romantic view holds that the artist creates from inner necessity in order to realize some internal vision.<sup>2</sup> While the artist’s low supply price facilitates in an obvious sense her cooperation with other inputs in a production process, the existence of tastes defined over the mode of production complicates the artist’s contractual participation in a complex creative activity. If the exercise of preferences about the creative process were contractible, they could be traded off against pecuniary compensation in a mutually agreed manner. When they cannot be specified and contracted upon, and when the creative urge also refuses to respect time constraints or commitments, the completeness of contracts between artists and other inputs and on the subsequent governance of such contracts is substantially limited. (We refer to non-artistic inputs lacking these tastes as “humdrum”.)

Another property that we impute to products with substantial creative inputs is great uncertainty about buyers’ reservation prices for any creative output; this uncertainty persists until all costs have been incurred and the finished output placed before them. This property is widely recognized in entertainment industries, where large sunk costs give rise to highly uncertain returns, by the slogan *nobody knows anything* [Goldman (1984)]. The force of this property depends on the interaction of its two conditions – the sunkness of costs and the uncertainty of the output’s market value. When net revenue depends heavily on distinguishing good from bad projects *ex ante*, great effort is

<sup>2</sup> Studies of the training of art and music students show that these imperatives are built into the curricula and absorbed into students’ attitudes. See Getzels and Csikszentmihalyi (1976) and Kingsbury (1988).



devoted to forecasting these outcomes by decision-makers who have invested heavily in knowledge about what failed and succeeded in the past. The *nobody knows* proposition is consistent with costly investments in forecasting by decision-makers who understand the wide variance around their point forecasts: a small improvement in the likelihood of distinguishing correctly between good and bad projects is worth a lot. Also pervasive in creative industries is the property of horizontal differentiation, associated with creative goods that seek uniqueness within sets of conventions that make many of them close substitutes for one another. We can call this property *infinite variety*. The closeness of substitution as an axiomatic property pertains to potential varieties of a creative good. Where it pertains to actual varieties depends on the incidence of fixed costs per variety relative to consumers' combined willingness to pay (overall, and for preferred varieties).<sup>3</sup>

Other axiomatic properties pertain to complex creative activities that require inputs from a number of suppliers each with *art-for-art's-sake* preferences. The resistance of artists to contractual commitments specifying their creative work complicates the organization of activities that demand the collaboration of several artists' inputs (along with humdrum inputs) – a *motley crew* of creative inputs. Creative inputs of any given type are differentiated vertically, that is, all agree that one artist's talent excels another's overall, although the better talent may not be worth its higher wage in every project. Artists' rankings are determined empirically in a costly consensus-based evaluation process involving the vertically differentiated artists themselves as well as others who employ, supervise or collaborate with them. This is the *A-list/B-list* property. It is closely related to another: that the ultimately perceived quality of a complex creative good tends to depend on each creative input performing at least up to some threshold level of competence. A handy way to represent this property is by means of a multiplicative production function: the failure or substandard performance of any input renders the project's whole output valueless. This is the *O-rings* property [Kremer (1993)]. Finally, the efficient execution of complex creative activities requires the close temporal coordination of key artistic and humdrum inputs, and this requirement complicates both the initial contracting and subsequent coordination of such projects (the *time flies* property).

## 2. Simple creative goods

Simple creative activities involve a single artist (source of creative input) dealing with one agent or enterprise that combines the artist's input with humdrum inputs and distributes (perhaps through intermediaries) the creative good to consumers. This seemingly simple relationship raises several major issues of organizational choice. First, would-be artists offering their talents to the market appear to be in chronic excess supply, so that the distributor assumes the role of a gatekeeper, selecting some but turning many

<sup>3</sup> On the determinants of the equilibrium number of units of creative goods on the market, see Baker (1991).

others away. Second, creative and humdrum inputs could be combined in several organizational settings. The humdrum entrepreneur might represent the artist or take part in a joint venture to develop and distribute the artist's product (talent), or the entrepreneur might hire the artist and assume decision rights over her creative activities. Third, the physical location of creative activities is subject to agglomerative pulls that depend in turn on how the dealings between the artist and humdrum inputs are organized and governed.

### 2.1. *Supply and returns on investment*<sup>4</sup>

We suppose that the artist seeks representation and employment upon completion of training or apprenticeship. The training process itself takes the would-be artist before a series of gatekeepers. As in other tournaments, the would-be artist competes at first with a random assortment of local aspirants. Success at the first stage brings the candidate into competition with others who have survived a first round. This series of elimination rounds proceeds through elementary and advanced training and continues through apprenticeship and the quest for commercial success. For the would-be artist the pursuit involves a series of investment decisions made under great uncertainty. While positive local certification is more informative than none, its value for predicting success in subsequent rounds is very low. That is partly because the proportion of initial contestants who achieve some ultimate success is tiny, partly because a student/apprentice's ability to benefit from additional training, conditional on the certification already attained, is not accurately predictable.

The apprentice artist's investment in training eventually realizes some rate of return, which we can think of as becoming known when she faces a commercial gatekeeper. The gatekeeper seeks to judge whether suitable humdrum inputs combined with the artist's developed talent will create enough value to cover their opportunity cost. Gatekeepers will on average do no better than covering their opportunity costs if gatekeeping is a competitive trade. What reward will flow to the artist, though, depends not only on the competitiveness of gatekeepers but also on the correlation of their *ex ante* assessments of the proffered talents. The artist's gross return to her talent is learned only after a contract is reached with the gatekeeper (or other partner) and the market's ultimate assessment realized. Because training and apprenticeship costs are sunk, these realized rates of return will fall into three ranges. First, some contenders get the nod from no commercial gatekeeper; their investments are clearly lost (aside from future consumption benefits). Second, the more successful apprentices admitted by the gatekeeper find that the market's willingness to pay for their talents will yield a positive return on their investment after the competitive gatekeeping enterprise has taken its normal profit. Third, the less successful contestants realize returns that cover the gatekeeper's opportunity cost but

<sup>4</sup> For further discussion of issues raised in this section, see Chapter 22 by Menger and Chapter 24 by Towse in this volume.

yield a rent on the artist's talent insufficient to produce a positive return on her sunk investment in training.

Do all would-be artists, taking the successful and the unsuccessful together, earn a normal return on their training investments? No hard data exist, but the enormous ratios of arts students graduated from U.S. colleges to professionals entering successfully into careers in a given year make it seem highly unlikely. How should we interpret this behavior and the apparent misallocation of resources that results? A high level of expected utility from purely consumption benefits of training is one factor that helps to rationalize the pattern. Another interpretation invokes a form of risk-loving behavior embodied in the attitude that a high level of creative success yields untold riches in utility (beyond that due to the cash takings). However, the budding artist is poorly positioned to make a rational decision about expected returns to training. The romantic conception of the artist's calling encourages the student to regard talent as a god-given asset that deserves unstinting dedication and effort. Furthermore, the teacher who faces the task of sustaining the student's motivation through years of arduous training and practice has every incentive to emphasize the glory of artistic greatness once achieved, and none to mention the paltry chances of achieving it [Towse (1993)].

## 2.2. Organizing supply of creative goods

The gatekeeping process rations and allocates the humdrum resources available to supplement the artist's input, but it does not dictate the organization of the process, which can take several forms. Consider for concreteness the visual artist whose work requires humdrum assistance to distribute and promote it. The artist could become an employee in a humdrum enterprise, producing works of art at the manager's direction under a conventional employment contract. The artist could prepare works sprung from her own inspiration, to be sold (off the park fence on Sunday morning?) piecemeal to whatever dealer takes a fancy to them. Finally, the artist could be represented by a single gallery on the basis of an exclusive-dealing arrangement.

The last arrangement clearly prevails for simple creative goods – between visual artist and gallery owner, between pop musician (group) and record label, between soloist in classical music and impresario, between author and publisher. Why is this so? Continue with the visual artist. In the context of the romantic ideal, artist, consumers, and intermediaries and certifiers (teachers, critics) agree that what matters is the artist's ability to create a sustained body of work that cumulates to a lifetime career. This requires that both the artist and the distributor of her work undertake many actions that amount to investments for long-run returns. The artist develops a body of work that will sustain periodic shows in the art gallery. The gallery operator interprets the work and promotes it to collectors, museum curators, critics, and periodicals, and lends works for shows in museums and other galleries, etc. Maximizing the value of this joint venture to develop the artist's career requires each party to undertake these actions to optimal degrees. For this arrangement to beat out other ways to organize distribution of the artist's work, it must be consistent with *art-for-art's-sake* tastes, which it clearly is. Artist and dealer

must be able to govern their relationship through explicit or implicit contracts. Formally, the dice are loaded against anything approaching a complete contract, which would require long-term commitments by both parties to ideal courses of action that are largely unknown in advance (dependent on random opportunities), incapable of formal contract and not practicably monitored by the parties even if contractible.

The arrangements prevailing in practice are simple (often handshake) contracts that divide gross revenues from sale of the artist's work between the two parties after certain costs are allocated to each. This contract falls short of an ideal incentive structure, which would require that each party exert effort up to the point where the last dollar's worth of effort adds just one dollar to the joint benefit. In practice the actual incentive for effort is weaker: each party expends effort only to the point where the last dollar's worth of effort adds another dollar to its share of the joint benefit. Categories of purchased inputs that are made one party's responsibility are likewise underfunded. However, any cost that is one party's responsibility, though reimbursed before revenues are divided, will be undertaken to the optimal degree (an action that maximizes revenue net of these assigned costs also maximizes either partner's fractional share of this revenue). Lacking a specific duration, these contracts run until either party chooses to end them, for example, when the artist's style changes in a way that eludes the dealer's sympathy. Reputation plays an important role in supporting the enforcement of the implicit terms, but contract failures do occur (for example, when the artist sells directly to collectors without compensating the dealer, or the dealer fails to report sales and make the associated payment to the artist). Evidently the contracts in these joint-venture type arrangements work well enough to dominate any fundamentally different organization of the distribution of visual art.<sup>5</sup>

The joint-venture types of contracts in the creative industries bear a relationship to the theoretical literature on incomplete contracts. The alignment appears neither neat nor simple, however, so we shall only point to some promising connections.<sup>6</sup> The first of these lies in the theoretical assumption that both parties to a relationship can observe and agree on the outcome of a transaction (or the quality of an input or state of nature that is occurring), but they cannot convey their understanding to a third party such as a court enforcing a contract. The assumption comports well with *art for art's sake* and other core properties of creative activities. The conditions in question are "observable but not verifiable". Under some circumstances – such as where the parties can observe each others' investments in the joint enterprise or the quality of intermediate inputs supplied – a first-best contract can still be sustained. Under others, only second-best is sustainable. The second assumption is that no asymmetry of information exists between the primary parties to the transaction. They possess the same information about actions

<sup>5</sup> A particularly interesting historical experience with the organization of the visual arts is the transformation of the French market around the time of Impressionism. An era of state certification through official salons gave way to a "dealer-critic" system of private certification and marketing. See White and White (1993), Jensen (1994), and Wijnberg and Gemser (2000).

<sup>6</sup> For background, see Hart (1989).

(investments) that either party has taken and the resulting quality of a product or level of its variable cost. Rather paradoxically this assumption aligns with the *nobody knows* property: the uncertainty around the values each party observes is great, but there is no general reason to expect one's accuracy or bias to differ from the other's. With this set-up, the theory of incomplete contracts may prove able to explain some empirical aspects of joint ventures in the creative industries: why they might contract on one variable (for example, a movie's screenplay and key actors) rather than another (the quality of the resulting film). It may also explain why one party reserves the right to decline to purchase a creative good (for example, the studio that chooses not to release a completed film), an action that both parties recognized as a possible outcome of their contract. Finally, the literature on incomplete contracts gives much attention to the possibility of renegotiation – think of the buyer's refusal to pay the agreed price for a finished good [Hart and Moore (1999)]. The creative industries illustrate the reputation mechanisms that so often seem effective for punishing those who violate understandings even where formal contracts and courts are not involved.

### 2.3. Prevalence of option contracts

The visual art market makes clear that the viability of an organizational structure compatible with *art-for-art's-sake* preferences depends on congenial long-term contracts and mechanisms that make them sustainable with only the lightest degree of formalization. Other arts and entertainment sectors depend on more formal contracts with distinctive recurring features. As in the visual art market, the creative product originates from some talent of the individual artist. It then goes through one or more steps (processes, transactions) before reaching the final consumer. The *nobody knows* property points to the great uncertainty about the ultimate reception of this product, both early and late in this series of fabrication stages. Because the right decision about fabricating a creative product can sometimes realize so much more value than a wrong decision, the participants find it worth investigating omens of successful or failure even if their information content is small. The incentive to make this investment in information and adapt to its message is at its maximum for the party next in line in the fabrication sequence to sink still-fungible resources into the project. Another key property is that inputs incorporated at any stage in the process (whether of creative or humdrum origin) are entirely sunk. With the input sunk, its supplier generally can make no further contribution to the value of the product; she might be asked for a rewrite (or the equivalent) when a partially completed project is judged to have gone off the rails, but that entails an additional contribution of resources and any allotment of decision rights tied to it. In general, the input sequence does not “cycle back”.

To focus incentives efficiently, a contract governing such a processing sequence should allocate decision rights in a way consistent with the parties' opportunities to affect the final product's value. That is exactly the property of the real option contracts in widespread use among creative industries. Consider the deal between an artist (screenwriter with a completed script, for example) and an agent able to supply the next round

of inputs to the process (film producer). The option contract between them has these features:

- The producer gets a period of time (six months perhaps) in which to investigate the possibility of filming the writer's screenplay. This investigation period may be renewable.
- Writer and producer agree on the full terms under which the writer will be compensated if the producer decides to purchase the script and make the film. Rights to modify and adapt the screenplay now pass to the producer.
- The writer is compensated for giving the option (forgoing other opportunities until the option runs out), often receiving a fraction of the agreed purchase price (the option payment likely credits against the purchase price if the option is exercised).

This contract provides an efficient structure of incentives because it respects the sunkness of previously installed inputs in the creative good's production process and assigns decision rights to the party poised to decide whether and how to continue fabrication. However it does carry an odor of unfairness to the artist who supplies the initial and often defining input to the product, only to see decision rights about its subsequent fate pass to the hands of humdrum decision makers. This violates *art for art's sake* in the sense of putting the realization of the artist's conception in the control of other parties. The artist can bargain to retain decision rights, of course, as when a film director retains the right of "first cut" – assembling the raw film into a completed motion picture. However retaining decision rights over subsequent steps likely costs the artist dearly. That is because it conveys an unlimited opportunity to hold up collaborators deciding how to proceed subsequently with the project, unless the scope of the retained decision power can be clearly delineated (as with first cut).<sup>7</sup>

The most apposite theoretical analysis of this option contract appears to be Nöldeke and Schmidt (1998), who addressed alternative ownership arrangements for a project that involves sequential production processes undertaken in turn by parties *A* and *B*. Suppose that *B* holds an option to buy the project after *A* has sunk his investment but before the resulting surplus is realized. *B*'s reservation value for the project increases with *A*'s investment (effort). This provision strengthens *A*'s incentive to invest, because underinvestment will deter *B* from exercising his option. But *A* also does not overinvest, because *B* becomes the owner of the firm and captures most of the benefit of any excess in *A*'s investment. If *B*'s ownership is sufficient to induce efficient investment by *B*, then the overall contract is first-best.

Other terms of contracts used in creative industries have related incentive and efficiency properties. Consider the advance against royalties commonly paid by publisher

<sup>7</sup> A few demonstrations can be found of trade-off between *art-for-art's-sake* preferences and the terms of financing creative work. Fee (2002) compared films financed by the major studio distributors ("production-finance-distribution deals") to those financed independently by foreign distributors, personal funds, etc. Securing independent financing is a considerable burden for the filmmaker, but leaves her creative control intact. Fee hypothesized and confirmed that films would be financed independently when the film-maker's *art-for-art's-sake* tastes were particularly strong. This he proxied by situations in which producer, director, and screenwriter are all the same person.

to author or record label to pop musician. The royalties subject to the advance implement a sharing of expected net revenues from the project between (say) publisher and author. However the royalty is literally based on sales revenue, so the publisher faces an impaired incentive to make promotional outlays that “buy” additional sales revenue that flows partly into the author’s pocket.<sup>8</sup> A royalty is traditionally regarded as working capital to provide the impecunious author with bread while the creative throes proceed. However it has an important incentive property for the publisher. Until the advance is earned back from realized net revenue, the publisher retains the whole of the profit dollar elicited by its effort to promote the author’s work. Since the publisher’s promotional decisions typically matter more for the work’s profitability than any contribution the author can make post-publication (e.g., book-signing sessions, appearing on TV talk shows), the advance improves the efficiency of the contract.<sup>9</sup>

The advance figures distinctively in contracts between pop musicians and record labels because it not only anticipates royalties on the record but also covers the musician’s cost of recording the master tape – a substantial outlay when the musician favors elaborate electronic procedures that require costly studio facilities. The musician, prone to perfectionism (*art for art’s sake*), thereby gains a pecuniary incentive to make efficient rather than excessive use of studio time. The incentive may not work as intended, though, on a risk-loving musician prone to bet all available resources on the chance of a gigantic success.<sup>10</sup>

#### 2.4. Agents and matchmakers

The artist–gatekeeper relationship frequently involves an agent who mediates between artists and the enterprises that realize the market value of their creations. These intermediaries perform several services, depending on the creative sector. One is matchmaking between artists with heterogeneous talents and creative enterprises with diverse capabilities and input needs. Another is negotiating terms between artist and gatekeeper. As a third, the agent himself functions as a gatekeeper when he selects artists to represent.

The service ostensibly provided by the agent is to represent the artist (author, say) to enterprises that might bring her work to market (publishers). This representation function is governed by an incentive contract that compensates the agent with a share (traditionally 10 percent but with upward perturbations) of the artist’s gross earnings.

<sup>8</sup> Record labels rectify this incentive by charging some promotional expenses against the artist’s royalties, which indeed induces the label to undertake excessive promotion (the artist pays, while label and artist share the additional gross revenue).

<sup>9</sup> Hansmann and Kraakman (1992) developed some related propositions about the efficiency of contracts carrying an advance; they can deter the publisher from opportunistically declining a manuscript when (bad) fresh news arrives at a later stage in the publishing sequence.

<sup>10</sup> Another distinctive feature of popular-music contracts is the incorporation of a series of “cheats” whereby the musician’s contractual royalty percentage is nibbled away by costs (including wholly artificial ones) charged against royalties. Industry observers [Passman (1994)] conjecture that musicians gain utility from the right to brag of a high royalty rate, implicitly agreeing to its dilution.



This contract (including the 10 percent figure) was established in the nineteenth century at the inception of the agency business, quickly displacing a fee-for-services contract because of authorial poverty as well as its incentive value [Hepburn (1968)]. Besides representation, however, the agent performs a gatekeeping service that would otherwise fall entirely on the publisher. The agent can profitably undertake to represent an author only if the time (effort) devoted to seeking an outlet for her work is expected to reap sufficient compensation from the resulting royalties. The agent may also invest time (effort) in editing and improving the author's work, to the point where a publishing-house editor can appreciate its potential. Now consider the dealings that occur between established agents and editors employed by publishing houses. They interact repeatedly, which increases the editor's credence in an agent's pitch on its his author's behalf. The credence due to their repeated interactions is supported by the editor's inference that the agent will suffer a pecuniary loss from devoting effort to an author of indifferent promise – a substantial up-front opportunity cost with poor long-run prospects for compensation. For the publisher, relying on agents' representations (their gatekeeping skills and quality signals) substitutes for dependence on what can be picked from the "slush pile" of unsolicited manuscripts. That is likely a less efficient matchmaking procedure because the publisher pondering an unchaperoned manuscript lacks the information that the agent draws from personal contact with the author.

While the agent's gatekeeping and representation functions benefit the publisher, the agent's skill at negotiating on the author's behalf is adversary. Publishers offer somewhat differentiated bundles of services, but none capable of generating substantial rents. The author's unique manuscript is the one input into the publication venture with rent-yielding potential. Thus over the years the publisher's one-time share of subsidiary rights for paperback, cinema film, and other such derivative products has eroded, as the agent representing the author came to pre-empt the publisher and take over the auctioning of subsidiary rights. The publisher's gains from the agent's gatekeeping function thus trade against the publisher's reduced share of rents from subsidiary rights.<sup>11</sup>

Akin to the gatekeeping role of agents is the function of certifiers who possess or invest in skills at making fine judgments on the quality of artists or their works. Theoretical research has recently turned to characterizing the market for certifiers' services, including the vertical differentiation of their services [Hvide and Heifetz (2001)]. The critic's economic function in creative industries has not been much studied, but on casual evidence seems to possess some analytically interesting features. Major acquisitions of visual art excepted, the individual's decision to consume a creative good is too small a transaction to warrant a large outlay on an advisor's services.<sup>12</sup> So critical opinion is commonly bundled into magazines or newspapers along with complementary

<sup>11</sup> The hard-back publisher's one-time substantial share of subsidiary-rights income clearly had an incentive value for the publisher's promotional efforts. Apparently authors (and agents) have adjudged the value to them of that incentive to be less than their gains from redistributing the rent stream.

<sup>12</sup> Large investments may be made in personal search and inspection, however. It is no doubt difficult for the consumer to convey her tastes to the advisor.



sorts of information. The amount of criticism supplied then depends on its marginal attraction to consumers of the bundle relative to their marginal valuations of other content. Critical services seem subject to vertical differentiation parallel to the differentiated involvement of consumers in various arts and entertainment industries. That is, the utility one gets from consuming creative goods increases with one's accumulated "cultural consumption capital" – built up from previous experience and both specialized and general training [Stigler and Becker (1977)]. Individuals vary in both aptitude and desire for building such stocks of consumption capital. As a result they tend to distribute themselves between the poles of "buff" and "casual" in their involvement. The judgments offered by critics and certifiers tend to display a parallel vertical differentiation, with reasoned and contextualized evaluations provided for the buffs, while the critic servicing the casuals tends to internalize their standards and opine whether or not they will like the work.

### 2.5. Agency and intermediation

Several unanswered analytical questions bear on how the agent's function is organized. The agent's primary function is analogous to intermediation or job-matching. A good deal of research has been done on the theory of intermediation.<sup>13</sup> It focuses on the factors giving the intermediary a productivity advantage over search by the individual primary sellers and buyers. It also addresses the determinants of the intermediaries' price-cost margin, which include their number and mode of competition with one another. In this literature the productivity of intermediation stems from transaction-cost advantages, inventory-holding advantages, and/or advantages in ascertaining quality and warding off adverse selection [for example, Biglaiser (1993)]. Unfortunately, none of this agrees very well with what seem the basic properties of matchmaking in the creative industries. Because *nobody knows*, impacted information seems not to be a problem, although collecting information in order to match the attributes of heterogeneous buyers and sellers involves significant costs. Also, explaining the equilibrium market price of the agents' services cedes place to the problem of explaining why a long-established revenue-sharing arrangement between agent and artist-client should seem conventionalized and immune to supply and demand disturbances. Consequently the following theoretical propositions are advanced tentatively.

Suppose that authors are to be matched to publishers. Members of each group possess a given set of differentiated attributes that are objective and can be determined at a cost by another party (no hidden information). Assume that the value created by pairing any author and publisher depends on the attributes of the two together. Assume that some optimal allocation of authors to publishers exists and can be calculated by any agent who has acquired information on every candidate party's attributes. Because the parties' haggling over terms holds no special interest, we assume that each author's outside

<sup>13</sup> Spulber (1999) provided a survey.

reservation price is zero, and that matched parties always reach a Nash bargaining solution. In these conditions the agent's matchmaking advantage is a natural monopoly: the attributes of each author and publisher need be collected but once by the agent who determines the optimal allocation, while decentralized match-making requires repeated collection of the same information. The gain from a central agent might be compromised by bounded rationality or its temporal equivalent – costs of delay while a central agent collects and processes all parties' information. Any such source of diminishing returns will tend to increase the number of agents and make the equilibrium population of agents an increasing function of the numbers of authors and publishers to be matched and the cost of gathering information from each. The average quality of the matches of course declines.<sup>14</sup>

It can matter who employs the match-making agent. If authors and publishers were like right and left shoes, the agent could simply enter the market as an entrepreneur, purchasing isolated rights and lefts, matching them and reselling pairs at a profit. Where units of human capital are being matched, this procedure clearly fails, and some party must recruit the agent and serve as principal in a governance relationship. This need for governance seems to inject an intrinsic asymmetry into the agent's activities. Can the agent work at random for parties on either side of the market? At first glance that arrangement seems viable but it is probably not. The marriage broker representing both brides and grooms has an incentive to provide a groom with a bride on his representation list, which need not maximize benefit to the groom (or the couple together). The agent negotiating a dowry incurs a clear conflict of interest if he represents both parties. Even without these governance problems, economies of specialization call for the agent to work for parties on only one side of the market. That being the case, we can ask what asymmetries or differences between the types of entities to be matched affect the question as to which of them more efficiently takes on the task of employing and monitoring the agent.

Let us return this question to the context of author and publisher. Suppose that publishers' attributes and policies are readily inferred from their backlists and reputations with authors for capability and integrity; authors' qualities on the other hand are more costly to identify. If the transacting parties on one side of the deal are represented by agents who pool information and economize on its transfer to the other side, pooling the more costly assessments of authors and their manuscripts beats pooling the less costly assessments of publishers' traits on behalf of authors. The same logic applies to the differential importance or value of the information to the other side. A publisher loses heavily if a celebrity author's book flops, but the celebrity might be nearly indifferent about which of several mainline trade publishers issues the book. Intensive collection of information about authors is then more valuable and lays claim to the agent's services, and the collection of information about publishers gets left to individual authors.

<sup>14</sup> The effect of costs of gathering information should depend on whether agents can segment the market's population and avoid duplicating collection costs.

Other influences also weigh in. Suppose that it costs the same for a publisher to size up a prospective author as for the author to evaluate a prospective publisher. Even after the gatekeeping agents have swept out the losers, the authors remain more numerous than the publishers (think of each house serving as exclusive publisher to a number of authors). If agents are to pool information on one side of the market only, they should pick the more numerous authors, thereby consolidating more information than if they represented the less numerous publishers.<sup>15</sup> The agent's gatekeeping function, which excludes many authors, is really a special case of this "differential numbers" effect.

Only descriptive information is available to provide empirical evidence on these predictions. Natural-monopoly tendencies were evident long ago in the booking of variety and vaudeville acts into local theaters [Poggi (1968, pp. 11–26); Sanjek and Sanjek (1991, Chapters 2, 3)]. The booking of big bands in the United States during the 1920s through 1940s provides a particularly interesting case because it illustrates the factors that served to cast up a near-monopoly intermediary, Music Corporation of America (MCA), to undertake the matching of bands to venues. Its dominance was owed, however, not so much to scale economies in matching as to two other factors: scale economies in managing bands' travel arrangements, and the credence value of a large organization with substantial fixed assets for ensuring the intermediary's responsibility in adhering to contracts and remitting payments [Walker (1964, Section 2, Chapter 5); Stowe (1994, pp. 103–106)]. MCA may also have benefited from forcing exclusive dealing on venue operators, requiring them to book only MCA bands if they obtained any of them [McDougal (1998, pp. 108, 128, 224)]. Finally, descriptions of the internal operations of large Hollywood agencies such as Creative Artists Agency (CAA) show that their individual agents work for limited numbers of artists, consistent with the hypothesis of size limits stemming from individual agents' bounded rationality and time costs. Scale economies stem importantly from the unfettered exchange of information among agents working for the same firm and the opportunity to assemble projects as packages of the various talents represented by a given agency [Slater (1997)].

## 2.6. Internal organization of gatekeeping firms

The gatekeeping process influences the internal organization of firms in creative industries because the gatekeeper (subject to the top manager's review) selects the projects undertaken by the firm and thereby makes its fundamental investment decisions. While the gatekeeper's primary qualifications presumably lie in detecting the potential of the artist's talent and perhaps working with the artist to ripen its fruits, the task of implementing the firm's investment decisions implies a wider scope for both responsibilities and incentives. An efficient compensation scheme will reward the gatekeeper for a good pick and penalize a bad one. The decision-maker responsible for selecting a project

<sup>15</sup> We neglect the plausible outcome of specialized agents on both sides of the market: plaintiffs' attorneys and defendants' attorneys.

therefore has an interest in the proficient performance of the subsequent processing steps that will affect the venture's final profitability. The implications of this organizational logic have worked themselves out in the U.S. publishing industry in modern times. The editor who performs the gatekeeping function was once just that – the party responsible for selecting the manuscript and polishing it to lapidary perfection. Promotional campaigns for books with the potential for wide popularity (“blockbusters”) have emerged as a central function of the publishing house, so the payout of the editor's pick becomes heavily dependent on the prowess of the promotional campaign, which therefore logically displaces copy-editing as the editor's primary entrepreneurial task [Whitesides (1981)]. Editors' compensation does not seem closely related to projects' profitability (as it might be for a sales representative or securities trader). However successful editors do realize important rewards (such as boutique labels of their own), and those who pick a run of unsuccessful projects can expect to be seeking new opportunities. Their counterparts in some other creative industries – motion-picture and Broadway stage producers, for example – receive explicit profit shares of successful ventures while being largely protected from the pecuniary downside of a failed project. That asymmetrical form of incentive compensation is likely optimal when the gatekeeper is risk-averse and the project's success highly uncertain even after the gatekeeper's best entrepreneurial efforts.

The logic of the gatekeeper's function – as editor, record-label talent scout, movie producer, etc. – also implies something about the external contacts of these persons and their mobility among firms in a creative industry. These patterns were first noticed by sociologists interested in the communication that occurs among skilled project-runners employed in different firms [Rogers and Larsen (1964); Powell (1985)]. Economists commonly assume that firms vigorously protect from prying eyes all information on their internal activities. Leakage of this information presumably allows competitors to copy the secrets of the firm's productivity or forestall its strategies in the market. However, creative industries deal with vast numbers of potential projects that will ultimately not go forward (*infinite variety*), and many others that are pursued but ultimately fail. Information on what projects have failed for a given firm, and what projects look promising but are not right for the firm considering them, largely lacks this value of confidentiality. Its access by a competing firm creates potential value for that firm without imposing any direct cost on the firm that divulges it. Proprietary information therefore becomes valuable trading stock among the gatekeepers employed by competing firms, for whom a transfer of useful information today creates a claim on some reciprocal tidbit in the future.<sup>16</sup>

Parallel to this mobility of information among a creative industry's gatekeepers is mobility of the gatekeepers themselves. Their human capital is not firm-specific, in that the differences among firms lie not in systems or technologies outside the gatekeeper's

<sup>16</sup> There appears to be room for theoretical research in this area. For a possible approach, see Baron and Besanko (1999).

control or authority but in the selection strategy that the gatekeeper himself implements. The gatekeeper's personal goodwill assets arise from his recurrent dealings with agents, artists, and other gatekeepers through the informal trade in information, and mainly resist appropriation by the firm that employs him. Accordingly gatekeepers in creative industries exhibit high levels of job mobility. Among publishing houses and record labels this mobility appears as simple job-hopping. In cinema films and TV programs the gatekeeper (producer and writer-producer, respectively) is usually an independent agent or the proprietor of a small firm who may reach a housekeeping deal with a large firm such as a movie studio. Such a deal provides the agent with a base for operations and seed money for developing new projects, while the host obtains the right of first refusal over the agent's projects for a period of a few years.

### 3. Structures and contracts in complex creative industries

A useful if rough distinction can be made between creative industries that rely on the works of a single class of artist and those whose products combine several of them. Producers of diverse creative inputs may interact in complex ways that would generally not occur among humdrum inputs. Artists may have *art-for-art's-sake* tastes that embrace preferences over how the common creative task is performed. Bringing several sets of divergent preferences into consonance poses a thorny bargaining problem, as the preferences likely defy exact codification and negotiation to reach a mutually satisfactory contractual agreement. Indeed accounts of the development of cinema films and stage plays and musicals commonly refer to disputes among artistic personnel over competing visions of the end product. These get resolved through the application of "muscle", the dominance of the preferences of whichever participant would most harm the project's prospects by withdrawing. The threat value of withdrawal is likely related to the participant's track record of superior performance (success) and/or distinctive and irreplaceable skills or attributes [Rosenberg and Harburg (1993, Chapter 7)]. In a creative project a compromise among several coherent but disparate conceptions of a creative project incurs an obvious danger of becoming a failed mishmash, whereas one or more of the pre-compromise visions might have succeeded on its own.

#### 3.1. Integration and disintegration

Among creative industries the U.S. motion-picture industry in the twentieth century provides a striking example how an industry's organization can be determined by the relative feasibility of different structures of contracts interacting with consumers' valuations of various types and qualities of creative goods. The major Hollywood studios emerged in the 1920s, integrated vertically from the production of films through their distribution and exhibition, and retaining under contract or regular employment many of the creative and specialized technical skills needed to produce cinema films. The efficient deployment of these film-making inputs under contract required that films be

turned out at a regular pace in order to keep the distribution and exhibition pipeline full, and each studio's portfolio of films was composed so as to make full-time use of the costly talents under contract. This assembly line rolled out a well-defined product – films relatively standardized in quality and style, even though varying in their reception by audiences and critics. A distinction was regularly made between “A” and “B” pictures, the latter of lower quality and shorter running time, intended to play as second features on a movie house's double bill. The B pictures were made by separate units of the major studios as well as by independent studios, and they served *inter alia* as training grounds for novice movie talents.

Contracts between the studios and key actors and other artists took the option form explained in Section 2.3 above. An actor was bound to the studio for a maximum period of seven years in half-year steps, with the studio holding the option either to renew (with a salary increase) or terminate every six months. The studio was thus motivated to make a substantial investment in the career of a promising performer, as it could collect the resulting rents over what was likely to be a substantial proportion of the actor's career. The studio retained decision rights over major creative choices such as which roles the actor undertook. Apparently the actor's pay was often renegotiated prior to her contract's expiration when she rose to star status, and stars gained access to the perks for which Hollywood is legendary. However even stars got no control over what roles they undertook. A device that no doubt helped to resolve the disputes arising under these contracts was that of lending out the performer to make a single picture at another studio. For this the contracting studio collected not only the fee due the performer but also a premium for its own treasury. Beyond its use in resolving disputes and tensions due to long-term option contracts, the loan-out procedure relaxed the constraint on film quality that was inherent in the studio's presumption that each film would be made by those inputs it had under contract and currently available for work.

The studio system was transformed in the 1940s and 1950s into a completely different organizational structure bound together by different contracts. It is a striking example of how the optimal organization of an activity can make a large, discrete switch following an exogenous disturbance affecting demand for the product and/or the feasible (legal) set of contracts. Several factors precipitated this change, and their respective necessity and sufficiency are not entirely clear. One was the introduction and diffusion of television, which provided to movie-goers a low-quality but cheap and convenient substitute for cinema films. With TV entertainment available at no pecuniary or travel cost, the cinema film (especially the B picture) faced formidable competition. The industry made the theoretically predictable adjustment of reducing the quantity of cinema films produced and raising their quality in the sense of employing costly inputs and elaborate special effects that distanced the cinema film from its small-screen competitor. Another major causal factor was the antitrust case *U.S. v. Paramount Pictures* 334 U.S. 131 (1948). Its principal consequence was to terminate vertical integration of the Hollywood studios into exhibition, which had been implemented mainly through ownership of then first-run downtown movie palaces. Although no studio by any means exhibited its films only in its own cinemas, this forward integration had effectively mandated a pace of movie

production to permit regular weekly changes of bill at each theater [Conant (1960)]. Disintegration tended to reduce the optimal rate of film production by the studios, and each film came to be marketed individually for exhibition in each city.

The mandated divestiture of exhibition and vertical differentiation led to the complete reorganization of studios by means of the divorcement of film production from exhibition. The central change was that each film was now assembled *à la carte* from the most suitable inputs available anywhere in the Hollywood talent pool. This change was congenial to the vertical differentiation of films from TV fodder, because it relaxed the constraint of reliance on the single studio's talent pool, and it facilitated the production of films differing widely in scope and ambition. Numerous fixed facilities were dismantled – each studio previously had its own production lot with pre-built sets, its own symphony orchestra, etc. Now, these and many other inputs and skills were hired temporarily as needed.<sup>17</sup>

The subsequent transformation of Hollywood's organization was traced in a series of papers by Christopherson and Storper [Christopherson (1992); Christopherson and Storper (1989); Storper (1989); Storper and Christopherson (1987)]. The following changes are documented in their research unless other sources are cited. Film production shifted away from the major distributors (studios) to smaller distributors and independent firms. Different bundles of inputs tended to come together for each film, with little "repeat business" [Lazarus (1985, pp. 94–95)]. Other packagers of film inputs such as talent agencies got into the business of organizing film production [Slater (1997)]. Many specialized independent service firms arose to provide film-making services on demand, and these firms diminished in average size as they became more numerous. The great increase in the number of arm's-length transactions involved in making a film entailed transaction costs that were mitigated if the input suppliers clustered closely in the Los Angeles area, and this centripetal force was evident. Personnel came more typically to work part-time for several employers, and the craft unions that had long siphoned substantial rents from the major movie studios found themselves unable to control access to competitive supplies of skilled labor. The factors explaining the year-to-year variation of studios' profits shifted from what stars they had under contract and theaters under control to the quality of films produced [Miller and Shamsie (1996)]. Independent markets for the exhibition of completed films sprang into existence in North America (Sundance, Toronto) as well as abroad (Cannes) [Donahue (1987)].<sup>18</sup>

<sup>17</sup> Other exogenous changes played minor roles in this transformation. During World War II high personal income-tax rates encouraged individual stars to form their own production companies to rent their services to a studio on a per-picture basis, and this tax dodge became a precedent for the later regime of one-off deals. Changes in camera technology made shooting films on location much easier, reducing the usefulness of studio lots. Also, when the *Paramount* decrees expired in the 1980s, there was no substantial return to the vertical integration of distribution and exhibition. Regarding the relative importance of these exogenous changes, it is noteworthy that the British film industry underwent the same disintegration as the American; the British film-makers also faced rivalry from TV, but they encountered no *Paramount* decision, suggesting the sufficiency of competition from television as explanation of the British disintegration.

<sup>18</sup> A somewhat similar disintegration has occurred in the television programming market in Great Britain; see Starkey, Barnatt and Tempest (2000).



### 3.2. *Contracts for complex creative goods*

When a complex creative industry's organization is transformed from extensive integration to "flexible specialization", we expect that a congenial set of contractual arrangements will emerge to govern the now-independent dealings. The motion-picture industry provides an attractive case study, along with its adjunct the market for TV program series. Cinema film projects usually start from a speculative script or a literary source from which a script is to be developed. Real option contracts provide workable governance for the producer (a film's entrepreneur) who seeks to develop a script. A series of steps is defined – a treatment (synopsis), a full draft, revision, polish, and so forth – with the writer paid for each step and the producer holding the option to continue with the next step. The screenwriters' organization, the Screen Writers Guild, supervises an arbitration procedure to determine the allocation of screen credit in the common case where a script passes through the hands of several writers.

With the script developed, the producer seeks to assemble the creative and technical inputs needed to make the film and to ensure that they are available at the right times for a closely coordinated series of sequential steps. This task can encounter hold-up problems. The last input committed to the coalition acquires some hold-up power due to the sunk negotiation costs and foreclosed alternatives already incurred by the other participants. Participants with high opportunity costs may insist on play-or-pay contracts that require them to be paid for their availability at a particular time even if the film is postponed or canceled. Play-or-pay commitments, though, can be renegotiated or traded off to another producer. Films sometimes fail because of defects in this contracting process, as when the coalition is assembled before all problems with the script are resolved or a key participant obtains a commitment to gratify some whim or preference that proves fatal to the film's overall success. However the process seems about as orderly as is feasible for investment decisions about which *nobody knows*, and the Hollywood studios at any time have hundreds of projects at some stage of development, so that those actually "greenlighted" have for better or worse survived an arduous winnowing process.

Motion-picture contracts commonly specify contingent compensation for major participants, bestowing shares of gross rentals received from exhibitors, net profits, or some variant on these. The exact incentive content of these deals is the subject of controversy. Participants sometimes take revenue or profit shares rather than straight compensation in order to assert their conviction about the project's merit. Contingent pay may also serve to induce effort, for example for the principal actor in the  $n$ th film of an action-adventure series [Chisholm (1997)].<sup>19</sup> However the compensation of highly-paid participants might be contingent simply because a gigantic fixed fee (the alternative) would shift enough of the film's overall risk on the other claimants to the film's cash

<sup>19</sup> Whether or not incentives are important in contingent motion-picture contracts, they are prominent in other contracts used in the industry. Goldberg (1997) showed how various contingent contracts reflect the timing of the opportunities open to various parties that can enhance the value of a deal.



flow to drive up their reservation prices substantially [Weinstein (1998)]. Whatever reasons for its use, compensation based on net profits suffers from the fact that the studio that distributes the film keeps the books in which net profits are determined. While the elements of creative accounting that enter studios' profit determinations generally have cogent economic bases, the studio has great scope for moral hazard by inflating judgmental costs and allocations that favor its own stake and invade the stakes of other participants – the notorious Hollywood accounting. The participants seek to avoid this moral-hazard problem by pre-negotiating the definition of net profits or demanding a share of gross revenue rather than profit, but transaction costs by themselves impair the use of contingent compensation.

### 3.3. Vertically differentiated talents

Complex creative industries employ teams of functionally differentiated creative talents, but their vertical differentiation – the *A-list/B-list* property – is also important for the industry's organization. Questions arise about how entrepreneurs selecting creative inputs and outputs deal with differences in inputs' quality. The obvious assumption is that the quality of a creative good perceived by the market generally increases with the quality of its creative inputs. An input's quality is optimally raised until the expected increment to revenue from the project falls equal to the incremental cost of quality. Another production relationship encountered in creative industries is that inputs' qualities interact multiplicatively rather than additively in determining the market's expected valuation – the O-rings production function. In these circumstances the failure or substandard performance of any one of them shrivels the value of the whole project. Also a high-quality input added to a project staffed with other high-quality inputs generates more incremental value than if it were added to a low-quality project. Faulkner and Anderson (1987) provide evidence supporting one implication: that the more successful producers, directors, and cinematographers tend to work together with significantly more than random frequency.

The importance of vertical differentiation in creative inputs raises the question as to what mechanism evaluates and records quality rankings of competing creative inputs. This is not quite the question of whom the gatekeeper admits and excludes, although admitted talents (visual artists, musicians, authors) undergo a ranking by critics' and consumers' evaluations. In complex creative industries such as cinema films, an economic mechanism underlies an evaluation process that continuously ranks the members of a given creative group. They rate each other's performances on professional prowess independent of the overall success of projects in which they participate. Each talent profits from having accurate knowledge of her own position on the quality scale: it reduces transaction costs and opportunity losses associated with competing for projects slotted above her quality ranking, or accepting projects targeted below it.<sup>20</sup> Knowledge of the prevailing rankings may generate value in other ways as well. For example,

<sup>20</sup> Analogous to the profit gained by a Cournot competitor from knowing accurately its variable costs relative to those of its rivals [Shapiro (1986)].

composers providing background musical scores for films sometimes become over-committed and need to subcontract jobs to newly arrived and/or lower-ranked talents [Faulkner (1983)]; hence there is also economic value to the talent from participating in the random interchanges with peer talents in which rankings are discussed and assigned and the emerging consensus reported. Furthermore, interest in the *A-list/B-list* ranking of a given class of talents extends to other types of talents who participate in the common set of projects. Each needs to know the talent rankings in other specialties in order to infer correctly the quality ranking of the project as a whole. Thus the whole system operates to create and update a consensus judgment on creative talents' rankings that is available to producers when they select inputs for a project.

### 3.4. *Distributing complex creative goods*

Complex creative industries, like other fabricating activities, face the problem of efficiently distributing their product to ultimate consumers. Distribution here embraces informing consumers about the varieties available as well as making them physically accessible. Creative industries such as publishing and sound recordings face distinctive problems due to two underlying structural features. First, the creative product is an "experience good" that cannot be accurately evaluated by final buyers short of actually consuming it. The supplier therefore faces a problem of how to disseminate information that will hold some value for signaling a match to consumers' tastes. Second, the product line offered by an industry embraces a large number of individual differentiated goods that are bought at retail in small dollar amounts. The marketplace must solve the problem of efficiently distributing many small units and holding them in inventory to await the exercise of consumers' uncertain demands.

A distinctive factor for book publishers and record labels is that retail inventories themselves perform a promotional function, because consumers learn about potential purchases through browsing in retail outlets. The retailer captures part of the extra sales revenue generated through enlarging his inventory and makes inventory decisions in light of that partial reward. The publisher, who shares this revenue, has reason to induce the enlargement of retail inventories. The publisher could use several policies to achieve this goal, and two of them hold particular importance in practice. One, employed by U.S. publishers since the 1930s, is to allow the retailer free return of unsold books. This privilege reduces the retailer's cost of holding inventory and increases its level. The retailer still incurs the cost of packing and shipping unsold books for return, but this bite out of net revenue is likely less than the alternative of knocking down the retail price until the book is sold. The publisher incurs costs of printing and one-way shipping of returned books as well as dealing with the returns (pulsing them, or disposing of them through specialized discount booksellers). An alternative policy, common until recently in countries outside the United States, is resale price maintenance (RPM), by which the publisher's contract with the retailer requires that books not be sold below the publisher's suggested retail price. RPM increases the bookseller's gross profit margin on books that due to retail competition might otherwise be sold at a discount – presumably

“best sellers”. It thereby increases the retailer’s optimal inventory of best sellers. It could increase the equilibrium number of retail bookstores, but it does not directly promote enlarged inventories of less popular titles. Thus, free returns and RPM both rectify what are regarded as market distortions from the publisher’s point of view, but each is prone to generate further distortions of its own. From society’s viewpoint there is no presumption about which policy to prefer.

The promotional role of retail inventories holds another implication for the distribution of creative goods. The retailer’s shelf space itself holds value for the publisher or record label because of the extra profit generated by additional copies sold to customers acting on impulse. That potential profit opens an opportunity for the retailer to charge the publisher for exhibition space in highly visible locations within the store, such as tables near the front door. Such “slotting allowances”, also familiar in grocery retailing, similarly apply to other distinguishable forms of promotion that the retailer undertakes [Shaffer (1991)]. To the publisher, paying the retailer directly for the value generated by these promotional policies is a partial substitute for the free returns and RPM already mentioned.

When failures occur in contracts between producers and distributors, a classic remedy candidate is vertical integration, which substitutes administrative direction for arm’s-length dealings. Integration has seen some use in creative industries, notably between record labels and wholesale distributors during the 1970s. It apparently arose from the labels’ need to coordinate the distribution and promotion of those pop recordings with the potential for large sales. For example, when a musician tours in order to promote a new recording, large stocks of her new release and past hits need to move in timely fashion through the distribution sector and into retail stores in cities on the tour. Without integration the label cannot readily induce independent wholesalers to stock at levels that will maximize their joint profits.<sup>21</sup> The solution of vertical integration has important implications for concentration among the record companies that are discussed in the next section.

Thus, the producer of creative goods such as books and sound recordings faces inter-related problems of informing consumers and efficiently distributing many small units of differentiated creative goods. The available instruments, including free returns, resale price maintenance, slotting allowances among others, are substitutes for one another in some settings, complements in others.

### 3.5. Structures of creative industries

Many observers have noted the high concentration of sellers in several creative industries – in the United States, and also for some worldwide. Concentration has been rising

<sup>21</sup> The scope and role of independent book wholesalers has changed repeatedly, reflecting in an interesting way organizational changes in both publishing and retail bookselling, as well as underlying costs and technology; see Miller (2003).

in some sectors, long stable in others. It shows no signs of transience. In motion-picture distribution roughly five to eight major distributors have dominated the industry since the 1920s. Although the market shares of individual distributors fluctuate considerably from year to year with the success levels of their films, the group's combined share is quite stable. Sound recordings have become concentrated since the 1970s, with the largest five or six companies recently accounting for about 80 percent of shipments worldwide and with more concentration in prospect. In book publishing, once unconcentrated, the largest four U.S. publishers in 1993 accounted for 30 percent of wholesale revenue.<sup>22</sup> A common explanation for concentration lies in the scale economies and sunk costs of the physical distribution system. The evidence for sound recordings was noted in the previous section. For motion pictures the distribution system manages the promotion of new films, negotiates exhibition contracts with individual theaters and carries out the physical distribution of prints of each film. Compelling circumstantial evidence holds that distribution systems determine the concentration of these creative industries. In motion pictures the disintegration of both exhibition and production left the distribution systems in place and unchanged in their concentration. In sound recordings the concentration of production coincided with vertical integration into distribution [Belinfante and Johnson (1982)]. In book publishing physical distribution plays a less central role, but the promotion of blockbusters is a function that fosters large-scale firms [Whitesides (1981)].

This hypothesis about concentration and its roots in creative industries is usefully probed by examining the responses of these oligopolies to major disturbances. The toys and games industry, with many marks of a creative industry, shows a noteworthy contrast to films and sound recordings, with their stable populations of dominant distributors and churning fringes. Toys and games exhibit the same scale economies in the promotion of blockbuster toys, where indeed the degree of dominance of the most popular designs if anything exceeds that of other creative industries. In toys and games, however, there are no substantial sunk assets in the form of production facilities, distribution systems, or promotional organizations. Manufacturing is generally contracted out. When a toy succeeds wildly, sales of the firm that introduced it can suddenly expand greatly, but no new fixed facilities or lasting goodwill assets result. When a toy expected to be popular fails in the market, or even when the popularity of a continuing success declines unexpectedly, quite large toymakers can run losses and disappear overnight [Stern and Schoenhaus (1990)]. The record industry's responses to major stylistic shifts in popular music provide a valuable laboratory experiment. When rock 'n' roll first came on the scene, it offended the ears of the established artist and repertory (A&R) executives, who kept their companies out of the business. The result was increased churning in the weekly Top 10 records and deconcentration of the record industry. The incumbent leading firms learned a lesson, however. They became much more open, starting or acquiring new record labels to exploit new musical styles as they emerged, making use of

<sup>22</sup> *Billboard*, January 21, 1995, p. 42; *Book Industry Trends*, 1995, quoted by Greco (1997, p. 58).

the firms' distribution and promotion capacities and raising the industry's concentration to its current high level [Peterson and Berger (1975); Lopes (1992)].

This source of concentration in creative industries clearly reflects one distinctive underlying feature of their structures, the proliferation of differentiated creative goods (*infinite variety*) and the associated scale economies in their distribution. More subtly, it reflects another – the incompatibility of the romantic concept of the artist, as the autonomous creator working from inner necessity, with the requisites of collaboration in a large and hence bureaucratic business organization. The typical creative industry contains large-scale firms organized around promotion and distribution, but also a large number of small firms organized around the gatekeeping functions of selecting and nurturing promising artists. The latter activities require personal rapport and suffer diseconomies of scale in effectiveness when organized on large scales. The large firms whose sizes are driven by scale economies in promotion and distribution can be labeled *promoters*, the small firms specialized in gatekeeping and nurture are *pickers*. The advantage of their separation lies quite simply in keeping artists and bureaucrats out of each other's hair. The picker/promoter distinction clearly appears in motion pictures, TV films and programs, first-run TV syndicators, record labels, art galleries, and book publishing. Mezias and Mezias (2000) identified essentially this distinction in the early (1912–1929) motion picture industry between the generalist firms vertically integrated in production and distribution and the innovative specialist firms operating at only one of these stages. It also appears in creative industries that fail to exhibit levels of concentration that are high by common standards. Picker art galleries are typically run by individuals with strong sympathy for and involvement in creative processes in the visual arts, hence well-attuned to dealing with artists. Promoters tend to represent artists who have achieved some measure of success and can be attracted away from a picker gallery by the offer of more effective promotion on a larger scale.<sup>23</sup> Agencies representing classical musicians similarly divide into small-scale units that take a hand in their artists' personal development and large-scale units suited to the job-matching tasks of assigning the top performers.

Another factor explaining differences in the organization of creative industries is the way in which the cost of quality varies with the quantity of output produced. If quality is a variable cost, high quality will enter into marginal cost. Firms offering different quality levels at correspondingly different prices will likely survive in the market. If quality is a fixed cost, however, it does not enter into marginal cost, and a high-quality firm will tend to undercut lower-quality competitors. The market will be highly concentrated with little variation of price with quality; where quality is a variable cost, more competitors will survive, and a range of qualities will be offered at diverse prices. Berry and Waldfogel (2003) showed that this difference in the variance of prices holds between two local creative industries – restaurants (quality enters into the meal's marginal cost)

<sup>23</sup> An important feature for the viability of these two classes of enterprise is the ability of the pickers to realize capital gains on works created by successful artists before their prices were lofted by large-scale promotion.

and newspapers (tabloid and “newspapers of record” incur the same marginal printing costs). Fixed costs of quality contribute to explaining many important features of creative industries, including the “superstar effect” [Rosen (1981)] and the dominance of the United States as a supplier in the world market for cinema films [Wildman and Siwek (1988)].

### 3.6. *Multi-activity firms in creative industries*

Large firms in creative industries commonly engage in numerous activities, earning them the label of “entertainment conglomerates”. The term is deceptive, however, because their parts are assembled in ways intended to generate rents – extra net revenue over what the same activities could earn if organized as free-standing single-business firms. Synergistic gains claimed by managers bent on merger often come into question when subjected to economic analysis.<sup>24</sup> One business model of gains from diversification stems from the observation that core creative works can yield rents from their embodiment in a number of different forms. The novel that arrives on the doorstep of a trade publisher is first issued as a hard-cover book. Then it appears in paperback. It is turned into a screenplay for a motion picture. The motion picture is “novelized”, the screenplay turned into a book with still photos from the movie. The soundtrack score is issued as a recording. A television series is conceived as a spin-off from the movie. The simplistic account of organizing these transactions holds that the firm possessing a business in each product line can simply pass the core creative input along from one of its divisions to the next, with cash gushing forth at each step. The problem with that program, of course, is that the author of the primary book manuscript is the legal owner of the core intangible asset, and she (or her literary agent) can readily stage an auction of this literary property in each of the markets where rents can be generated. Standard hard-cover publishers’ contracts have sought to divide the rents from subsidiary rights evenly between author and publisher. Indeed, that distribution might be agreeable to (optimal for) a novice author whose novel lacks manifest signs of major subsidiary-rights potential, so that the value of these rights depends substantially on the publisher’s efforts to develop and promote the work. However acceptance of that arrangement is the author’s option. If the author chooses to stage her own auction of subsidiary rights, it is not obvious why bidders affiliated with the hard-cover publisher should be able to wring more value from the project than independent bidders and therefore to prevail in the bidding. An auction is a highly efficient device for extracting value from bidders and a successful conglomerate needs some special trick to generate more value than independent firms bidding in the author’s auction.

<sup>24</sup> That risk-spreading is not featured to explain the diversification of firms in creative industries may seem surprising. The case for imputing a risk-averse preference function to the publicly-traded firm is always shaky, and little empirical evidence ties the entertainment conglomerates’ behavior to risk aversion. Smaller firms in creative industries, whose uncertain cash flows directly affect the welfare of potentially risk-averse individuals, commonly behave as if they are risk-loving due to *art for art’s sake* (see Section 2.1 above).

This is not to deny synergistic gains for entertainment conglomerates, but rather to point to the stringent conditions for realizing them. When core creative elements can be embodied in many different forms, some ways to create value may not be fully exploited by sequential independent auctions. The cash-flow streams of the diverse embodiments likely are interdependent. Coordination at the outset in styling the element for its first use raises its value in subsequent uses as well. Outlays to promote one embodiment yield spillover benefits for others that must be taken into account in order to maximize value. Internalization can facilitate this coordination, which is likely difficult at arm's length.

Another feature of creative industries that promotes multi-activity firms grows from the public-good character of some core creative products. The cinema film, the television sitcom series, the pop song or recording once created can be used repeatedly at no marginal cost, or at only a small incremental cost of putting the good in another form. However the creative good must earn rents in these various uses if its original fixed cost is to be recovered. No matter how many creative goods compete for any given use, nor how closely competitive are their providers, the equilibrium price must exceed the zero marginal cost.<sup>25</sup> This property of all information-type goods creates incentives for vertical integration between businesses that provide creative content and those that distribute it. Consider one of the vertically integrated firms that account for most U.S. television networks, say, Disney and ABC. If ABC shows a film from (say) Paramount's library, the payment is a cost to ABC and a rent to Paramount. If ABC instead draws from Disney's library, any payment is purely an internal transfer, and Disney/ABC incurs no cash cost. The incentive to internalize the transaction applies to Paramount-UPN and any other integrated firm comprising a content provider and distributor. The individual firm's incentive to internalize this rent transfer might be augmented by a strategic consideration. A non-integrated firm that depends on vertically integrated rivals for content or distribution feels itself vulnerable to foreclosure by its integrated rival. Foreclosure and refusal to deal are by no means necessarily profitable to an integrated aggressor, but states of nature can occur in which they would be; in such vertically related oligopoly industries firms commonly explain their actions as if they believe that the likelihood of a costly hold-up is substantial.<sup>26</sup>

This incentive to internalize transactions and ensure against foreclosure would apply in any industry that produces information, or any other good with a fixed but no marginal cost. It holds further significance in creative industries, though, because of two of their distinctive properties: the great uncertainty about consumers' valuations of a yet-uncompleted good (*nobody knows*); and the prevalence of large numbers of products (movies, sitcoms, pop recordings) that are close but imperfect substitutes (*infinite variety*). When the ABC network restricts itself to showing Disney programs and

<sup>25</sup> If competition between vendors of creative goods drives their prices below average cost, equilibrium requires that some sellers exit until a price no less than average cost is sustainable.

<sup>26</sup> For a rigorous demonstration of how a firm could gain by vertical foreclosure, see Ordovery, Saloner and Salop (1990). A connection to incomplete contracts is made by Bolton and Whinston (1993).



movies, it narrows its choices and presumably offers viewers a less attractive menu than if ABC picked freely among all studios' offerings. This is a real cost of the internalization, though (with close substitution) likely a small offset to Disney–ABC's business-stealing gain from internalization. Another real cost arises from the internalization because the distributor of programming probably has some useful experience for assessing the prospect of new entertainment investments. Dealing at arm's-length with content providers, the distributor might create some value by occasionally warding off a turkey-in-the-making. If Disney instructs ABC to take program series produced by Disney, this critical input is stifled unless the distribution arm can preserve considerable bureaucratic autonomy in the face of the basic incentive to steal business from rival content providers.<sup>27</sup>

#### 4. Role of non-profit enterprise<sup>28</sup>

An important aspect of industrial organization in the U.S. creative industries is the incidence of non-profit organizations (NPOs). They dominate the performing arts (except for Broadway theater) and share dominance of "cultural storage" activities – museums, libraries – with public-sector firms. Substantial fringes of NPOs appear in other sectors such as publishing, music recording, and broadcasting. This pattern invites an explanation why NPOs appear where they do. Two distinctive features of creative industries, interacting with each other, seem to offer one. The first is the combination of high fixed and low marginal costs that prevails in many creative activities. The second is the prevalence of *art-for-art's-sake* tastes, particularly those pertaining to the quality and variety of creative goods.

##### 4.1. Non-profits and governance in creative industries

High fixed costs pose a problem for any industry if sellers are constrained to charging a single linear price. In order to cover average costs, the equilibrium price must substantially exceed marginal cost, even when large numbers of sellers compete. Indeed, if fixed costs are sunk at the outset in a two-stage market game, the maximum viable number of sellers may be limited to the few that can cooperate and avoid price competition in the second stage (after the fixed costs are sunk). The gap between price and marginal cost inflicts a deadweight welfare loss, and high-fixed-cost industries have an incentive to avert it by means of non-linear prices or price discrimination. Non-linear pricing involves charging each customer a combination of fixed and variable prices, with the variable component ideally equal to marginal cost and the fixed component

<sup>27</sup> See Caves (2005, Chapter 7) on the consequences of removing regulations that limited the internalized ownership of TV program series.

<sup>28</sup> For further discussion of this topic see Brooks' Chapter 15 in this volume.



high enough that total revenues cover total costs but without excluding any customer willing to buy at marginal cost. In addition, or instead, the enterprise may be able to align prices with customers' diverse levels of willingness to pay. This might be done by allowing customers to choose different bundles of services (à la carte admission fees or an annual membership – second-degree price discrimination). It might be done by charging prices aligned to individuals' differing levels of willingness to pay (discounts for seniors and students – third-degree price discrimination). All these devices of course turn up in many markets dominated by profit-seeking enterprises. What is it about creative industries that pushes them toward NPO status when high fixed costs are present?

The answer seems to lie in the difficulty in creative industries of managing the contracts that implement non-linear prices in an enterprise run by a profit-seeking manager. The fixed charge (membership fee, say) must be committed at the start of the season before the product (performance) is actually on display. The consumer's willingness to prepay depends on the expected variety and quality of performances to be offered during the coming season. The list of plays to be performed might be contractible, but not the myriad stylistic and quality-related choices involved in presenting them. What then keeps the manager from promising high quality, offering lower quality, and pocketing the profit? The manager's *art-for-art's-sake* tastes may come to the rescue, if the organization is non-profit and the manager can credibly display a preference for high-quality and innovative performances. Moral hazard is then kept at bay. This mechanism may also facilitate the contract between the manager and performing artists, whose own *art-for-art's-sake* tastes make them willing to sacrifice pecuniary compensation for the opportunity to reach for artistic innovation and excellence. Their employment precommitment may turn on the manager's shared tastes, like the audience members' season tickets.<sup>29</sup>

The NPO is thus hypothesized to succeed because the manager's publicly espoused tastes mitigate the governance problem facing consumers, when they are asked to cover the fixed cost now and enjoy the performance later. The mechanism seems even more persuasively coherent when we think of the ongoing NPO in which the manager mounts this season's performances and then passes the hat for contributions as well as requesting membership renewals. Audience members' enthusiasm for the preceding season's offerings presumably determines not only the incidence of renewal but also the generosity of donations.

#### 4.2. Two-part pricing in practice: The donor-supported non-profit organization

The specific form of organization that provides the empirical embodiment of the NPO is the donor supported NPO, characterized by a self-perpetuating board of directors.<sup>30</sup> Its

<sup>29</sup> Hansmann (1981) developed the theory behind this model. For an empirical study of the alignment between regional theater directors' values and outside interest groups, see, Voss, Cable and Voss (2000).

<sup>30</sup> Sociologists get credit for identifying this form [DiMaggio (1982, 1986)] and researching the mechanisms that make it effective [Ostrower (1996), Odendahl (1990)].

board consists chiefly of wealthy consumers of the NPO's services, whose board memberships reward past financial contributions but also carry the expectation of continuing support. These persons are not necessarily effective monitors or managers of what may be a large and complex organization, but they are strongly motivated to ensure the viability and success of the organization by a sort of "social equity" that they hold, and that yields a return of esteem when the organization performs well and loss of face when it does not. Studies of such NPO board members show that their fund-raising efforts and administrative service to the organization are important relative to their role as personal donors [Ostrower (1996)]. This institution arose in nineteenth-century America, and the symphony orchestras, museums, and other such cultural enterprises organized in this manner were markedly successful in delivering "quality" in the organization's product relative to the for-profit enterprises and cooperatives that preceded the NPOs.<sup>31</sup>

Research on the ecology of donor-supported NPOs confirms this model of governance and its implications for the population of these firms. The model implies that donations should function as a regular, planned component of the organization's resource intake, as distinguished from an emergency topping-up when the organization suffers a negative fiscal shock. Financial flows over time indeed do behave as if both direct box-office revenues and donations maintain a stable average relationship to the organization's costs, subject to random year-to-year shortfalls and surpluses. Negative shocks may well occasion special flurries of fund-raising, because the NPO generally has no equity capital to absorb such shocks (an endowment, of course, may perform this function). Other relevant research deals with variations in the population of NPOs from city to city. The weight of fixed costs for such organizations implies that no market can efficiently support a large number. When small and large metropolitan areas are compared, the number of NPOs in a given class tends to increase with city size, though less than proportionally, and where larger numbers of NPOs of a given type tend to persist, they are commonly differentiated in their functions, reflecting in turn the cultural diversity of the city in question [Blau (1986)]. Finally, the aggregate U.S. population of arts and culture NPOs has grown over time at rates responsive to major changes in tax policy and in the ecological niches opened by the growth of the national economy [Bowen et al. (1994)].

## 5. Conclusions

A standard research line in industrial economics seeks to explain the level of or changes in the number and size distribution of business units within an industry. Some of our conclusions follow in that tradition. Like others, the creative industries' structures tend

<sup>31</sup> Symphony orchestras in nineteenth-century America provide a particularly interesting case study. NPOs competed with and vanquished other organizational types, i.e. for-profit firms and cooperatives [Hart (1973); Shanet (1975)].

to be driven by the efficient scales with which creative goods are produced and distributed (very large for record labels and movie studios, small for art galleries). Firms tend to sort themselves into those focused on the distribution of creative goods (“promoters”) and those concerned with identifying and nurturing creative talents (“pickers”). Large enterprises also include the “entertainment conglomerates”, which seek synergistic gains that depend theoretically on quite special conditions; foreclosing others and avoiding foreclosure by them may be principal motives.

What distinguishes the organization of the creative industries, however, is the prevalence of distinctive types of contracts. These contracts govern collaborations between artists and other parties in arm’s-length relationships or within an enterprise. These contracts’ structures devolve from a few bedrock properties of creative work and creative products. Two broad types of contracts hold particular importance in the creative industries. Some amount to joint ventures in which artists and “humdrum” inputs (perhaps one of each, perhaps many) collaborate simultaneously to obtain some valuable output. When these production processes occur not simultaneously but in sequence, real option contracts pervasively govern the sequential steps. Option contracts can leave the artist an autonomous creative agent (pop musicians and record labels) or govern the use of their talents to an employment relationship (classic Hollywood studios).

Our interpretation of these contracts and patterns of enterprise organization invokes the self-interested economic actors adapting to the fundamental elements of tastes and technology that characterize creative industries. That statement might appear in any analysis of an industry’s organization. Indeed, it usually does. The creative industries are striking, though, for possessing specific properties of tastes and technologies – the axiomatic properties listed at the start of this chapter – that supply theoretically coherent explanations for the structures of contracts and organization of enterprises. It has not been feasible to frame these interpretations with formally testable hypotheses, but the many points of agreement between casual empirical evidence and predictions based on these basic properties are certainly encouraging.

Besides deal structures and enterprise populations, the organization of an industry also embraces the prevalent type of firm – in this case, the role of non-profit enterprises. They dominate a number of arts activities, apparently for two interrelated reasons. These activities incur high fixed but low marginal costs, pressing them to employ two-part prices and club arrangements to ensure the coverage of fixed costs. When product quality is endogenous, however, non-profit status may be necessary for the manager credibly to forswear degrading quality once the customers’ fixed payments are in hand. Non-profit organizations supported by donation streams thus enjoy functional advantages.

While the creative industries are no fecund source of data bases, they do provide many opportunities for further research. The following list is confined to points that received rather conjectural treatment in this survey – theoretical propositions that could change greatly when formally worked out, or readings of qualitative empirical evidence that might turn out to be faulty generalizations.

- The organization of agents' activities and their roles as matchmakers and intermediaries have been little studied. The existing literature on intermediation does not match up well with the empirical questions that arise in the creative industries.
- Sticky, conventionalized prices present a puzzle. Are they really sticky, or do "list" and "transaction" prices diverge? If the stickiness is real, do the theoretical conjectures offered here stand close examination?
- The recent wave of vertical integration in the "entertainment conglomerates" suggests issues of foreclosure, especially in industries with zero marginal costs, that have not been worked out.
- The talent guilds that prevail in the creative industries arose long ago in response to contract failures. However it is not clear how their objectives in the ongoing entertainment industries might be characterized.
- The literature of contract theory assigns great importance to renegotiation, and renegotiation (shading into repeated interactions) seems to play an important role in the ongoing balancing of equities in the creative industries. Can empirical evidence be developed and related to theory?

## References

- Baker, A.J. (1991). "A model of competition and monopoly in the record industry". *Journal of Cultural Economics* 5, 29–54.
- Baron, D.P., Besanko, D. (1999). "Informational alliances". *Review of Economic Studies* 66, 743–768.
- Belinfante, A., Johnson, R.L. (1982). "Competition, pricing and concentration in the U.S. recorded music industry". *Journal of Cultural Economics* 6, 11–24.
- Berry, S., Waldfogel, J. (2003). "Product quality and market size". Working Paper No. 9675. National Bureau of Economic Research.
- Biglaiser, G. (1993). "Middlemen as experts". *RAND Journal of Economics* 24, 212–223.
- Blau, J.R. (1986). "The elite arts, more or less *de rigueur*: A comparative analysis of metropolitan culture". *Social Forces* 64, 875–906.
- Bolton, P., Whinston, M.D. (1993). "Incomplete contracts, vertical integration, and supply assurance". *Review of Economic Studies* 60, 121–148.
- Bowen, W.G., Nygren, T.I., Turner, S.E., Duffy, E.A. (1994). *The Charitable Nonprofits: An Analysis of Institutional Dynamics and Characteristics*. Jossey-Bass, San Francisco.
- Caves, R.E. (2000). *Creative Industries: Contracts between Art and Commerce*. Harvard University Press, Cambridge, MA.
- Caves, R.E. (2005). *Switching Channels: Organization and Change in TV Broadcasting*. Harvard University Press, Cambridge, MA.
- Chisholm, D.C. (1997). "Profit-sharing versus fixed-payment contracts: Evidence from the motion-picture industry". *Journal of Law, Economics, and Organization* 13, 169–201.
- Christopherson, S. (1992). "The origins of fragmented bargaining power in entertainment media industries". In: *Proceedings of the Forty-fourth Annual Meeting*. Industrial Relations Research Association, pp. 10–17.
- Christopherson, S., Storper, M. (1989). "The effects of flexible specialization on industrial politics and the labor market". *Industrial and Labor Relations Review* 42, 333–347.
- Coase, R.H. (1937). "The nature of the firm". *Economica* 4, 386–405.
- Conant, M. (1960). *Antitrust in the Motion Picture Industry: Economic and Legal Analysis*. University of California Press, Berkeley, CA.

- Di Maggio, P.J. (1982). "Cultural entrepreneurship in nineteenth-century Boston: The creation of an organizational base for high culture in America". *Media, Culture, and Society* 4, 33–50.
- Di Maggio, P.J. (1986). *Nonprofit Enterprises in the Arts: Studies in Mission and Constraint*. Oxford University Press, New York.
- Donahue, S.M. (1987). *American Film Distribution: The Changing Marketplace*. UMI Research Press, Ann Arbor, MI.
- Faulkner, R.R. (1983). *Music on Demand: Composers and Careers in the Hollywood Film Industry*. Transaction Books, New Brunswick, NJ.
- Faulkner, R.R., Anderson, A.B. (1987). "Short-term projects and emergent careers". *American Journal of Sociology* 92, 879–909.
- Fee, C.E. (2002). "The costs of outside equity control: Evidence from motion picture financing decisions". *Journal of Business* 75, 681–711.
- Getzels, J.W., Csikszentmihalyi, M. (1976). *The Creative Vision: A Longitudinal Study of Problem Finding in Art*. Wiley Interscience, New York.
- Goldberg, V.P. (1997). "The net profits puzzle". *Columbia Law Review* 97, 524–550.
- Goldman, W. (1984). *Adventures in the Screen Trade: A Personal View of Hollywood and Screenwriting*. Warner Books, New York.
- Greco, A.N. (1997). *The Book Publishing Industry*. Allyn and Bacon, Needham Heights, MA.
- Hansmann, H. (1981). "Nonprofit enterprise in the performing arts". *Bell Journal of Economics* 12, 341–361.
- Hansmann, H., Kraakman, R. (1992). "Hands-tying contracts: Book publishing, venture capital financing, and secured debt". *Journal of Law, Economics, and Organization* 8, 628–655.
- Hart, O. (1989). "Incomplete contracts". In: Eatwell, J., et al. (Eds.), *Allocation, Information, and Markets*. Norton, New York.
- Hart, O., Moore, J. (1999). "Foundations of incomplete contracts". *Review of Economic Studies* 66, 115–138.
- Hart, P. (1973). *Orpheus in the New World*. Norton, New York.
- Hepburn, J. (1968). *The Author's Empty Purse and the Rise of the Literary Agent*. Oxford University Press, Oxford.
- Hvide, H.K., Heifetz, A. (2001). "Free-entry equilibrium in a market for certifiers". Working Paper. Norwegian School of Economics and Business.
- Jensen, R. (1994). *Marketing Modernism in Fin-de-siècle Europe*. Princeton University Press, Princeton, NJ.
- Kingsbury, H. (1988). *Music, Talent, and Performance: A Conservatory Cultural System*. Temple University Press, Philadelphia.
- Kremer, M. (1993). "The O-rings theory of economic development". *Quarterly Journal of Economics* 108, 551–575.
- Lazarus, P.N. (1985). *The Movie Producer: A Handbook for Producing and Picture-Making*. Barnes and Noble, New York.
- Lopes, P.D. (1992). "Innovation and diversity in the popular music industry, 1969 to 1990". *American Sociological Review* 57, 56–71.
- McDougal, D. (1998). *The Last Mogul: Lew Wasserman, MCA, and the Hidden History of Hollywood*. Crown, New York.
- Mezias, J.M., Mezias, S.J. (2000). "Resource positioning, the founding of specialist firms, and innovation: The American feature film industry". *Organization Science* 11, 306–322.
- Miller, L.J. (2003). "The rise and not-quite fall of the American book wholesaler". *Journal of Media Economics* 16, 797–820.
- Miller, D., Shamsie, J. (1996). "The resource-based view of the firm in two environments: The Hollywood film studios from 1936 to 1965". *Strategic Management Journal* 39, 519–543.
- Nöldeke, G., Schmidt, K.M. (1998). "Sequential investments and options to own". *RAND Journal of Economics* 29, 633–653.
- Odendahl, T. (1990). *Charity Begins at Home: Generosity and Self-interest among the Philanthropic Elite*. Basic Books, New York.
- Ordovery, J.A., Saloner, G., Salop, S.C. (1990). "Equilibrium vertical foreclosure". *American Economic Review* 80, 127–142.

- Ostrower, F. (1996). *Why the Wealthy Give: The Culture of Elite Philanthropy*. Princeton University Press, Princeton, NJ.
- Passman, D.S. (1994). *All You Need to Know about the Music Business*. Simon & Schuster, New York.
- Peterson, R.A., Berger, D.G. (1975). "Cycles in symbol production: The case of popular music". *American Sociological Review* 40, 158–173.
- Poggi, J. (1968). *Theater in America: The Impact of Economic Forces, 1870–1968*. Cornell University Press, Ithaca, NY.
- Powell, W.W. (1985). *Getting into Print: The Decision-Making Process in Scholarly Publishing*. University of Chicago Press, Chicago.
- Rogers, E.M., Larsen, J.K. (1964). *Silicon Valley Fever: Growth of High-tech Culture*. Basic Books, New York.
- Rosen, S. (1981). "The economics of superstars". *American Economic Review* 41, 845–858.
- Rosenberg, B., Harburg, E. (1993). *The Broadway Musical: Collaboration in Commerce and Art*. New York University Press, New York.
- Sanjek, R., Sanjek, D. (1991). *American Popular Music Business in the 20th Century*. Oxford University Press, New York.
- Shaffer, G. (1991). "Slotting allowances and resale price maintenance: A comparison of facilitating practices". *RAND Journal of Economics* 22, 120–135.
- Shanet, H. (1975). *Philharmonic: A History of New York's Orchestra*. Doubleday, Garden City, NY.
- Shapiro, C. (1986). "Exchange of cost information in oligopoly". *Review of Economic Studies* 53, 433–446.
- Slater, R. (1997). *Ovitz: The Inside Story of Hollywood's most Controversial Power Broker*. McGraw-Hill, New York.
- Spulber, D.F. (1999). *Market Microstructure: Intermediaries and the Theory of the Firm*. Cambridge University Press, Cambridge.
- Starkey, K., Barnatt, C., Tempest, S. (2000). "Beyond networks and hierarchies: Latent organizations in the U.K. television industry". *Organization Science* 11, 299–305.
- Stern, S.L., Schoenhaus, T. (1990). *Toyland: The High-stakes Game of the Toy Industry*. Contemporary Books, Chicago.
- Stigler, G.J., Becker, G.S. (1977). "De gustibus non est disputandum". *American Economic Review* 67, 76–90.
- Storper, M. (1989). "The transition to flexible specialisation in the U.S. film industry". *Cambridge Journal of Economics* 13, 273–305.
- Storper, M., Christopherson, S. (1987). "Flexible specialization and regional industry agglomeration: The case of the U.S. motion picture industry". *Annals of the Association of American Geographers* 77, 104–117.
- Stowe, D.W. (1994). *Swing Changes: Big-band Jazz in New Deal America*. Harvard University Press, Cambridge, MA.
- Towse, R. (1993). *Singers in the Marketplace: The Economics of the Singing Profession*. Edward Elgar, Cheltenham.
- Voss, G.B., Cable, D.M., Voss, Z.G. (2000). "Linking organizational values to relationships with external constituents: A study of nonprofit professional theatres". *Organization Science* 11, 330–347.
- Walker, L. (1964). *The Wonderful Era of the Great Dance Bands*. Howell-North, Berkeley, CA.
- Weinstein, M. (1998). "Profit-sharing contracts in Hollywood: Evolution and analysis". *Journal of Legal Studies* 27, 67–112.
- White, H.C., White, C.A. (1993). *Canvases and Careers: Institutional Change in the French Painting World*. University of Chicago Press, Chicago.
- Whitesides, T. (1981). *The Blockbuster Complex: Conglomerates, Show Business, and Book Publishing*. Wesleyan University Press, Middletown, CT.
- Wijnberg, N.M., Gemser, G. (2000). "Adding value to innovation: Impressionism and the transformation of the selection system in visual arts". *Organization Science* 11, 323–329.
- Wildman, S.S., Siwek, S.E. (1988). *International Trade in Films and TV Programs*. Ballinger, Cambridge, MA.
- Williamson, O.E. (1985). *The Economic Institutions of Capitalism*. Free Press, New York.

## THE MEDIA AND ADVERTISING: A TALE OF TWO-SIDED MARKETS

SIMON P. ANDERSON

*University of Virginia, USA*

JEAN J. GABSZEWICZ

*CORE, Université catholique de Louvain, Belgium*

### Contents

Abstract	568
Keywords	568
1. Introduction	569
2. Background	572
2.1. Conceptual background	572
2.2. Statistical background	574
2.3. Platform competition	579
3. The model ingredients	581
3.1. Viewers and readers	581
3.2. Advertisers	583
3.3. Platforms	586
4. Equilibrium	587
4.1. Short-run equilibrium with advertising	587
4.2. Short-run equilibrium with pricing	590
4.3. Short-run equilibrium with pricing and advertising	591
4.4. Long-run equilibrium (free entry)	593
5. Welfare analysis	594
6. Product selection: Choosing program type	596
7. Press concentration and advertising	599
8. Conclusions	603
Acknowledgement	606
Appendix A	606
References	610
Further reading	613

## Abstract

Media industries are important drivers of popular culture. A large fraction of leisure time is devoted to radio, magazines, newspapers, the Internet, and television (the illustrative example henceforth). Most advertising expenditures are incurred for these media. They are also mainly supported by advertising revenue. Early work stressed possible market failures in program duplication and catering to the Lowest Common Denominator, indicating lack of cultural diversity and quality. The business model for most media industries is underscored by advertisers' demand to reach prospective customers. This business model has important implications for performance in the market since viewer sovereignty is indirect. Viewers are attracted by programming, though they dislike the ads it carries, and advertisers want viewers as potential consumers. The two sides are coordinated by broadcasters (or "platforms") that choose ad levels and program types, and advertising finances the programming. Competition for viewers of the demographics most desired by advertisers implies that programming choices will be biased towards the tastes of those with such demographics. The ability to use subscription pricing may help improve performance by catering to the tastes of those otherwise under-represented, though higher full prices tend to favor broadcasters at the expense of viewers and advertisers. If advertising demand is weak, program equilibrium program selection may be too extreme as broadcasters strive to avoid ruinous subscription price competition, but strong advertising demand may lead to strong competition for viewers and hence minimum differentiation ("la pensée unique"). Markets (such as newspapers) with a high proportion of ad-lovers may be served only by monopoly due to a circulation spiral: advertisers want to place ads in the paper with most readers, but readers want to buy the paper with more ads.

## Keywords

advertising finance, two-sided markets, platform competition, pensee unique, circulation spiral

*JEL classification:* D43, L13, L82, M37, Z11



## 1. Introduction

Sociologists, political scientists, lawyers, historians, and philosophers all have their views about the media. The wide scientific interest in media reflects the growing importance of entertainment and communication in today's information society. Citizens in developed countries devote the lion's share of their leisure time to consuming mass media such as television and newspapers. It may not be too large a stretch of the imagination to say that leisure time use (job satisfaction aside and ignoring eating pleasure in those cultures with fine cuisine) determines much of the quality of life: by extension, the quality of life for many people is thus underpinned by the quality of the media!<sup>1</sup> In this respect the media industries, and the broadcasting industry in particular, take on an overall importance to the national well-being far beyond the dollar or euro magnitude of the sector in the national accounts.<sup>2</sup>

Much of today's popular culture derives from television programming. Children at school copy the actions and characters of their heroes seen on TV the evening before, adults retell jokes and rehash story lines, and the hairstyle of the leading lady in *Friends* becomes a topic of national debate. Media are also the source of news of current affairs and political actions. The way the news are presented can also shape public opinion and, by influencing citizens' voting behavior, can even establish or depose governments and presidents.

Surprisingly enough, the media were long ignored by economists, despite the fact that media content cannot exist without some physical medium (TV sets, newspapers, magazines) that is produced and exchanged in a market. Yet the media are not traditional products like butter, gasoline, or sugar. First, media firms (in most cases) produce and distribute a *public good*: one person's consumption of a media product does not diminish the ability of another to consume it (non-rivalrousness).<sup>3</sup> Second, media products in many countries are viewed as *merit goods*, a category of goods where the state makes a paternalistic judgment that consumption is "good". Such consumption is often encouraged by public spending (whereas "merit bads" are discouraged by taxes or regulations and restrictions). With merit goods, "public" evaluation is seen as different from the private one, so rejecting a purely individualistic view of consumer benefits. This stance derives from the fact that media constitute a powerful instrument of education whose nature and diversity considerably shape the collective values of society. Finally, most

<sup>1</sup> The average American watches over four hours of TV per day. In Japan, the figure is three hours and thirty minutes, and in Europe only slightly more. Subtracting hours of sleep, hours worked, hours commuting, and hours eating from the daily total of 24 hours we conclude that leisure time is mostly devoted to watching TV.

<sup>2</sup> The intrusion of American cultural values and icons into European homes through the television screen is one reason why many countries (such as France with the "exception culturelle") restrict non-local content of programming.

<sup>3</sup> Some media products also share the other property common to public goods, non-excludability, like free newspapers or television broadcasting. Other media products, like cable broadcasting or magazines, are excludable, see [Samuelson \(1964\)](#) for further discussion.

media companies finance their activities (at least partially) by *advertising*. Media firms need advertisers to make the production of media content worthwhile, while advertisers need media firms to make their products known to potential consumers.<sup>4</sup> Consequently, the media industry sells a joint product to two different categories of buyers: the medium itself to advertisers, and the medium *content* to media consumers (readers, TV-watchers, web-surfers, etc.).

Media firms thereby operate in two different industries and get their profits from both. From this two-sided interest, the cultural content offered to media consumers is shaped by the desire to offer advertisers a vehicle that reaches as many prospective consumers as possible: “when news sell ‘eyeballs’ to advertisers, the question becomes what content can attract readers or viewers rather than what value will consumers place on content” [Hamilton (2004)]. This potential bias in the type of programming or reading content offered may bias popular culture as well. The ads themselves are the subject of cult followings, and characters in ads may lead fashions and fads. The dollar amount spent on ads is the tip of a larger economic iceberg: insofar as new product introduction needs or is facilitated by advertising, product turnover and product generation is determined by ads. Some might say tastes too are influenced by ads. Ads can certainly create hype and fashions. Advertising also forms and reflects popular culture. It is important economically not only because of the fraction of GDP that it represents directly (around 2%) but also because it may facilitate the introduction of new products to market and so underscore a larger fraction of GDP.

Competition for advertising revenues therefore governs market performance; commercial television needs advertising revenue to survive (subscription pricing aside). Competition for advertising revenues therefore governs market performance. The willingness to pay of advertisers to contact viewers of particular demographics thus determines the type and range of programs offered in a free market system. This is very different from a traditional market structure where the principle of consumer sovereignty governs the type and range of products offered on the market. In conventional economic markets, consumers “vote” with their dollar purchasing power for the products they want, and firms, seeking profits, have the incentive to provide what consumers want. In the commercial television context, viewer sovereignty is filtered and muted. Viewers “vote” with their eyeballs for the programs they want to watch, and broadcasters need to deliver eyeballs to advertisers. However, different eyeballs get different vote weights in the sense that advertisers care about the type of viewers who are delivered – those most inclined to change their purchase behavior and buy copious quantities of the product on display are those of most interest to the advertiser. In addition to this type of distortion (whose consequences we elaborate upon below), media market performance can be sub-optimal for more subtle reasons even when all viewers are equally weighted

<sup>4</sup> The degree of advertising in media financing varies across media and countries. Public broadcasting services financed only by public subscription exist in England or Japan, while other media are fully financed by ads, like free newspapers and commercial TV broadcasting.

by advertisers. The reason stems from the particular market interaction inherent in the commercial television market, which forms a leading example of a “two-sided market” with network externalities.<sup>5</sup> In a two-sided market, two groups interact through an intermediary, or platform, that accounts for the externalities between the groups. In the media context, the platform is the broadcast company (or companies) and the two interacting groups are advertisers and viewers. Advertisers like more viewers to receive their messages. Viewers though find advertising a nuisance insofar as it detracts from time available to watch a program. The more advertisements are carried, the more the viewers are disappointed, so the former impart a negative externality on the latter. However, the viewers do not pay a direct price for the entertainment that they receive.

A similar structure governs commercial radio. Many Internet sites are also financed solely by advertising revenues from click-throughs and pop-up ads, which are also frequently a nuisance to surfers (at least, those who do not click through!). Magazines and newspapers are founded on a similar business model, and derive much of their revenue from the advertisements they carry. However, they also typically charge a direct price to their readers. This is true now for pay-per-view television, and for premium television shows too. Cable television, which involves a local service provider bundling together selections of channels, is an intermediary type of structure insofar as it typically carries to the household many programs that do carry ads themselves. The ability to price programming alters the market outcome by drawing in some direct competition for viewers.

The business model for newspapers and magazines has similar elements, although arguably advertisements are not as much of a nuisance as they are with television, radio, or web-pages.<sup>6</sup> Readers can skip past the ads without having to pay much attention to them, while they interrupt and postpone a television program. Readers may even find a positive net benefit from ads. This is especially true for classified ads in newspapers, and for products displayed in specialist magazines (motorcycles, golf, sailing, etc.). If readers do get positive net benefits, then the market interaction may be fundamentally different. If a medium attracts more readers or viewers, the more are advertisers willing to pay to get their messages across (this is true regardless of whether the readers or viewers are attracted to the messages per se). When readers want to get ad exposure (“ad-loving” behavior), then the market may loosely be described in terms of a “positive spiral”.<sup>7</sup> That is, the more readers there are, the more advertisers want to advertise in the paper or magazine, but then the more readers want to subscribe to it. This reinforcing effect may mean that only a monopoly can survive in the market. This conclusion

<sup>5</sup> Although most two-sided markets studied in the literature involve bilateral positive externalities, broadcasting instead typically involves negative externalities to viewers from advertisers and positive externalities on advertisers from the number of viewers.

<sup>6</sup> The existence of “Infomercials” on television indicates that advertising is not a nuisance to all viewers, too.

<sup>7</sup> Modeling this can be quite intricate. Caillaud and Jullien (2001) note that they “attempt to capture a fundamentally dynamic process by way of a static model, hence some imperfection”.

though ought to be tempered if there is product differentiation (so that several different types of magazine can survive, offering different specialties, or newspapers may provide different political viewpoints). Another caveat here concerns whether advertisers can reach readers through different media, and whether advertisers tout their wares in several papers or magazines. These issues are discussed further below.

In what follows we shall refer to the television context, and speak for the most part of viewers who watch broadcasts on channels. Differences for other media are pointed out where pertinent.

## 2. Background

We first present some conceptual background, and then some statistical background. This is followed by a description of the basic two-sided market paradigm, as applied to media markets.

### 2.1. Conceptual background

Perhaps the earliest model of television program choice is due to [Steiner \(1952\)](#). Steiner assumed simply that viewers will watch the (single) program type they prefer, and that different viewers have different preferences.<sup>8</sup> To take an example, suppose that 67% of the population will only watch game shows, and the rest only will watch sports. Then if there are two channels operated by competing firms, they will both offer game shows and so divide the larger pool of viewers. This is the Principle of Duplication, and is arguably prevalent on afternoon and prime-time network television. It implies that the market system does not cater to the minority taste. A monopoly though, with two channels, would not cannibalize its own game audience by providing a second game show, but would instead provide a sports show and then cater to the whole market. Implicit in the above description is that television broadcasters wish to maximize viewers. This makes sense when viewers do not mind ads, ads are sold at a fixed price per ad per viewer, and there is a binding cap on ad levels (as in the E.U. currently). Otherwise, and as we develop in the models below, broadcasters need to worry about viewers switching over or off, and extracting advertising revenues optimally.

A similar idea to Steiner's Duplication Principle is arrived at with a different variant of the model. Suppose [following [Hotelling \(1929\)](#)] that viewers' ideal tastes are distributed along a unit interval. Each viewer watches the channel closest to her ideal taste point. There are two broadcasters who choose "locations" in the unit interval, with the objective purely of maximizing own viewership. Then the equilibrium is what [Boulding](#)

<sup>8</sup> See [Cabizza \(2004\)](#) for a model with a similar preference structure. Her paper addresses the extent that programs cater to minority tastes under private or public broadcasting, and in a mixed system. She also notes that, in addition to [Steiner \(1952\)](#), [Rothenberg \(1962\)](#) and [Wiles \(1963\)](#) indicate the tendency for duplication of program types that attract large audiences.

(1955) christened the Principle of Minimum Differentiation. Both broadcasters choose exactly the same program type and split the market, just as in Steiner's analysis.<sup>9</sup>

An alternative specification of the program scheduling problem is formulated by Cancian, Angela and Bergstrom (1995).<sup>10</sup> These authors consider two TV channels that must decide (non-cooperatively) when to broadcast their evening television news. Viewers prefer to watch the news as soon as they get home from work. The times when viewers get home are distributed on an interval of time. Broadcasters strive to maximize audience size, and each is to choose a broadcast time. This game has no pure strategy Nash equilibrium. Indeed, whenever its opponent chooses a broadcasting time past the median of the distribution, each network's best response is to broadcast its show just before its competitor's to get over half the viewers. Its best reply when its competitor's expected broadcasting time is before the median is to choose the latest possible time and so again get over half the viewers.<sup>11</sup>

A second early concept that still resonates today is that of the Lowest Common Denominator (LCD), proposed in this context by Beebe (1977). Beebe took issue with Steiner's assumption that viewers will not watch if they are not offered their most preferred program type – and hence took issue with Steiner's conclusion that monopoly outperforms competition in terms of catering to diverse tastes. Suppose for illustration that viewers have diverse first preferences, but all would watch a game show if nothing else were available. Then a monopoly would have no reason to offer more than one program, and it would air a game show. This is, by construction, the LCD program type. Competing broadcasters though would offer different program types in order to attract viewers from rivals.<sup>12</sup>

These basic analyses are important as far as they go, but they miss the crucial tension in the market. In these models, viewers are not deterred by ads, and advertisers have the same willingness to pay for communicating with viewers. The important insight from the economics of platform competition is that the platform (broadcaster) needs to get both sides of the market on board – viewers must be delivered to advertisers, and advertisers are the direct revenue. How much they are willing to pay depends on the number of viewers delivered on the other side of the market.

<sup>9</sup> See Eaton and Lipsey (1975) for an extension to many firms, a consideration of non-uniform consumer densities, and other extensions.

<sup>10</sup> See also Nilssen and Sjørgard (1998).

<sup>11</sup> Gabszewicz, Laussel and Sonnac (2004) analyze an extension of the basic Hotelling game with single-homing advertisers and competition for viewers who dislike ads. Surprisingly, this extension also leads to non-existence of a pure strategy equilibrium, albeit in a more complex (two-stage) game where firms choose broadcast times and then ad levels.

<sup>12</sup> Beebe (1977) presents several numerical examples of group sizes and preference structures to determine equilibrium offerings under competition and under multi-channel monopoly. He does so for both a fixed number of channels, and for an endogenous number of channels determined by fixed costs of airing a channel.

## 2.2. Statistical background

This sub-section substantiates the view that media consumption takes up a lot of available leisure time, by providing some data about the time spent by consumers (readers or viewers) with various media. It also indicates the economic importance of advertising in the US as a fraction of total GDP and as a function of medium type, and shows that performance concerns due to large amounts of advertising might be quite well-founded. A break-down of what advertising time on various media is worth to advertisers is given later in the text.

Table 1 shows how much time is spent by households in the US watching television. The table documents the rise in the importance of television watching over the last 50 years. The current household (not individual!) average hours watched is an astounding eight per day. Arguably this rise (from four and a half in the 1950's) is due to habit changes and technology changes (such as cheaper television sets). The 1970's and 1980's saw households owning multiple sets, as well as the advent of color televisions. In the 1990's, the set of program options (including many 24 hour programming options) increased immensely with the increased popularity of cable, satellite, etc.

Individual watching rates are quite a lot lower than the household rates, but still around 4 hours a day in the US (a detailed break-down by medium type is given in

Table 1  
Time spent viewing per household, US

Year	Time spent per day
1950	4 hrs. 35 mins.
1955	4 hrs. 51 mins.
1960	5 hrs. 06 mins.
1965	5 hrs. 29 mins.
1970	5 hrs. 56 mins.
1975	6 hrs. 07 mins.
1980	6 hrs. 36 mins.
1985	7 hrs. 10 mins.
1990	6 hrs. 53 mins.
1995	7 hrs. 17 mins.
1996	7 hrs. 11 mins.
1997	7 hrs. 12 mins.
1998	7 hrs. 15 mins.
1999	7 hrs. 26 mins.
2000	7 hrs. 35 mins.
2001	7 hrs. 40 mins.
2002	7 hrs. 44 mins.
2003	7 hrs. 58 mins.

Source: [www.tvb.org](http://www.tvb.org), based on data from Nielsen.

Table 2  
Average annual hours spent for TV and work, 1994–1997

Country	TV-viewing hours per adult	Work hours per person in employment	Work hours per adult
Norway	878	1413	850
Switzerland	882	1608	1036
Sweden	893	1600	981
Finland	929	1725	1043
Germany	1148	1563	903
France	1166	1616	878
Japan	1324	1885	1116
Spain	1334	1813	903
Italy	1340	1637	761
UK	1387	1738	1065
USA	1462	1954	1264

Source: Corneo (2001).

Table 2). While the US rate is the highest in the world, some other countries come close. Corneo (2001) presents evidence that people spend roughly the same amount of time working as they do watching television so there is a positive correlation across countries.<sup>13</sup> Surprisingly, Norway has the lowest number of hours watching (60% of the US figure).<sup>14</sup> On average, people spend only 30% more time working than they do watching TV.

Corneo estimates an OLS regression of the correlation between hours worked per adult ( $W$ ) and hours watching television ( $TV$ ). The estimated equation is  $TV = -614 + 1.05 W$ . Both the constant and the coefficient are statistically significant, and the  $R^2$  value is .51.

Table 3 provides a breakdown across media (television, newspapers, radio, magazines, and the Internet) of time spent. The importance of demographics to advertisers is implicit in this table, given the break-out of occupations, income, and education levels.<sup>15</sup> Poorer people tend to watch more television, as do the retired. Magazine and Internet use is highest among richer individuals.

<sup>13</sup> To explain this, Corneo (2001) develops a simple model in which adults choose between 3 activities, work, TV watching, and “socially enjoyed leisure” (activities enjoyed with others). To explain the positive correlation, Corneo invokes multiplicity of (Pareto-ranked) equilibria. Given that others are working long hours, it does not pay an individual to invest greatly in social ties. In a related vein, Rogerson (2005) explains the big difference in continental European (France, Germany and Italy) hours worked as due to different tax rates: the Europeans consequently indulged in more “home production”.

<sup>14</sup> Surprising at least, because one might imagine that long winter nights would be spent watching television.

<sup>15</sup> Wildman, McCulloch and Kieshnick (2004) show empirically that implicit prices for access to different individual types in a program’s viewership have different prices. See also Goettler (1999) and Wilbur (2004b).

Table 3  
Adults: Time spent yesterday in minutes with major media

Adults	Television	Newspapers	Radio	Magazines	Internet
<i>Age</i>					
18+	258.4	32.4	120.7	18.3	65.8
18–34	236	16.4	141.3	16.9	71.4
18–49	234.3	23.7	131.2	15.6	79.5
25–49	234.8	26	132.2	13.2	84.6
25–54	239.4	27.2	132.6	14.2	85.3
35–64	254.6	34.1	116.3	16.3	76.9
65+	317.1	58.1	94.9	27.3	18.6
<i>Household income</i>					
Under \$25K	318.7	27.6	101.7	18	30.3
\$25–50K	277.7	31.9	139.5	18.8	81.4
\$50–75K	234.5	24.6	129.6	12.3	56
\$75–100K	212.6	42.1	111	17.5	92
\$100K+	203	39.7	91.4	19.2	79.5
<i>Education</i>					
HS Grad	287.8	27.9	133.3	16.9	44
Some college	273.9	30.5	134.9	19.4	74.9
College grad+	204.7	39.6	109	17.4	88.6
<i>Occupation</i>					
Prof/Tech/Mgr/Owner	199.8	31.2	134.3	18.1	101.1
Admin/Clerical/Sales	238.5	24.7	145	16.7	69.2
Trade/Service	223.8	21.2	148.7	18.1	39.4

Source: [www.tvb.org](http://www.tvb.org), based on data from Nielsen.

To get the data in Table 3 (which pertain to January 2003), adults were asked about their prior day's usage ("yesterday"). The philosophy behind this methodology is that people remember well what they did on the previous day. Interviews were evenly conducted over a 2-week period, so that the data aggregates week-end and week-day figures (see [www.tvb.org](http://www.tvb.org) for more details on the methodology).

The cost of accessing attention differs quite substantially across media. More detailed data are given in Table 4. For instance, the current rate for a 30 second commercial on prime-time translates into a rate of 2 cents per household reached. For spot television commercials, the rate is somewhat higher, at 2.7 cents per home reached. For comparison, the newspaper rate for the year 2000 works out at nearly seven cents per home for a half-page advert. We turn now to the share of advertising expenditures in the economy, and how much of this is on TV. The share of advertising in GDP has remained roughly 2% over the last forty years.

A break-down in terms of advertising volume across media (also from [www.tvb.org](http://www.tvb.org)) is given in Table 5. The last year (2003) is broken out into shares and the percentage change over the previous year. Television (summing broadcast and cable) accounts for



Table 4  
Gross domestic product, total ad volume, and television ad volume 1960–2003

Year	GDP*	Ad volume		TV ad volume	
	\$Billions	\$Millions	GDP (%)	\$Millions	Ad volume (%)
1963	617.7	13,100	2.12	2032	15.50
1973	1382.70	24,980	1.81	4460	17.90
1983	3536.70	76,000	2.15	16,879	22.20
1993	6657.40	140,956	2.12	32,471	23.00
2003	10,987.90	245,477	2.23	60,746	24.70

\*Sources: GDP, [bea.doc.gov](http://bea.doc.gov); ad volume and TV ad volume [www.tvb.org](http://www.tvb.org), based on data from Nielsen.

about a quarter of the total spending, and, using the data from the previous table, therefore accounts for about half a percent of GDP on its own.

Finally, the amount of advertising is also a performance concern. In broadcast media, especially, ads are hard to avoid Shields (2004) reports a study by MindShare for 2003 “ad clutter” meaning non-program (commercials, promos, PSAs, etc.) minutes. The data are recorded in Table 6. ABC, NBC, and Fox all passed 15 mins./hr.; and CBS has increased its clutter, but is still below the 15-minute mark at 14:18 minutes.<sup>16</sup> All are trending upward, as are Cable networks, though they generally carry less clutter. At one extreme, MTV carries 15:25 minutes every hour, while ESPN carries “only” 11:48 minutes of non-programming. By contrast, the EU has advertising caps that restrict the level of advertising to 9 minutes per hour. This level is very similar to the 9.5 non-programming minutes per hour that were standard in the US twenty years ago. This amount was a limit on commercials that was agreed upon by the National Association of Broadcasters and maintained by a voluntary code.<sup>17</sup> There is evidence the actual programming is being subverted with messaging too.

Shields suggests that total viewer demand for TV is pretty inelastic: “overall TV viewing is not that different [over the last decade] suggesting that clutter isn’t driving viewers from the set, but may be causing them to flip [channels]” and notes that network ratings have dropped for prime time programming as fragmentation spreads viewership over hundreds of channels.

<sup>16</sup> Some popular programs are among those with the highest clutter (for the fourth quarter of 2003). These include “The Bachelor” (18:08 minutes), “My Wife and Kids” (17:40), “Everybody Loves Raymond” (16:15), “Survivor: Pearl Islands” (16:05), and “Friends” (16:06) [source: Shields (2004)]. However, daytime television carries even more clutter than prime-time. In November 2001, NBC’s “Days of our lives” carried 23:23 minutes per hour, ABC’s “All of our children” carried 22:59 minutes, and ABC’s “General hospital” had 22:31 minutes. These figures, and much further interesting data on clutter, can be found in the American Association of Advertising Agencies and the Association of National Advertisers, Inc. joint 2001 Television Commercial Monitoring Report.

<sup>17</sup> Unfortunately (and interestingly), Nielsen does not publish data on advertisement ratings (number of people watching the ads) even though it has the technology to do so.

Table 5  
Estimated annual US advertising expenditures (in millions of dollars)

	1953 (\$ m.)	1973 (\$ m.)	1983 (\$ m.)	1993 (\$ m.)	2003 (\$ m.)	2003 (%)
<i>Newspapers:</i>	2632	7481	20,582	32,025	44,843	18.3
National	606	1049	2734	3620	7357	3
Local	2026	6432	17,848	28,405	37,486	15.3
<i>Magazines:</i>	627	1448	4233	7357	11,435	4.7
Weeklies	351	583	1917			
Women's	158	362	1056			
Monthlies	118	503	1260			
Farm publications	71	65	163			
<i>Broadcast TV:</i>	606	4460	16,879	28,020	41,932	17.1
Network	320	1968	6955	10,209	15,030	6.1
Spot (nat'l)	145	1377	4827	7800	9948	4.1
Spot (local)	141	1115	4345	8435	13,520	5.5
Syndication			300	1576	3434	1.4
<i>Cable:</i>			452	4451	18,814	7.7
Cable network			376	3295	13,954	5.7
Cable (non-net)			76	1156	4860	2
<i>Radio:</i>	611	1723	5210	9457	19,100	7.7
Network	141	68	296	458	798	0.3
Spot (nat'l)	146	400	1038	1657	3540	1.4
Spot (local)	324	1255	3876	7342	14,762	6
<i>Yellow pages:</i>			4400	9517	13,896	5.7
National			489	1230	2114	0.9
Local			3911	8287	11,782	4.8
<i>Direct mail</i>	1099	3698	11,795	27,266	48,370	19.7
<i>Business papers</i>	395	865	1990	3260	4004	1.6
<i>Out of home:</i>	176	308	794	1090	5443	2.2
National	119	200	512	605	2298	0.9
Local	57	108	282	485	3145	1.3
<i>Internet</i>				0	5650	2.3
<i>Miscellaneous:</i>	1523	4932	9954	18,513	31,990	13
National	846	2562	6952	13,534	24,550	10
Local	677	2370	3002	4979	7440	3
<i>Total National</i>	4515	13,700	42,660	81,867	152,482	62.1
<i>Total Local</i>	3225	11,280	33,340	59,089	92,995	37.9
<i>Grand Total</i>	7740	23,210	76,000	140,956	245,477	100

Source: [www.tvb.org](http://www.tvb.org), based on data from Nielsen.

Table 6  
2003 prime-time clutter (minutes:seconds)

	Network commercial minutes		Non-program minutes	
	2002	2003	2002	2003
ABC	10:15	10:15	15:16	15:31
CBS	9:03	9:19	14:06	14:18
Fox	9:04	9:11	14:47	15:13
NBC	9:41	9:19	14:49	15:07

Source: [www.tvb.org](http://www.tvb.org), based on data from Nielsen.

2.3. Platform competition

The key to the basic market model is that advertisers want to reach viewers, but viewers dislike ads,<sup>18</sup> and the size of each of these two segments matters both for the equilibrium arrangement and the optimum. The platform, or intermediary, is the broadcast company (or companies) that renders the ads palatable by bundling them with programs that are the viewers’ ultimate objective. That is, entertainment is provided free of a direct price, and this sugar-coats the consumption of ads the prospective consumer would otherwise not choose to watch. However, the platform recognizes the trade-off between higher ad levels that lead to more revenue per viewer, and the loss in viewer base from ramping up ad levels too high. Thus the platform has to coordinate the two sides of the market to get them both on board in the numbers that maximize revenue, and recognizing how both sides benefit or suffer from the interaction. When the market structure has more than one platform, competition from other platforms must also be factored into each platform’s calculus.

The economics of two-sided markets was developed after researchers into credit card markets recognized that this is a market not immediately amenable to traditional analysis.<sup>19</sup> Instead, credit cards and other prominent examples are two-sided in the sense that the benefits one receives on one side of the market depend not directly on the number of other agents on one’s own side, but rather on the number of agents on the other side. Thus the benefits to a shopper from holding a card depend on the number of stores that take it, and the benefits to a store from taking a card depend on the number of

<sup>18</sup> In the ad-loving variant, they do like ads.

<sup>19</sup> Indeed, even though it was sometimes suggested that credit cards constitute a market with network externalities, the prevailing model of network externalities at the time was one in which these externalities were “one-sided”. That is, a consumer’s benefit from carrying a card depends on the number of other card holders. Some reflection suggests that this is true to the extent that more shops are likely to take a card if more prospective consumers carry it. However, this mechanism ought to be modeled directly. The one-sided prototype might fit well such networks as fax machines or computer software [see Economides (1996), for a review].

consumers who carry it. The credit card example has two-sided positive externalities. Here the card-issuing company is the platform, the intermediary that coordinates the two sides of the market.

The economics of two-sided markets were pioneered by [Caillaud and Jullien \(2001, 2003\)](#) and [Rochet and Tirole \(2003\)](#), and further synthesized and extended by [Armstrong \(2004\)](#) and [Rochet and Tirole \(2004\)](#). The latter authors address the tricky task of defining two-sided markets.<sup>20</sup> [Wright \(2003\)](#) provides a useful service in indicating several examples of fallacious reasoning that would be ill-inspired from thinking about traditional markets in a context that was actually two sided. For example, one might think pricing below marginal cost would be indicative of predatory intent. However, in a two sided market, such pricing is quite natural, and stems from the need to get one side on board in order to extract surplus from the other side. In the TV context, viewers watch for “free” but advertisers pay for access. Indeed, [Armstrong](#) suggests that creates more benefits for the other side is the one that will enjoy low prices (for joining the platform). [Wright](#) though warns against thinking of this as a cross-subsidy from one side to the other, pointing out that with traditional subsidies, the side “paying” would prefer that the other were eliminated, along with the implicit tax. In the television context, the advertisers definitely would NOT like to see the viewers barred from the market! Rather, it is the low (or zero) price that attracts the viewers and therefore provides the surplus to the advertisers.<sup>21</sup>

[Caillaud and Jullien \(2003\)](#) also allow for “multi-homing”,<sup>22</sup> meaning that sometimes agents from one side of the market may use more than one platform – men (or indeed women) could use two different dating services, say. In the television context below, [Anderson and Coate \(2005\)](#) have multi-homing by advertisers, meaning that advertisers can place ads on several channels, while [Ferrando et al. \(2004\)](#) have single-homing.<sup>23</sup>

<sup>20</sup> The typical firm must get both worker and consumer sides “on board” in the sense of coordinating different agent groups, but this should not be considered a two-sided market problem. As [Armstrong \(2004\)](#) notes, “agents from one group generally do not care how well the firm performs in the market for the other group, but only about their own terms for dealing with the platform”. Two-sided markets also involve cross-group network effects absent in the simple firm context.

<sup>21</sup> In a similar vein, the earlier papers by [Caillaud and Jullien \(2001, 2003\)](#) situate the problem as competing “cybermediaries” (internet matchmakers) that coordinate groups of agents that wish to transact. The “Chicken and Egg” problem to which they refer alludes to getting both sides on board. In the simplest variants of the models, there are just entrance fees to the parties who may then interact. In a more complex version, agents transact if they find a match (which they do with an exogenous probability), and the platform can charge a price on that transaction too. This would be analogous to a royalty on sales following an ad on TV, which is an arrangement not seen in practice in ad markets.

<sup>22</sup> The term comes from usage on the Internet, meaning to have more than one Internet Service Provider.

<sup>23</sup> [Gabszewicz and Wauthy \(2004\)](#) contribute further to the debate.

### 3. The model ingredients

Going beyond the powerful, but rather rudimentary analyses of Steiner (1952) and Beebe (1977) means getting more explicit about the tastes and objectives of the three principal groups of agents who interact in the market. For concreteness, consider the case of television. The agents are the viewers, the advertisers, and the broadcast companies who have the central role of coordinating the two sides of the market. We describe these in turn.

#### 3.1. Viewers and readers

The model is built up from the basic ingredients that form its micro underpinnings. This means describing the tastes of diverse individuals to then generate their choices. We then aggregate up over individuals to find the viewership of each program type offered.<sup>24</sup>

Viewers make a discrete choice of which station to watch. At any moment, a viewer can reasonably only watch one station. Although a viewer might switch channels over a given hour (and we describe below how to allow such behavior), our starting point is to have viewers watch a single channel. We must also modify the model in the case of newspapers for which it is conceivable that a reader may buy and read several alternative papers: a fortiori for magazines (multi-homing by readers).<sup>25</sup> Table 3 above showed a break-down across media of time spent by individuals. In parallel, Table 7 shows the fraction of adults that each medium reaches (i.e., it gives a break-down by category of how many individuals are accessing the medium).

A simple way to model tastes is to suppose that each viewer has a conditional utility function for each option  $i$ , and this utility depends on the *match value*, which is the intrinsic benefit of entertainment, and may differ across viewers. From this we subtract the full price paid for the option. The full price consists of the monetary (or subscription) price,  $s_i$ , from watching channel  $i = 1, \dots, n$ , plus any nuisance from advertising. Supposing the advertising level is  $a_i$ , the simplest way to capture the nuisance cost is to

<sup>24</sup> Of course, this is not the only way to proceed. There is a long-standing tradition in Industrial Organization (and more recently on Macroeconomics) to use representative consumer models to portray the aggregate taste. These have also been used in Media Economics: see Barros et al. (2005) for an application to the Internet and vertical integration. Cunningham and Alexander (2004) study an equilibrium model and find that greater concentration (the inverse of the number of firms broadcasting) may decrease the total amount of programming broadcast, and a decrease in consumer welfare.

<sup>25</sup> Models in which viewers are assumed to mix between channels include Gal-Or and Dukes (2003) and Peitz and Valletti (2005). Anderson and Neven (1989) analyze the welfare properties of such a set-up in the context of product differentiation. Indeed, while without mixing the socially optimal locations are the quartiles of a linear location space, with mixing the optimal locations are the extreme ones. The positive analysis is the same if there is a linear likelihood of buying a product after seeing an ad; when it comes to the pricing analysis, the analogy is to a model of pay-per-view rather than flat subscription pricing. Another interesting issue that arises is that mixing viewers may be reached on two different channels, thus eroding the monopoly bottleneck that channels have over delivering viewers.

Table 7  
Adults reached yesterday by major media (%)

Adults	Television	Newspapers	Radio	Magazines	Internet
<i>Age</i>					
18+	90	65.2	72.8	48	51.1
18–34	87.6	48.5	80.1	48.1	55.6
18–49	88.5	58.7	79.9	48	58.1
25–49	89.8	64	81.1	45.1	57.3
25–54	90.5	65.9	80.2	47.7	58.9
35–64	91.5	71.4	74.8	49.9	57.9
65+	89.9	77.5	51.9	41.5	19.9
<i>Household Income</i>					
Under \$25K	87.3	57.4	59.7	40.2	27.8
\$25–50K	91.6	66.3	72.6	48.6	48.9
\$50–75K	89.8	66.9	77.1	45.6	56.5
\$75K+	89.8	71.2	83	55.2	74.8
\$100K+	92.2	72.5	85.5	61.2	75.9
<i>Education</i>					
HS Grad	90	60.9	69.4	42.3	35.5
Some college	92.3	67.2	73.6	51.4	54.3
College grad+	89.4	72.2	80.7	53	71.9
<i>Occupation</i>					
Prof/Tech/Mgr/Owner	89.6	65.3	84.8	54.1	73.5
Admin/Clerical/Sales	88.7	64.2	79.2	48.2	47.4
Trade/Service	85.8	65.3	78.7	52.4	38.1

Source: [www.tvb.org](http://www.tvb.org), based on data from Nielsen.

assume it is linear in the advertising level, at a rate  $\gamma$  per ad. This nuisance rate may be constant across the population, or may differ across different viewers. We also deduct from the full price any expected surplus the consumer may expect from trades inspired from the ads seen. Loosely, such surplus serves to reduce the effective  $\gamma$ , and may even render it negative. Such might be the case with classified ads for which the consumer actively searches out information and so advertising provides a positive net benefit. In the sequel we shall assume that nuisance costs are the same for all viewers, and that viewers expect no surplus from the goods they see advertised. We denote the full price of option  $i$  as  $f_i$  and it is thus given by

$$f_i = s_i + \gamma a_i. \quad (1)$$

The match utility is inspired from the standard stable of discrete choice models of product differentiation. From models of vertical (or quality) differentiation [Mussa and Rosen (1978), Gabszewicz and Thisse (1979), Shaked and Sutton (1982)] we draw a taste for a measure of quality,  $\theta q_i$ , with  $\theta$  an individual specific marginal willingness to pay for quality, and  $q_i$  the quality of option  $i$ . From models of spatial competition we draw the distance disutility  $\tau(\cdot)$  that measures how disappointed is the consumer from

not getting her ideal horizontal product specification. This depends on the “distance” between the viewer (at  $x$ ) and the program offering (at  $x_i$ ). From models of probabilistic discrete choice, we draw an alternative way to conceptualize horizontal differentiation, via match values  $\mu\varepsilon_i$ . These are typically assumed independently and identically distributed across consumers, and so (if the other potential sources of differentiation are not present), products are symmetric substitutes and competition is “global” in the sense that each product competes symmetrically with each other one. This is to be compared to the “local” competition inherent in the spatial model: each product competes directly with only its two neighbors.

In sum, the utility of consumer with preference draws  $\{\theta, x, \varepsilon\}$  (i.e., located at  $x$  and buying from a firm “located” at  $x_i$  with quality  $q_i$  and setting price  $s_i$  with ad level  $a_i$ ) then becomes:

$$u_i = y - [s_i + \gamma a_i] + \theta q_i - \tau(|x - x_i|) + \mu\varepsilon_i, \quad i = 1, \dots, n, \quad (2)$$

where the term in square brackets is the full price (i.e.,  $f_i$ : recall Equation (1)) and  $y$  is consumer income (which we suppose is the same across all consumers since it anyway plays no role in the choice model). In the sequel, we shall typically only deal with one type of differentiation in Equation (2) at a time, and the others will be suppressed.

### 3.2. Advertisers

The Economics of Advertising are quite controversial when it comes to the normative analysis. This is because successful advertising shifts demand, and therefore (presumably) consumer surplus. A comprehensive survey of the Economics of Advertising is provided in [Bagwell \(2003\)](#), and some salient points are discussed following the presentation of the model used here.

The simplest formulation for advertiser demand is that it is perfectly elastic. This means that there is no producer surplus to worry about. This assumption was used by [Spence and Owen \(1977\)](#). These authors assume that ad demand is flat and also that broadcast firms run into regulatory caps, so effectively there is neither an ad level decision to make, nor is there any ad surplus to worry about. Many papers treat this (simple) case of perfectly elastic demand, including [Gabszewicz, Laussel and Sonnac \(2001\)](#) and [Hansen and Kyhl \(2001\)](#). Other authors treat a downward sloping ad demand but do not treat ad surplus in the welfare analysis. This makes sense if all ads are viewed as pure social waste, and just serve to reshuffle demand. Nonetheless, such analyses do impart a socially important role for advertising spending – that of financing the media. On the other hand, if ads do generate some expected surplus to consumers, this surplus ought to be added to the utility function when deciding which channel to watch.<sup>26</sup> In order to disentangle the market performance in the media market per se from that in the ad

<sup>26</sup> This rather complicates matters. [Anderson and Coate \(2005\)](#) refer to analysis of this issue.

market, we shall analyze a benchmark case in which the private and social demand for advertising coincide.

One consistent story that generates such an ad demand is as follows. Suppose all advertisers are independent producers of new goods. Ads communicate the existence of these products to prospective consumers who could not otherwise find out about these goods and so can only buy them if they see an ad. Such ads are not persuasive but informative, and so are readily amenable to welfare analysis. The independence assumption implies that there can be no business stealing. We further suppose that each good is sold at a price that extracts all consumer surplus. This therefore closes down the other possible channel for deviation of social and private advertising.<sup>27</sup> Moreover, it obviates having to deal with the consumer surplus from goods in the viewer choice model. A simple formulation that ensures there is no consumer surplus arises when viewers, if interested, buy one unit of the good up to a reservation price that is common to all consumers.<sup>28</sup>

The aggregate demand for ads is then determined as follows. Following [Anderson and Coate \(2005\)](#), assume that advertisers differ in the probability that consumers are interested in their products, but advertisers are otherwise identical. More generally, it suffices to rank advertisers by a scale of high to low profit from contacting a viewer. All viewers are the same except for their tastes for programs, so there can be no targeting of ads correlated to programs.<sup>29</sup> Then the demand for advertising is simply the mass of advertisers who find that the expected benefit from communicating with viewers exceeds the price stipulated by the channel. We also assume that viewers “single-home” (watch one channel), and that there is a single period only. A viewer only needs to see an ad once in order to be informed of the product.<sup>30</sup> Taken together, the above assumptions imply that all that matters to an advertiser is the price for contacting a viewer. If the advertiser’s demand price exceeds the price quoted, the advertiser will advertise on the

<sup>27</sup> The broad Industrial Organization principle that governs the optimality of various economic magnitudes [following [Spence \(1976\)](#)] is that the bias depends on where the balance tilts between two opposing forces. First, firms do not take into account incremental consumer surplus that they cannot capture when they decide the level of an activity. This is termed Consumer Surplus Non-Appropriability. Second, firms do not account for the fact that they reduce other firms’ profits. This effect is commonly termed Business Stealing. These principles are usually applied to entry decisions [[Spence \(1976\)](#)] but apply equally well to advertising levels. In general then we should not expect the ad level to be optimal for these reasons. [Dukes \(2004\)](#) highlights business stealing by using the framework of [Grossman and Shapiro \(1984\)](#).

A neutral benchmark case naturally arises when we close down both of these channels of discrepancy of equilibrium from optimal levels. This has the advantage that we can then concentrate directly on the distortions inherent in finance by ad support, without yet worrying about the ad benefit per se. Put another way, the backdrop is one in which the private demand for ads coincides perfectly with the social demand.

<sup>28</sup> Alternatively, we could envisage a perfectly discriminating monopoly using two-part tariffs.

<sup>29</sup> Allowing targeted ads is a potentially important extension given the importance of demographic variables in ad demand.

<sup>30</sup> [Shields \(2004\)](#) reports that less clutter (i.e., non-program material) may result in a greater impact for advertisers. Ford sponsored the season-premiere of “24” on FOX, which was otherwise commercial-free. Ford’s brand recall score was over twice the average for the time period, according to IAG’s Reward TV data.



channel and reap a surplus equal to demand price minus advertising price per viewer, all multiplied by the number of consumers reached. This also means that active advertisers will typically advertise on all channels available, in order to reach the consumers who are delivered only through those channels. Equivalently, the advertisers engage in “multi-homing”.<sup>31</sup>

Let then the inverse advertising demand be given by  $p(a)$  per viewer when a channel carries  $a$  ads. The corresponding revenue per viewer is then  $R(a) = p(a)a$ . We assume below that this revenue function is log-concave. This means that the revenue function is quite “well-behaved” (the assumption includes a concave revenue as a special case).

At this juncture, we briefly review alternative views of advertising proposed in the literature on the economics of advertising. This literature traditionally distinguishes persuasive from informative advertising. Persuasive advertising is viewed as shifting consumer tastes [see, for example, the somewhat controversial paper by [Dixit and Norman \(1978\)](#), and the comments thereon in later issues of the *Bell Journal*]. [Dixit and Norman \(1978\)](#) take an agnostic view of how advertising works, but they take as a primitive that it shifts demand. The problem then for the practitioner of welfare economics is which demand curve to take as the true one. [Dixit and Norman](#) argue that over-advertising is the norm in both cases, whether one takes the pre-advertising demand or the post-advertising demand as the “true” one. However, explaining why demand shifted frequently leads us back to the complementary goods story or the informative advertising one described in more detail below.

Informative advertising works by telling consumers something about the product that then makes them more likely to buy, or to buy at a higher price. Informative advertising can be further split into that which indirectly informs consumers, and that which directly communicates product characteristics, quality, or price. Indirect information is communicated in signaling models [such as [Milgrom and Roberts \(1986\)](#)] in which advertising allows consumers to infer high quality in an adverse selection context.<sup>32</sup>

Directly informative advertising has been the topic of many studies. A major result in this context is due to [Butters \(1977\)](#), who finds that the market provides the socially optimal amount of advertising. His model was extended by [Stegeman \(1991\)](#), who suggests that the market tends to err towards under-advertising. An important contribution by [Grossman and Shapiro \(1984\)](#) introduces product differentiation via a circle model

<sup>31</sup> We address single-homing advertisers in Section 6 below.

<sup>32</sup> The view of advertising as a signal of product quality goes back to Nelson and to Klein and Leffler. A more formal treatment was undertaken by [Milgrom and Roberts \(1986\)](#), and several subsequent papers have extended this line of inquiry. The basic view is that advertising communicates quality by the firm “putting its money where its mouth is”. A low quality producer would not conspicuously spend large sums of money in promoting a product that no consumer would ever buy again. Thus reassured, consumers buy the product knowing that it is of high quality and they will repeat their purchases. The firm recoups its advertising expenses on the profits from the repeat purchases, and the advertising, which equivalently is public “money-burning”, serves an indirectly informative role, though no direct information about the product is actually transmitted in the advertisement.

[as in Vickrey (1964) and Salop (1979)]. This formulation was used by Dukes (2004) in the broadcasting context.

Another view is associated with the Chicago School [see, for example, Stigler and Becker (1977), and Becker and Murphy (1993)], and holds that advertising can provide a complementary good to the physical product. Advertising can be seen as enhancing the perceived product quality, for example, by fostering a brand image that consumers appreciate being associated with. Stigler and Becker (1977) argue first that “persuasive” advertising can be addressed within economic models, and, further that the level of advertising is socially optimal.

Depending on which view one takes, the conclusions below as regards the optimality of industry performance need to be tempered. For example, if there is over-advertising on broadcasts in the benchmark model, and it is believed that advertising levels are themselves excessive (so that the private benefit overstates the social benefit), then a fortiori advertising is excessive. Matters become more delicate when the conclusions from the separate parts run in opposite directions.

### 3.3. Platforms

The platforms are the TV stations that intermediate between the advertisers and the viewers. They are, of course, crucial to the two-sided market because they coordinate and balance the two sides on the platform. By bundling entertainment with the ads, they sweeten the delivery of the message in order to get it across.

The platforms that are covered by the current analysis are broadcasters (television and radio), publishers (newspapers and magazines), and web portals. Each follows the basic business model of delivering prospective customers to advertisers by attracting the viewership/listenership, readership, or web-surfer with news or entertainment content that carries a set of messages, superfluous (and possibly annoying) to the person enjoying the content. The relative size of the costs and benefits to agents who participate on the platform differs across applications. With television and radio, the advertisements break into the content and supplant it. In newspapers and magazines, the reader can easily bypass the ads so that the nuisance cost per se is likely negligible. While the overall modeling framework applies to various different markets, parameter values will differ across market applications.

Table 8 gives a time series of rates for prime-time network TV.

These rates are somewhat lower than the spot television rates, which are given in Table 9.

For comparison, magazine rates and newspaper rates are given below in Section 7.

In what follows we look at several variants of the model. We first consider a short run analysis in which the number of platforms is fixed. In the long-run, the number of platforms is determined by a free entry condition. We shall also first take the product locations as fixed, then look at them as being endogenously determined.

Table 8  
Network television cost, primetime (Mon–Sun) average program\*

	Households viewing (avg. min.)	Cost per 30 sec. (\$)	Cost per 1000 homes (\$)
1965	9,968,000	19,700	1.98
1975	13,500,000	32,200	2.39
1985	14,510,000	94,700	6.52
1995	10,860,000	95,500	8.79
2004	6,070,000	120,500	19.85

\*Source: [www.tvb.org](http://www.tvb.org), based on data from Nielsen.

Table 9  
Spot television cost for top 100 markets/30-second commercial primetime (Mon–Sun)

	Households per rating primetime	Cost per households rating primetime (\$)	Cost per 1000 homes (\$)
1982	703,092	6235	8.87
1985	732,211	7360	10.01
2005	941,219	24,181	25.69

Source: [www.tvb.org](http://www.tvb.org), based on data from Nielsen.

## 4. Equilibrium

We now find the equilibrium for the model, starting with a short-run analysis (a fixed number of firms) and then moving to a zero-profit (free entry and exit) equilibrium. In the next sub-section, we consider the case when advertising is the only revenue source. We then look at subscription pricing alone, and finally at pricing and advertising together. The model below encompasses many of those used in the literature as special cases, and we derive the equilibrium values for these special cases for comparison purposes. The particular models include the duopoly at the ends of a Hotelling line with linear transport costs [Anderson and Coate (2005)]; the quadratic transport cost version [Gabszewicz, Laussel and Sonnac (1999, 2002)]; and the Vickrey–Salop circle model with linear transport costs [Choi (2003); Crampes, Haritchabalet and Jullien (2004)].

### 4.1. Short-run equilibrium with advertising

Suppose that there are  $n$  platforms and let  $K$  be the fixed cost in setting up a platform. Given the revenue per consumer,  $R(a_i)$ , the profit to broadcaster  $i$  is

$$\pi_i = R(a_i)N_i(f_i, f_{-i}) - K \quad (3)$$

where we recall that  $f_i = s_i + \gamma a_i$  denotes the full price of broadcaster  $i$ ;  $f_{-i}$  denotes the vector of full prices of all broadcasters other than  $i$ , and the viewership functions  $N_i(\cdot)$  are determined from the particular viewer model assumed (i.e., the specification of Equation (2)). We go into more detail on this below (and give derivations in [Appendix A](#)), but for the moment write the own viewership derivative with respect to full price as  $N'_i < 0$  (i.e.,  $\partial N_i(f_i, f_{-i})/\partial f_i$ ). Then, suppressing arguments, we have, for an interior solution:

$$R'(a_i)N_i + R(a_i)\gamma N'_i = 0, \quad (4)$$

or, in elasticity form:

$$\frac{R'(a_i)}{R(a_i)}a_i + \frac{\gamma N'_i}{N_i}a_i = 0.$$

As long as  $R$  is log-concave,  $R'/R$  is a decreasing function. Likewise, as long as the viewership demand function is well behaved in the sense that  $N_i$  is log-concave, then  $N'_i/N_i$  is a decreasing function. These conditions suffice to ensure a unique solution for ad levels. In the case of a symmetric demand model, the equilibrium is symmetric. Supposing that there is a unit mass of consumers, the equilibrium viewership is  $N_i = 1/n$ . Denoting the common value of the viewership derivative (with respect to full price) by  $N'$  (i.e.,  $\partial N_i(f, f)/\partial f_i$ ), the equilibrium ad level,  $a^*$ , is defined implicitly from Equation (4) as

$$\frac{R'(a^*)}{R(a^*)} + \gamma n N' = 0. \quad (5)$$

In [Appendix A](#) we derive some common solutions for the viewer demand switch-over rate,  $N'$ .

It is useful to point out the case of  $\gamma = 0$ , meaning that viewers and readers are neutral about ads. In that case, due to product differentiation, each media firm will have its own market share, and is the exclusive channel for reaching the corresponding prospective consumers. Each channel will then price ads at the point where the marginal revenue from ads is zero (which is the marginal cost to the firm of airing an ad, and corresponds to the point of unit elasticity of the ad demand function).

If  $\gamma > 0$ , then viewers or readers find ads to be a nuisance. Advertising levels are then lower than if ads are not a nuisance. The reason is that competition is in nuisance levels, and firms strive to reduce the nuisance (all the while recognizing that the “nuisance” is the source of their income). It therefore makes sense that more competition (higher  $n$ ) results in lower equilibrium nuisance, just as more competition typically leads to lower prices (also a nuisance!) in standard models of product differentiation. These conclusions are borne out in [Appendix A](#) for some standard formulations of viewer choice. This interpretation in terms of competition for nuisance is important because one might usually expect the total “output” (ad level, by analogy) to be higher when there are

more firms.<sup>33</sup> Along with a lower level of ads the more firms there are, the approach concurrently predicts that the price per ad per viewer should be higher (with fewer firms advertising, the demand price is higher). Empirical evidence presented in [Brown and Alexander \(2005\)](#) gives the opposite relation, and so disagrees with the set-up above, but is consistent with the representative consumer model presented in [Cunningham and Alexander \(2004\)](#). More work is needed here to evaluate the hypotheses of the alternative approaches.

If  $\gamma < 0$ , then viewers or readers actually appreciate ads. While this does not seem relevant for the case of television or radio for the majority of people (since ads then necessarily displace entertainment content that presumably drew the individual in the first place to watch or listen), with magazines or newspapers there is no such presumption. The value of  $\gamma$  then represents the expected (net) surplus per ad seen by the reader. If this is negative, there is a desire for ads (ad-lovers), and the intermediary (television or radio broadcaster, newspaper or magazine publisher, web-master) must take this into account when determining how many ads to run. Now, more ads will actually attract more readers or viewers, but running more ads will also bring the broadcaster or publisher into the region where marginal revenue is negative. It follows from this logic, and from the equations presented previously, that now ad levels rise with the number of firms. This at least might sound more intuitive – “output” is larger with more firms and each has less market power to keep down the advertising level and so keep up the advertising price. Then the competition among firms is not in nuisance but rather in the attractiveness that is afforded by carrying many (desirable) ads.

[Dukes \(2004\)](#) emphasizes strategic interaction among advertisers. He models the product market with a circle framework using the oligopoly informative advertising model of [Grossman and Shapiro \(1984\)](#), and so allows for an explicit business stealing effect (at the cost of assuming that one industry supports the medium). The media are modeled with the circle framework, as above. He shows that less product differentiation or more media differentiation lead to a higher market level of advertising.<sup>34</sup> His result that there are more ads per station when products are closer substitutes contrasts with the standard [Grossman and Shapiro \(1984\)](#) finding, and underscores the importance of jointly considering the advertising and product markets. In the other direction, [Dukes and Gal-Or \(2003\)](#), [Gal-Or and Dukes \(2003\)](#) and [Gal-Or and Dukes \(2006\)](#) show that several features of the media industry may be explained by the joint incentive of both

<sup>33</sup> Indeed, though, the ad level per firm is lower. These properties vis-a-vis the workhorse Cournot model were first brought out by [Masson, Ram and Reynolds \(1990\)](#).

<sup>34</sup> [Anderson and Coate \(2005\)](#) get the latter result, but cannot treat the former in their ad specification. [Dukes \(2005\)](#) assumes instead that advertising is not informative but is used to differentiate a product from competing products. As usual, lower levels of media market competition lead to more advertising. However, here more advertising leads to higher surpluses in product markets since more advertising leads to more product market differentiation.

media and advertisers to limit the extent of advertising in order to limit competitive product information from consumers.<sup>35,36</sup>

#### 4.2. Short-run equilibrium with pricing

At this point it is worthwhile deriving several results that are quite standard to the economics of product differentiation regarding the pricing of differentiated commodities. These results are useful both in their own right for describing the properties of equilibrium when there is no advertising, and as an ingredient for the analysis of the next sub-section, which treats subscription prices along with advertising.

So consider now a broadcaster's (or a newspaper's, or a magazine's) profit if it uses only subscription prices. Assuming zero marginal cost for reaching viewers or readers,<sup>37</sup> its profit is

$$\pi_i = s_i N_i - K,$$

and so the price equation (the pricing first-order condition) is

$$N_i + s_i N'_i = 0,$$

which has a similar form to the advertising Equation (4) above.<sup>38</sup>

Now, under demand symmetry,  $N_i = 1/n$ , and this pricing equation reduces to a simple form

$$s = \frac{-1}{nN'}$$

where again the notation  $N'$  denotes that the viewer share derivative is taken at a symmetric solution (recall that this is negative, and so the subscription price is positive!). The values of  $N'$  for the commonly used models of product differentiation are given in [Appendix A](#), and hence the solutions for the symmetric subscription price are readily derived.

<sup>35</sup> These results stem from the fact that informative advertising is a competitive externality for competing producers [see [Grossman and Shapiro \(1984\)](#)]. Then competing producers supply too much advertising relative to the joint profit maximization. Competitive conditions in the commercial media industry determine the extent of this externality.

<sup>36</sup> [Dukes and Gal-Or \(2003\)](#) modify the [Dukes \(2004\)](#) model to investigate the incentives for broadcasters to sign exclusive contracts with advertisers, whereby competing advertisers are excluded from advertising. While such exclusion unambiguously benefits an advertiser, certain conditions must be satisfied before a broadcaster will offer such a contract. Exclusive contracts are offered when media have sufficient power in the market for advertising vis-à-vis advertisers since they are able to capture rents from consumers who are excluded from informative advertising. Exclusive contracts are more likely to occur when media markets are less differentiated or when consumers are unlikely to be informed about products in the absence of advertising.

<sup>37</sup> Positive marginal costs are addressed below, and are given a separate development because of their importance in the analysis of pricing and ad level choice which follows.

<sup>38</sup> Equivalently, the elasticity of the own viewer demand is  $-1$ .

It is useful for the analysis of the next sub-section to now present an intermediate result, dealing with the case when the subscription revenue received by the broadcaster or publisher is augmented by a fixed sum per viewer (or listener, as the case may be) of  $\bar{R}$ . This could arise if there is a step demand for ads per viewer, but we shall see below that it belies a more general principle.

The broadcaster's (publisher's) profit then is

$$\pi_i = (\bar{R} + s_i)N_i - K,$$

and, following the same steps as above, the price equation under demand symmetry reduces to a simple form

$$\bar{R} + s = \frac{-1}{nN'}.$$

We can now re-introduce this into the profit function to give the equilibrium value of profits as

$$\pi^* = \frac{-1}{n^2N'} - K.$$

The key property here is that this profit level is independent of  $\bar{R}$ . This is because whatever extra rents may be attached to the consumer are competed away at the equilibrium pricing decision. We term this result the revenue-neutrality property.<sup>39</sup> The revenue-neutrality result arises because markets are fully covered (all consumers watch/buy a magazine) and because each reader/viewer buys one magazine or TV channel each. Note from the form of the profit function that  $\bar{R}$  enters just like a (negative) average cost per unit would. In this light, it is unsurprising that average cost levels in such models do not affect equilibrium profit levels. The pricing equation simply determines mark-ups, which are the revenues earned per reader or viewer delivered. We return below to the wider impact of the revenue-neutrality property.

#### 4.3. Short-run equilibrium with pricing and advertising

We now allow the platforms to price along with their ad levels. When both price and advertising are positive, the value of advertising,  $a_p$ , solves<sup>40</sup>

$$R'(a_p) = \gamma.$$

<sup>39</sup> This result that has been noted previously by several authors, including Armstrong (2004) and Anderson and Coate (2005). Wright (2002) gets it in the context of mobile phone telephony.

<sup>40</sup> The idea behind this result can be seen by thinking of the broadcaster as maximizing revenues per viewer for a given level of full cost per viewer. That is, recalling the full cost is  $s + \gamma a$ , a broadcaster that maximizes the revenue per viewer,  $R(a) + s$ , subject to this constraint, will optimally choose to set  $R'(a) = \gamma$ . If negative pricing is not permissible or feasible, then the price is zero (there is no subscription fee even if one is feasible), and the ad level is determined by the market interaction over ad levels alone, as per the analysis of the preceding sections.

It is noteworthy that this result is independent of market structure. The intuition behind the result is that it parlays nuisance costs into ad revenues. Indeed, suppose that  $\gamma$  were above  $R'(a)$ . Then  $a$  could be reduced by a small amount,  $da$ , while  $s$  could be raised by a small amount  $\gamma da$  so that the full price to consumers is constant. Ad revenues would go down by  $R'(a) da$ , but by supposition this is less than the rise in subscription revenues ( $\gamma da$ ).

Given that  $a_p$  is determined by  $a_p = R'^{-1}(\gamma)$ , we can substitute this relation into the profit function and write it as

$$\pi_i = (R(a_p) + s_i)N_i - K,$$

and think of the broadcasters choosing just the subscription levels (or indeed, the full prices), since  $R(a_p)$  is tied down by the above relations.<sup>41</sup> This though means that the revenue  $R(a_p)$  plays just the same role as the fixed revenue  $\bar{R}$  in the analysis of the previous sub-section. The implication is then from the revenue-neutrality property that short-run profits are independent of the strength of advertising demand as long as  $s > 0$  (so that  $R'(a_p) = \gamma$ ). They are also independent of  $\gamma$ . Note lastly that ad levels are independent of the number of firms.

We have now to determine when subscription pricing will be used in equilibrium. As we argued above, a broadcaster will monetize the nuisance if  $R'(a) < \gamma$ . Conversely then, we can say that a broadcaster will not monetize any nuisance if  $R'(a) > \gamma$ . Most importantly, if the equilibrium without subscription pricing involves  $a^* = R'^{-1}(\gamma) < a_p$ , then introducing the ability to price will not have any effect and the equilibrium will be as in the first sub-section above. That is, pricing will not be used, and the equilibrium will remain a free-to-air commercial television, free newspapers or web-sites, etc., if  $a^* < a_p$ . If though  $a^* > a_p$ , the equilibrium when pricing is feasible will have positive subscription prices and the equilibrium level of advertising  $a_p$ . Pricing has important distributional effects. First, profits rise. This can be seen from comparing profit levels with and without it. With pricing, profits are  $\pi_p^* = \frac{-1}{n^2 N'} - K$ . Without pricing, they are  $\pi_a^* = \frac{R(a^*)}{n} - K$ , with  $a^*$  determined from  $\frac{R'(a^*)}{R(a^*)} = -n\gamma N'$ . Hence  $\pi_p^* > \pi_a^*$  as  $\frac{-1}{n^2 N'} > \frac{R'(a^*)}{R(a^*)}$ . Since  $R'(a^*) < \gamma$  in order for pricing to be used, then profits are higher when pricing is chosen in equilibrium.<sup>42</sup> Advertisers lose out because the ad level is lower, and so they lose some surplus. Viewers lose out because the full price (the sum of the advertising nuisance plus subscription price) rises.

The strong conclusion above is that the properties of the market equilibrium are just the same as the subscription-price-only model when prices are positive. They therefore depend only on the product differentiation specification used to describe the consumer

<sup>41</sup> See Anderson and Coate (2005) for more details on equilibria with ads and pricing.

<sup>42</sup> It is obvious that the ability to price raises profits for a monopoly, but it is not a priori clear that this is so for oligopoly since competition might be expected to be more severe once firms compete in more dimensions.



preferences in the market.<sup>43</sup> It is important to reiterate that these properties stem from full market coverage, unit demand per consumer, and also that parameter values ensure the equilibrium is in the regime with positive subscription prices along with advertising.<sup>44</sup> The influence of the latter condition may be rather subtle. As we shall see below, the equilibrium product selection may change drastically as parameters change, and the reason revolves around this condition.

#### 4.4. *Long-run equilibrium (free entry)*

The long-run equilibrium we consider involves zero profits for all firms.<sup>45</sup> The corresponding numbers of firms are determined from setting the profit expressions in the sub-sections above equal to zero. This is straightforward for the most part, but the effects of allowing pricing (as compared to the advertising-only equilibrium) does bear comment.

As noted above, when pricing is actually used along with advertising, short-run profits are independent of the strength of advertising and the nuisance to consumers. This

<sup>43</sup> Peitz and Valletti (2005) consider a two-stage duopoly game in which location along the unit Hotelling line is chosen first, then broadcasters compete for viewers. Otherwise the set-up is the same as in Anderson and Coate (2005), so that one contribution can be seen as endogenizing locations. Peitz and Valletti (2005) consider two symmetric games; one without subscription pricing, and the other one with subscription pricing. Their objective is to analyze the welfare properties of these two formulations. With subscription pricing, the location game involves extreme differentiation. Under free-to-air broadcasting, there is always some provision of ads (unless platforms are located at the very same point, which does not happen at equilibrium). As expected, ads decrease with nuisance, and increase with transportation costs (at a given location). Candidate location equilibria go monotonically from minimal differentiation (when nuisance is zero or transportation cost is infinity) to maximal (when nuisance is high enough or transportation cost is low enough). For given locations, welfare is only affected by ads. As expected from the Anderson and Coate (2005) analysis, pay-tv is better than free-to-air when nuisance is high (since free-to-air overprovides), or when locations are sufficiently close (under such duplication, both systems underprovide, but more so with free-to-air since competition escalates in providing few ads that cannot be compensated by pricing). Finally, for endogenous content provision, pay-tv is better with high nuisance costs: both systems provide extremely differentiated content, but pay-tv offers efficient ad levels, while free-to-air overprovides. When the nuisance parameter is very small: free-to-air almost minimally differentiates content, and pay-tv maximally differentiates (but these have same welfare losses), so free-to-air is worse because it severely underprovides ads given locations are almost at the center. Similar effects arise with respect to transportation costs.

<sup>44</sup> It is possible that advertising is chosen to be zero, which will happen if  $R'(0) \leq \gamma$ , meaning that the marginal revenue from the advertising sector starts out no higher than the nuisance cost. Then advertising is so annoying that broadcasters would price it out of the market. Note then that the optimal level of ads is also zero in this case.

<sup>45</sup> Spectrum constraints limit the number of broadcasters in many markets. Even if there are no such constraints (as with newspapers and magazines) the number of firms should be an integer, so the equilibrium number is the largest number making non-negative profits (implicitly assuming profits per firm decline with firm numbers), while the optimum number is not so constrained with a floor. The integer problem is ignored below, although explicitly considered in Anderson and Coate (2005), albeit for at most two firms. We also do not consider here the possibility of equilibria with entry deterrence: see Eaton and Wooders (1985) and Anderson and Engers (2001) for a description of such possibilities.

turn means that the long-run (zero-profit) configuration of firms is independent of these variables, and indeed is just the same as when there is no advertising or only subscription prices are chosen. As compared to the equilibrium with advertising only, if then pricing is rendered feasible, the long-run equilibrium number of broadcasters will be greater because of the higher profits associated with the ability to price.

## 5. Welfare analysis

Suppose that parameters ensure that all markets are served.<sup>46</sup> Then the *optimum* advertising level,  $a_o$ , has the marginal social cost,  $\gamma \geq 0$ , equal to the marginal social benefit, which is the advertising demand price. Thus it solves

$$p(a_o) = \gamma.$$

It is therefore immediately clear that the advertising level with pricing is below what is optimal. The marginal revenue curve that determines the equilibrium level is below the demand curve that determines the optimal level.

Without pricing though, either relation is possible, as the following discussion makes clear. For low  $\gamma$ , virtually all the advertisers ought to be communicating with the viewers. The equilibrium has the ad level provided by each broadcaster bounded above by the level  $R'^{-1}(0)$ , where marginal revenue is zero. This is effectively the “competitive bottleneck” property [see [Armstrong \(2004\)](#)] that each broadcaster has a monopoly in delivering its viewers and so prices access to those viewers monopolistically.<sup>47</sup> This is a feature of two-sided markets when one side single-homes. At the other extreme, if  $\gamma \geq p(0)$ , the optimum has no advertising because the nuisance cost exceeds the demand price (social benefit) of all ads. The equilibrium though always has advertising, because ads are the only source of revenue for broadcasters.

We now look at the entry dimension of performance. We continue to suppose that subscription pricing is infeasible (or indeed that it is not used in equilibrium). It is insightful to suppose that  $\gamma = 0$  and retain the assumptions of fully covered markets and unit demands by consumers. Then there is a total disconnect between the equilibrium and the optimum. The optimum has the number of firms as described in the previous section, which depends on the product differentiation parameters. It also has a level of advertising determined by  $p(a) = 0$ : given that viewers and readers are not disturbed

<sup>46</sup> Unserved markets are addressed in [Anderson and Coate \(2005\)](#).

<sup>47</sup> This monopoly position is due to the assumption in the models described that viewers are single-homing (choosing just one channel to watch). While it is true that at any given time a viewer may only watch one program, there still may be competition in delivering viewers in a multi-period context when viewers switch channels. [Anderson and Coate \(2005\)](#) provide a preliminary analysis of two-period competition with broadcasters, while [Armstrong \(2004\)](#) analyzes (simultaneous) readership of multiple magazines along the lines of [Caillaud and Jullien \(2003\)](#).

at all, the social optimum should have all advertisers with positive demand price communicate with the prospective buyers of their products. On the other hand, if  $\gamma = 0$ , there is no conduit for competition between firms. Each will choose the level of ads such that  $R'(a) = 0$ , and so maximizes the revenue per reader or viewer delivered. Advertisers (leastwise, those with demand prices for ads above  $p(R'^{-1}(0))$ ) choose to advertise on all channels so the number of prospective buyers reached is independent of the number of broadcasters in the market. What this means is then that the total revenue,  $R(a^*)$ , is a “prize” that is fully dissipated by the  $n$  broadcasters entering the market. The equilibrium number of broadcasters is then  $R(a^*)/K$ . Hence, for example, doubling the number of advertisers (at each level of willingness to pay), will double the number of firms in the market at equilibrium. But the optimal number will remain unchanged.

In summary, the advertising level when  $\gamma = 0$  is too small at the equilibrium and advertising revenue is a pure rent split by the number of firms. Hence there are too few ads and the number of firms may bear no relation to the optimal number. A weak ad demand will mean the market cannot be served; a strong one will be massively over-served. Anderson and Coate (2005) already note that if there is little ad demand, then the free market cannot provide much programming. This is clear in a system that needs ad revenues to survive. On the other hand, the market may over-provide too. For example, Anderson and Coate (2005) show the market may be served by two firms when it is optimal to only have one. This possibility of over-entry clearly extends to circle model with free entry, as is borne out by results in Choi (2003). It is also apparent for the explicit advertising model with business stealing as used by Dukes (2004). Indeed, parallel to the finding in Anderson and Coate (2005), Dukes finds that advertising is above the optimal level when media differentiation (measured by  $\tau$ ) is high enough.

Consider briefly the case when ad demand is perfectly elastic with demand price  $\beta$  per advertiser per viewer reached. Then ads are formally like standard prices in product differentiation models. At the optimum, though, we should have no ads shown if  $\gamma > \beta$ . Conversely, if  $\gamma < \beta$ , all advertisers should be allowed to advertise. The equilibrium number of ads varies continuously with  $\gamma$  though.

Finally, we return to the case when pricing is feasible along with advertising. Then, as shown in the short-run analysis, ad levels are independent of firm numbers. Ad levels are insufficient, because the optimum under covered markets sets  $p(a) = \gamma$ .<sup>48</sup> The number of firms in the market is the same as in the model when only prices can be used, and we know that the number of firms is typically too large in models of product differentiation, and these conclusions transfer directly. These conclusions differ quite drastically from those of the equilibrium without pricing (i.e., advertising only). It is worth recalling though our earlier caveat that we have assumed that each viewer watches one program, and that parameters ensure the equilibrium is in the region with fully served markets.

<sup>48</sup> Anderson and Coate (2005) deal with uncovered markets.

## 6. Product selection: Choosing program type

We now address the issue of “breadth” provided by the market, by which we mean the horizontal differentiation between products selected. To ease readability, we develop the model from the beginning to make this section free-standing. Given the motivating example is newspapers, we refer throughout to papers and readers.

Suppose then that there are two newspapers. Each produces at unit cost  $c \geq 0$ , and each sells advertising space to advertisers. The newspapers are sold at prices  $s_i$ ,  $i = 1, 2$ , to readers. Each reader buys only one paper (“single-homing”). Readers’ political opinions range from the extreme left to the extreme right. This taste diversity is represented by the unit interval  $[0, 1]$ . In standard fashion, the further the newspaper’s stance from the reader’s ideal point, the higher the disutility of the reader. Following Gabszewicz, Laussel and Sonnac (2002), we suppose that this disutility is  $t(x - x_i)^2 + s_i$  for a reader of type  $x$  buying a newspaper offering opinion  $x_i$  (see Equation (2)).

Let  $x_1$  and  $x_2$  denote the locations of the papers. The demand functions  $N_1$  and  $N_2$  for the newspapers are then easily derived as

$$N_1(s_1, s_2) = \frac{x_1 + x_2}{2} + \frac{s_2 - s_1}{2t(x_2 - x_1)}$$

and

$$N_2(s_1, s_2) = \frac{2 - x_1 - x_2}{2} + \frac{s_1 - s_2}{2t(x_2 - x_1)}.$$

The corresponding editorial revenues are then

$$\pi_i = (s_i - c)N_i(s_1, s_2), \quad i = 1, 2.$$

This model of the press industry is the standard Hotelling location model with quadratic transportation costs, and for this problem we know that firms always locate at the two extremes of the unit interval at the unique sub-game perfect equilibrium of the game in which firms select price and location [see d’Aspremont, Gabszewicz and Thisse (1979)]. Thus, in the absence of advertising revenues, the media’s ideological messages reflect maximal political diversity at equilibrium.

We now introduce the second source for financing daily press, revenue accruing from advertising. Of course, if advertising rates and volumes are assumed to be fixed and independent of the number of readers, the above conclusion still holds and advertising revenues simply add to revenues. However, the larger the readership, the more attractive should be the newspaper to advertisers, and the more they are willing to pay for exposure to a larger block of readers. For simplicity, we suppose that the demand for advertising per reader reached is perfectly elastic, and let the demand price (per advertiser per viewer) be  $\beta$ .

Then profits accruing to paper  $i$  from newspapers’ sales to the readership and advertising space to the advertisers now amount to

$$\pi_i = (s_i - c)N_i(s_1, s_2) + \beta N_i(s_1, s_2), \quad i = 1, 2.$$

Substituting from the reader demand functions above, we get

$$\pi_1(s_1, s_2) = (s_1 + \beta - c) \left( \frac{x_1 + x_2}{2} + \frac{s_2 - s_1}{2t(x_2 - x_1)} \right)$$

and

$$\pi_2(s_1, s_2) = (s_2 + \beta - c) \left( \frac{2 - x_1 - x_2}{2} + \frac{s_1 - s_2}{2t(x_2 - x_1)} \right).$$

These revenue expressions are just the same as those obtained by the firms in a spatial competition model with quadratic transportation costs, when a constant unit subsidy equal to  $\beta - c$  is added to the newspaper's price. This subsidy is equal to the difference between the unit receipt originating from advertising sales and the unit production cost of each copy of the newspaper. We may now identify the sub-game perfect equilibrium of the sequential game in which editors select, in the first stage, their political images  $x_1$  and  $x_2$  (opinion game) and in the second stage, their newspapers' prices  $s_1$  and  $s_2$  (price game). In the price game, payoffs are given by the equations above. However, prices are constrained to be non-negative. When the price game has an interior solution with positive prices,<sup>49</sup> these are given by

$$s_1^* = c - \beta + t(x_2 - x_1) \frac{2 + x_1 + x_2}{3}$$

and

$$s_2^* = c - \beta + t(x_2 - x_1) \frac{4 - x_1 - x_2}{3}.$$

Substituting these sub-game equilibrium values of the price game back into the profit functions enables us to now solve for the equilibrium to the papers' location game. This yields the conclusion that the equilibrium outcome depends crucially on the size of  $t$  and  $\beta$  [see Gabszewicz, Laussel and Sonnac (2002), for full details]. Indeed, when political preferences are strong ( $t$  "large") and/or when advertising receipts are weak ( $\beta$  "small"), the opinion game has a unique equilibrium with maximal political diversity. This is much as the game in pure subscriptions, which might be expected with weak advertising demand.

However, with weak political preferences and/or significant advertising receipts, the opinion game has a unique equilibrium with *minimal* political diversity.<sup>50</sup> Thus, when

<sup>49</sup> This assumption simplifies the presentation of the results here. It is not true though that the price game always has an interior solution. The values for prices given in the text can become negative, for instance, when  $\beta$  is large compared with  $c$ , in which case the subscription price must be equal to zero. See Gabszewicz, Laussel and Sonnac (2002) for more details.

<sup>50</sup> There is also an overlap region where both are equilibria. Gal-Or and Dukes (2003) offer an additional explanation for duplication. By offering similar programming (i.e. duplicating) media induce stiffer competition for viewers, thus coordinating on lower levels of advertising and thereby raising advertisers' surpluses in product markets.

papers seek advertising revenue, there are considerable consequences on the equilibrium of the opinion game: the tendency to offer readers maximal political diversity is fully reversed when political preferences are weak and/or advertising receipts are sufficiently important.<sup>51</sup> This conclusion reveals that the dependence of advertising rates on the readership's size may well induce editors to influence the political content they display to their readership so as to develop higher advertising resources.

The welfare economics of the above model are quite straightforward. The social optimum locations are at the quartiles.<sup>52</sup> The equilibrium though involves either minimum differentiation, or maximum differentiation.<sup>53</sup> These locations are equally bad from the social perspective. Recall that if the subscription prices are allowed to be negative, the only equilibrium is maximal differentiation. Negative prices could be reflected in giveaways like free gifts, although in that case one might expect readers could pick up several copies of free papers along with free gifts, so rendering such negative prices infeasible.<sup>54</sup> In practice, even though there do exist newspapers that are given away free, one might expect the non-negativity constraint to be reflected in a small nominal price to obviate outright wastage.

The equilibrium determined above was derived from two specific assumptions on nuisance costs (they are zero) and advertising revenues (they are constant per viewer). We might though expect somewhat similar results with positive nuisance costs and a revenue function that exhibits decreasing average returns per ad per viewer. Indeed, as long as the subscription price is positive, newspapers carry the ad level that satisfies  $R'(a) = \gamma$ , and so profits are independent of the ad revenue and competition is effectively competition in subscription prices alone, as we have seen above. Such competition leads to maximal differentiation. However, when locations are "too close", the subscription price is capped at zero. Then the equilibrium is in ads alone. This consideration leads us to briefly consider the case of competition in ads alone.

Suppose then that  $R(a) = \beta a$ , so that the ad demand is the same as above, yet now with  $\gamma > 0$ . Then the result is maximal differentiation because the ad competition model is formally equivalent to the pure subscription price one.<sup>55</sup> This means that there

<sup>51</sup> The price floor (non-negative price) is crucial to this result. Without it, equilibrium prices can take on negative values, dissipating advertising revenues to the benefit of readers and maximum differentiation would continue to prevail. With a price floor, the editors can, beyond some point, choose a political position closer to their competitor's, without further exacerbating competition.

<sup>52</sup> These locations minimize the average distance traveled. Each location is at the mid-point of the market it serves, and market sizes are equal.

<sup>53</sup> The maximum differentiation result is rather an artifact of the assumption that firms must locate in the unit interval. If locations are unrestricted, they choose to locate at  $(-\frac{1}{4}, \frac{5}{4})$ , which locations are outside the unit interval, although not as far apart as they could possibly go. These locations are more extreme than the tastes of any reader, and are socially less desirable than minimum differentiation.

<sup>54</sup> The "free gift" could be interpreted as the comics pages, so individuals only want one copy.

<sup>55</sup> To see this, note that the ad competition profit function is  $\pi_i = \beta a_i N_i(\cdot)$  and the profit function under price competition is  $\pi_i = s_i N_i(\cdot)$ . The argument of the readership function in both cases is the full price, so the two problems give the same solutions.

is a discontinuity in behavior between the cases of  $\gamma = 0$  (minimum differentiation) and  $\gamma > 0$  (maximal differentiation).

Suppose now that  $R'(a)$  is strictly concave. Note first that if  $\gamma = 0$ , both papers carry the ad level such that  $R'(a) = 0$ , and papers just strive to maximize the number of readers. Then they minimally differentiate. The analysis is quite complex if  $\gamma > 0$ . Preliminary results suggest that the equilibrium locations may get arbitrarily close to minimal differentiation. However, verifying that the second order conditions hold for the first stage (location) game remains elusive.

An overall evaluation of the state of the art on location competition as applied to the economics of media industries is as follows. The standard specification of quadratic disutility costs and price-only competition leads to maximal differentiation. This result constrains welfare analysis because it predicts that locations are always excessively far apart, contradicting casual empiricism and flexibility of the solution. Allowing for advertising competition offers the tantalizing proposition that locations could vary between maximal and minimal differentiation (according to parameters), and so the solution is not a priori constrained by excessive diversity. However, actually proving that the candidate equilibrium is the solution to the problem is a difficult mathematical problem.

## 7. Press concentration and advertising

Specialists in media economics have often viewed the advertising market as responsible for *concentration* in the press industry. This is backed up with empirical work that relates advertising rates to the circulation of newspapers.<sup>56</sup> Earlier theoretical contributions that ascribe the growth of concentration within the press industry to the interaction between advertising and newspapers' markets are due to Furhoff (1973), Gustafsson (1978), and Engwall (1981). We now understand these interactions as those of a two-sided market. The market for printed media is a particularly significant example of this phenomenon. Newspapers sell some space to advertisers and the larger the demand for advertising, the higher the share of advertising revenues in their total profits. On the other side of the market, readers' attitudes toward printed media advertising are quite ambiguous. Although it seems generally accepted that TV-viewers dislike advertising [see Brown and Rothschild (1993), and Danaher (1995)], it seems that readers of printed media have mixed views, and some have a positive perception of press advertising while others are negative.<sup>57</sup> If we take this at face value, then the utility of the readers is related to the size of advertising demand, positively for some and negatively for others. This means there are different types of network effects at play between the printed media and the advertising markets for the readership too. Indeed, some think that advertising could foster the circulation of newspapers [see Blair and Romano (1993), Gustafsson (1978),

<sup>56</sup> See Dertouzos and Trautman (1990), Reimer (1992) and Kaitatzi-Whitlock (1996).

<sup>57</sup> In a recent opinion poll, 37% of French readers claimed to be ad-averse (*Le Monde*, November, 9, 2002).

Table 10  
Magazine advertising cost and daily newspapers, US, circulation, costs, and cost per 1000 readers

			1965	1985	1997	2000
Magazine	Combined circulation	(000)	147,080	159,978	129,623	
	Combined page rate	B&W	\$595,143	1,695,541	3,358,235	
		4C	\$826,879	2,288,036	4,437,329	
	Cost per page per 1000	B&W	\$4.05	10.59	25.91	
		4C	\$5.53	14.3	34.23	
Newspapers	Total daily CIRC	(000)	60,358	62,766		55,773
	Cost 1/2 page each daily		\$312,112	1,515,163		3,712,650
	Cost per 1000 CIRC		\$5.17	24.14		66.57

Source: [www.tvb.org](http://www.tvb.org), based on data from Nielsen.

and Rosse (1980)]. Others believe that it slows it down [see Musnick (1999) and Sonnac (2000)].

Some statistics on advertising costs through reaching readers by magazine ads and by newspapers are given in Table 10.

We now analyze the interaction between the newsprint media and advertising industries when there are readers of both stripes. Let there be two editors producing differentiated newspapers or magazines (for instance, news-magazines proposing different political opinions) that take the extreme positions on a unit segment. Readers' tastes are distributed uniformly on  $[0, 1]$ . Newspaper 1 is located on this spectrum at point 0, while newspaper 2 is located at point 1. Editors also sell some proportion of their newspaper's surface to advertisers who buy it to promote the sales of their products. At each point  $x$  of the unit interval  $[0, 1]$ , a fraction  $\lambda$  of readers are *advertising-avoiders* and a proportion  $1 - \lambda$  are *advertising-lovers*. The advertising-avoiders lose utility when there are more ads in the paper, while the advertising-lovers gain. More precisely, suppose editor  $i$  quotes a price  $s_i$  for the newspaper and sells a proportion  $a_i$  of it to advertisers. For an advertising-avoiding reader located at  $x$ , the total loss in utility when buying newspaper 1 (at 0) is

$$x^2 + \gamma a_1 + s_1, \quad \gamma > 0,$$

while the total loss in utility when buying newspaper 2 (at 1) is  $(1 - x)^2 + \gamma a_2 + s_2$ . Similarly, for one of the  $1 - \lambda$  advertising-loving readers at  $x$ , the total loss in utility when buying newspaper 0 is

$$x^2 - \gamma a_1 + s_1, \quad \gamma > 0$$



and the total loss in utility when buying newspaper 2 is  $(1 - x)^2 - \gamma a_2 + s_2$ . Note that, for simplicity, it is assumed that the ad-loving propensity on one side is exactly equal to the ad-avoiding cost on the other side.

Define  $\tilde{k} = \gamma(2\lambda - 1)$  (so that the case of all ad avoiders corresponds to  $\tilde{k} = \gamma$ ). The reader demand function for newspaper  $i$ ,  $i = 1, 2$ , is then

$$\begin{aligned} D_i(s_1, s_2, a_1, a_2) &= 0 \quad \text{for } s_i \geq 1 + s_j + \tilde{k}(a_j - a_i); \\ D_i(s_1, s_2, a_1, a_2) &= 1 \quad \text{for } s_j + \tilde{k}(a_j - a_i) - 1 \geq s_i \geq 0; \\ D_i(s_1, s_2, a_1, a_2) &= \frac{1}{2}(1 + (s_j - s_i) + \tilde{k}(a_j - a_i)) \quad \text{otherwise.} \end{aligned}$$

The difference  $(a_i - a_j)$  between the advertising volumes in the papers plays a crucial role in the demands for the newspapers. At equal prices, the paper with the more advertising benefits from a larger readership if and only if there is an ad-loving majority in the reader population ( $\lambda < \frac{1}{2}$ ).

Total revenues also include advertising revenues from sales of advertising space. We now develop a model of the advertising market to derive the demand for advertising space as a function of the advertising rates charged by the editors in this market. Let  $q_i$  denote the unit price of an ad charged to advertisers by paper  $i$ ,  $i = 1, 2$ . Advertisers are ranked in the unit interval  $[0, 1]$  by increasing willingness to pay for an ad. Assume that each advertiser  $\theta$ ,  $\theta \in [0, 1]$ , buys an ad in only one of the two newspapers, at the exclusion of the other (thus we assume single-homing for advertisers). Assume that advertiser  $\theta$ 's benefit from inserting an ad in newspaper  $i$  at a rate  $q_i$  is  $D_i\theta - q_i$ , where  $D_i$  is the readership of paper  $i$  as given above. Since  $\theta \in [0, 1]$ , the advertising market is never covered.<sup>58</sup> This representation of the advertising market sets it up as a vertically differentiated industry [Gabszewicz and Thisse (1979)]; here, the ‘‘high quality’’ product firm is the newspaper with the larger readership.

Consequently, if  $a_i$  advertisers buy their ads in newspaper  $i$ , paper  $i$ 's total profit  $\pi_i$  is

$$\pi_i(s_1, s_2, q_1, q_2) = s_i D_i(s_1, s_2, a_1, a_2) + q_i a_i, \quad i = 1, 2. \tag{6}$$

We consider a two-stage game played between the papers. At the first stage, they select newsstand prices  $s_1(a_1^a, a_2^a)$  and  $s_2(a_1^a, a_2^a)$  conditional on the expected volumes  $a_1^a$  and  $a_2^a$  of advertising which will be determined in the second stage. Payoffs in the first stage depend on the expectations of both editors and readers about the difference  $a_i^a - a_j^a$  between the advertising volumes sold by the editors in the second period. These payoffs are given by Equation (6) with  $a_i - a_j = a_i^a - a_j^a$ .

The second stage strategies are the advertising prices  $q_1$  and  $q_2$ . Entering in this stage, prices  $s_1$  and  $s_2$  have been already selected determining readerships,  $D_i(s_1, s_2) = D_i$ . Based on the above model of the advertising market, payoffs in the second stage game are derived as a function of advertising rates  $q_1$  and  $q_2$ . Denoting by  $s_i^*(a_1^a, a_2^a)$ ,

<sup>58</sup> The case of a covered market is treated in Gabszewicz, Laussel and Sonnac (2004).

$i = 1, 2$ , the equilibrium values in the first-stage game, conditional on expectations  $a_1^a$  and  $a_2^a$ , and by  $(q_1^*, q_2^*)$  the equilibrium of the second-stage game, we further require  $a_i(q_1^*, q_2^*) = a_i^a$ : the value of the demand function of each editor in the advertising market at the second-stage equilibrium is consistent with first-stage expectations on these values.<sup>59</sup>

The equilibria of the game are as follows [and broadly substantiate parallel results by [Caillaud and Jullien \(2003\)](#)]. First, whatever the value of  $\lambda \in [0, 1]$ , there always exists an equilibrium corresponding to symmetric expectations ( $a_1^a = a_2^a$ ), with prices and market shares equal in both markets. This is an equilibrium with symmetric expectations about the advertising market shares. This equilibrium leads to Bertrand competition in the advertising market and, consequently, to equal prices and market shares in the newspaper market. In the case of ad-repulsion ( $\lambda > \frac{1}{2}$ ), no other equilibrium exists than the symmetric one.

Second, consider the case of a majority of ad-lovers ( $\lambda < \frac{1}{2}$ ). Then, if ad-attraction is strong ( $-6 \geq \tilde{k} = \gamma(2\lambda - 1)$ ), there are two asymmetric equilibria. At each, one editor eliminates the rival completely, and the eliminating editor is the one who is expected to sell more advertising. If ad-attraction is weaker ( $\tilde{k} > -12$ ), there are also two asymmetric equilibria with both editors enjoying strictly positive market shares in both the readership market and in the advertising markets. The paper which is expected to sell more advertising has higher prices and larger market shares in both markets. This latter result is akin to the base intuition of [Furhoff \(1973\)](#) revealing why ad-attraction can drive concentration growth in the daily newspaper industry. This intuition was described by [Gustafsson \(1978, p. 1\)](#) in the following terms: "The larger of two competing newspapers is favoured by a process of mutual reinforcement between circulation and advertising, as a larger circulation attracts advertisements, which in turn attracts more advertising and again more readers. In contrast, the smaller of two competing newspapers is caught in a vicious circle, its circulation has less appeal for the advertisers, and it loses readers if the newspaper does not contain attractive advertising. A decreasing circulation again aggravates the problems of selling advertising space; so that finally the smaller newspaper will have to close down." The equilibria under strong advertising attraction can be viewed as the limit of the market dynamics underlying this description. The paper which is expected to sell a larger number of ads makes itself more attractive than the rival one. The more ads the former inserts, the more this reinforces the attractiveness. This strengthening finally leads to the eviction of the latter from both the press and advertising markets. There also exist other equilibria corresponding to situations of weaker ad-attraction. At these equilibria, the paper with the larger expected share in the advertising market does not completely evict its rival. Nevertheless, the initial asymmetry about expected advertising market shares makes the paper with the larger expected share the leader in both industries since it sells more in both, and at higher prices.

<sup>59</sup> For a detailed equilibrium analysis, see [Ferrando et al. \(2004\)](#).

The interaction between the reader and advertising markets is rather complex when the two-sided network effects between these two industries are explicitly taken into account. It leads however to an important conclusion: *under ad-attraction, concentration in the press industry should be expected as a direct and natural consequence of the advertising market.* Since newspapers constitute a major vehicle for spreading political and social information to citizens, it is important to recognize the shortcomings in the business model that might lead to high concentration.

## 8. Conclusions

Television indubitably drives much of popular culture today and radio, magazines, and the Internet are other important drivers. Households in the US currently watch some 8 hours of television a day. A common business model describes several media markets, including television and radio, the Internet, and newspapers and magazines. Entertainment and content are the bait to get prospective purchasers of consumer goods to be exposed to advertisements. This chapter has described the economics of this business model. What makes broadcasting different from other goods is that the broadcast delivers two goods, the program to viewers and the audience to the advertisers. This is why it is a two-sided market. Put another way, the advertising is piggy-backed onto the program that interests the viewers.

We have examined several dimensions of performance of the market. These include the range, quality, and breadth of the offerings (magazines, television programs, websites) provided. Since these are classic dimensions studied in the economics of product differentiation, we borrowed heavily from the economic theory of product differentiation. The dimension of interest may also be a political measure, such as the difference between the political stances represented in newspapers. There are other dimensions apart from the measures of the diversification of offerings that are important to the economics of media industries. All of these dimensions are important determinants of the cultural level. However, the economic models must be interpreted with care. For example, a higher “quality” in the model is one that more people choose, *ceteris paribus*. This may not correspond to some paternalistic view of what people “should” be watching, reading, or listening to.

Media industries exhibit several market imperfections. Broadcasting is a public good that is nonetheless provided by the market system, and the reason it is provided is advertising finance. But the ads are a nuisance to viewers (negative externality). Broadcast firms often historically formed a tight oligopoly. Given this knot of imperfections, one might have a poor expectation of performance in the industry. Indeed, one usually expects public goods to be under-provided, market power to also cause underproduction, and negative externalities to cause overproduction. Surprisingly then, media markets are

able to deliver optimal performance configurations. Careful empirical work is needed to determine the values of key parameters in the structural model and hence to determine how far current practice deviates from optimal, and in what direction. For valuable progress in this direction, see [Wilbur \(2004b\)](#).

There is another type of market failure inherent to the market provision of broadcasting through advertising finance. Television offerings are disciplined by indirect consumer sovereignty: viewers “vote” with their eyes, and broadcast companies want to deliver viewers – of the right demographics – to advertisers. However, there is no incentive under advertising-based finance to cater to the tastes of viewers who would not buy the products advertised. This means there will be bias even without paternalistic views of what people “ought” to be watching. This bias explains the targeting of TV programs to those demographics that deliver most expected revenues to advertisers. Loosely, one would expect these to be the twenty- and thirty-somethings with high disposable incomes. Those with as-yet unformed tastes and large discretionary spending are the most lucrative targets for advertisers. Insofar as characters in television programs often reflect their audience (people like to watch characters similar to themselves), then one can indeed see many programs with characters in the 20–50 year old range. The reason is quite straightforward: the advertising dollars are in this range. This causes quite an important cultural bias. [Goettler \(1999\)](#) finds that shows attracting more homogeneous viewerships (in terms of age and gender) elicit higher advertising prices. This makes sense because then ads may be better targeted. The concurrent bias that this effect suggests is that there is a tendency for programming to be too narrow. Second, he finds that shows watched by 35–49 year olds command higher advertising rate premia, suggesting a bias in catering to this group. Third, he finds that the advertising price increases in the size of the audience in a convex manner. The associated bias is toward programming with mass appeal, and the Lowest Common Denominator concern of [Beebe \(1977\)](#) arises from this incentive.

Of course, whenever an audience is watching or paying attention to any event, there is the incentive to try to reach them with a message. This is all the more true in the age of TiVo and other ad avoidance technologies, and ever greater demand from advertisers to get their messages across. Commercial placement is being seen increasingly in movies and programs (e.g., BMW’s in James Bond films in place of the traditional Aston Martin), and it will alter the scripts of the movies themselves, as writers have to write in the sponsoring products. One might view such placement as a Trojan Horse carrying in undesired elements. A similar phenomenon has been happening in US schools. Channel One provides programming free to schools and even pays for the hardware (televisions, satellite dishes, etc.). The Trojan Horse is the advertising that comes with the programming. The schools must guarantee 90% of students watch, and they are not allowed to turn off the supposedly educational programming. The programming is designed rather like MTV programming, and, although purportedly covering current affairs, is really de-

signed as a vehicle for advertisements.<sup>60</sup> Advertisers in turn pay high premia for slots: up to twice the amount spent to reach adults.<sup>61</sup>

Children's viewing habits are also the concern underlying the FCC's requirement (since 2001) that each new television be equipped with a V-chip (see <http://www.fcc.gov/vchip/>). The chip enables parents to block programming deemed unsuitable, as rated by the "TV Parental Guidelines", that were established by the National Association of Broadcasters, the National Cable Television Association and the Motion Picture Association of America. However, use of this blocking device does not seem to have caught on (in contrast with Internet blocking filters). As Hazlett (2004) puts it: "Perhaps parents adopted Internet filters and spurned the V-chip because watching television is generally a more public activity: It's easier to keep an ear on what your kids are watching in the living room than to keep an eye on every Web site they see."

Several recent papers have empirically investigated models of two-sided markets in various different contexts. The pioneering paper in the print area is Rosse (1970), who estimated cost curves in the newspaper industry in the context of a model that included newspapers' feedback effects between advertising and readership. Rysman (2004) considers welfare properties of various market structures in the yellow-pages market, and finds that oligopoly is preferred to monopoly. Since consumers use yellow pages to find information, his conceptual framework could be applied to a newspaper industry with ad-loving readers. Kaiser and Wright (2004) take Armstrong's (2004) model of two-sided market competition to data on the German magazine industry, assuming single-homing on the part of both advertisers and readers.

There have also been several recent empirical studies of the broadcasting industry. Berry and Waldfogel (1999a, 1999b, 2001) use data on radio listenership to look at the effects of entry and concentration in radio. Sweeting (2004) tackles the difficult problem of estimating an equilibrium model of the timing of radio stations' commercial breaks. Wilbur (2004a) uses advertising levels and ad prices and viewing data to estimate a two-sided model of the television advertising market. He finds ads cause viewing losses: preliminary estimates indicate that 30 seconds of ads on top of the current level will decrease watching by 1.1% of viewers (or 0.7% of households) per hour.

One direction for future research is to take a deeper look at actual market structures. Most systems are mixed, and include a variety of different firm types. Anderson (2003) provides a preliminary analysis in this direction by looking at the coexistence of advertising financed and subscription financed television, while Peitz and Valletti (2005) also compare these two types of regime. If we look at actual market structures there are various types operating in the market, with varying degrees of public support, regulation, advertising finance, etc. In the US, public television, supported by the Federal

<sup>60</sup> "Channel One is more commercial than network TV; its hipper, faster-moving, full of loud rock music and directly or indirectly, its always selling something." [Fox (2004)]

<sup>61</sup> "Each 30 second ad costs advertisers nearly \$160,000, more than twice the cost of a commercial on prime time television news." [Gange (2004), in a review of Fox (2004)]

Government and by private donations, coexists with commercial channels, pay channels and religious channels. In the UK, the BBC is supported by television license fees and Government grants, and is not allowed to carry ads. There are six major channels in the television market in France. These vary by the level of Government support and the number of ads they are allowed to screen, as well as by content of programming, and one is a pay channel during prime time and into the night. One thorny problem for future research concerns the appropriate modeling of the behavior and objectives of a Public Broadcaster.

On the empirical side, it is worth investigating more deeply the extent to which programming is indeed duplicated along the lines suggested by Steiner (1952), and the work on the *Pensée Unique*, or whether indeed programs are more differentiated. Very useful work in this direction is the study by Goettler and Shachar (2001). These authors suggest that broadcast firms instead differentiate their offerings quite substantially.

Another issue of cultural concern is the “quality” of programming defined from the perspective of the local community. The FCC in the US stresses concerns about “localism” in the decision to grant a license. Similarly, local content rules (as in France and Australia, among others) are designed to retain and foster national programming. These are related issues because they stress a concept of quality that reflects and bolsters community appreciation of its integrity. If an objective such as protecting community identity is valued, then it would presumably need special protection (or subsidy) when faced with a Lowest Common Denominator type programming of mass appeal (“Hollywood” to the protagonists). The appropriate policy stance in this regard remains an open research issue.

### Acknowledgement

The first author gratefully acknowledges research funding from the NSF under grant SES-0137001. We thank many conference participants for their comments, insights, and presentations. We also thank our co-authors for several of the insights on which we draw, Victor Ginsburgh for editorial help, Juan Carlos Bisso for research assistance, and Anthony Dukes, Serguey Khovansky, Alan Krueger, Nathalie Sonnac, Cat Tyler, and Ken Wilbur for comments and suggestions.

### Appendix A

Several special cases treated in the literature may be derived from a circle model with power transport costs. We therefore determine the equilibrium for that model. Assume the circle has circumference  $L$ , and the transport cost function is  $\tau(\cdot)$ . We seek a symmetric equilibrium. Let the indifferent viewer between broadcaster  $i$  (at location 0) and its clockwise neighbor (at  $L/n$ ) be at  $\hat{x}$ , so  $i$ 's viewership is simply  $2\hat{x}$ , and our task is to determine the value of  $N'$ .

If the other broadcasters all set a common ad level  $a$ , while broadcaster  $i$  sets  $a_i$ , demand satisfies equality of “full nuisance” or

$$\gamma a_i + \tau(\hat{x}) = \gamma a + \tau\left(\frac{L}{n} - \hat{x}\right).$$

Evaluating at a symmetric solution,

$$\frac{d\hat{x}}{da_i} = \frac{-\gamma}{2\tau'\left(\frac{L}{2n}\right)}.$$

For power transport costs,  $\tau(x) = tx^\alpha$ ,  $\alpha \geq 1$ , this becomes

$$\frac{d\hat{x}}{da_i} = \frac{-2^{\alpha-2}\gamma n^{\alpha-1}}{\alpha t L^{\alpha-1}}$$

so that ad levels solve (by Equation (5) and recalling that in a symmetric equilibrium,  $N_i = L/n$ , and  $N'_i = \frac{2}{\gamma} \frac{d\hat{x}}{da_i}$ )

$$\frac{R'(a^*)}{R(a^*)} = \frac{2^{\alpha-1}\gamma n^\alpha}{\alpha t L^\alpha}.$$

Since the left-hand side is falling in  $a$ , the important result here is that the equilibrium ad level is *decreasing* in the number of competing platforms. This makes sense here because competition is effectively contested over the nuisance value to consumers, and more competition means less nuisance, and so less advertising on each channel.<sup>62</sup>

There are some important special cases to this. First, suppose that ad demand is perfectly elastic. Then  $R = \beta a$ , where  $\beta$  is then the demand price for ads. Then  $\frac{R'}{R} = \frac{1}{a}$

<sup>62</sup> We assume in the main text that the strategic variable is price per ad per viewer. This is equivalent to ad levels being the strategic variables (this equivalence property follows from the feasibility of multi-homing for advertisers and the assumption that the willingness to pay for an ad is a linear function of the number viewers delivered by a channel). This strategic assumption corresponds to broadcasters that choose the size of their advertising breaks and then sell the space to prospective advertisers. An alternative strategic assumption is that broadcasters take as given the ad prices of rivals. This could be considered a Bertrand assumption where the standard one is a Cournot assumption. The implications of the alternative (Bertrand) assumption may be thought of as follows. When a broadcaster changes its price per ad, and a rival's price is assumed fixed, the rival must adjust its ad levels to keep its price per ad constant. More concretely, suppose a broadcaster raises its price per ad. Fewer advertisers want buy ads there, and so more viewers watch (which dampens the initial effect somewhat). This means that the rival needs to cut back its ad level because otherwise the rival is delivering fewer consumers per ad. It must make up for that by improving the “quality” of its ads – by delivering more viewers. This then means that the ad levels move together when broadcasters use ad price as strategies. This means more collusive behavior. Since more collusion implies MORE ads, price strategies lead to more ads, and closer to monopoly levels. It is noteworthy that Bertrand competition in standard differentiated products markets typically leads in stead to more competitive outcomes than Cournot competition. [Crampes, Haritchabalet and Jullien \(2004\)](#) also uncover such an effect. [Nilssen and Sjørgard \(2003\)](#) compare price and quantity strategies, using a representative consumer approach to advertising demand.

and we can solve for  $a^*$  directly as

$$a^* = \frac{\alpha t L^\alpha}{2^{\alpha-1} \gamma n^\alpha}. \quad (\text{A.1})$$

Ad levels rise with product differentiation ( $t$ ) and fall with the number of firms. They fall as viewers get more annoyed by them too.

The special case of  $\alpha = 1$  gives the standard Vickrey–Salop [Vickrey (1964), Salop (1979)] result of<sup>63</sup>

$$a = \frac{tL}{\gamma n}.$$

Noting that the standard duopoly line model is like  $L = 2$  (the firms are one unit distance apart), then (setting  $n = L = 2$  and  $\alpha = 1$  in Equation (A.1)) we have for that case<sup>64</sup>

$$a = \frac{t}{\gamma}.$$

In a similar fashion, for the standard version of the line model with two platforms and quadratic transport costs, set  $n = L = \alpha = 2$  in Equation (A.1) to give exactly the same result as for linear transport costs:

$$a = \frac{t}{\gamma}.$$

<sup>63</sup> Gal-Or and Dukes (2006) investigate the incentives for non-consolidating media mergers for firms on the circle, using the framework of Dukes (2004) described further below. On the one hand, a merging media firm improves its market power vis-à-vis advertisers since they now have a larger set of viewers. However, reduced competition for viewers induces higher equilibrium levels of advertising, which, in their model, lowers product market surpluses. They show that media mergers are profitable when the media market is sufficiently competitive so that, post-merger, the market power benefit exceeds the losses associated with increased advertising levels. Their results contrast with traditional product markets, where mergers are more profitable with less competition [Deneckere and Davidson (1985)]. Choi (2003) considers mergers of neighboring firms in the context of the model of this section. He shows the more familiar result [see also Eaton and Wooders (1985)] that such mergers raise the profits of all firms, and these profits are lower for firms further from those that merge.

<sup>64</sup> The monopoly case is rather interesting. Anderson and Coate (2005) show that the monopoly ad level for the Hotelling specification may be higher or lower than the duopoly one as the market is or is not served, respectively, under monopoly. This result appears quite specific to the Hotelling model with unserved markets though, as we argue below.



Rather similar qualitative results hold for the logit model of demand considered by Anderson (2000).<sup>65</sup> Here we have simply that

$$N' = \frac{1}{\mu} \frac{1}{n} \left( \frac{1}{n} - 1 \right),$$

and so the equilibrium ad level is given by

$$\frac{R'}{R} = \frac{\gamma}{\mu} \left( \frac{n-1}{n} \right).$$

With perfectly elastic ad demand, as above, this reduces to the closed form:

$$a = \frac{\mu}{\gamma} \left( \frac{n}{n-1} \right)$$

which increases with product differentiation as measured by  $\mu$  and also decreases with the number of firms.<sup>66</sup>

The logit form may also readily be extended to allow for viewers who do not watch. Then, with the logit above, and an outside option,<sup>67</sup> we get an implicit form for the equilibrium ad level of:

$$\frac{R'}{R} = \frac{\gamma}{\mu} \left( 1 - \frac{\exp(-\frac{\gamma a}{\mu})}{n \exp(-\frac{\gamma a}{\mu}) + \exp(\frac{V_0}{\mu})} \right)$$

where  $V_0$  denotes the relative quality of the outside option.

Rewriting,

$$\frac{R'}{R} = \frac{\gamma}{\mu} \left( 1 - \frac{1}{n + \exp(\frac{V_0 + \gamma a}{\mu})} \right).$$

Since the right-hand side is increasing in  $a$ , there is a unique solution; since the right-hand side is increasing in  $n$ , higher  $n$  always means lower ad levels for all  $n$ , including

<sup>65</sup> The logit model is given by assuming the  $\varepsilon_i$  terms in Equation (2) are i.i.d. double exponential. Suppressing the transport cost and quality components, the viewership demand function is

$$N_i = \frac{\exp(-\gamma a_i / \mu)}{\sum_{j=1, \dots, n} \exp(-\gamma a_j / \mu)}.$$

<sup>66</sup> These properties hold for a general class of discrete choice models with i.i.d. idiosyncratic tastes with a log-concave distribution.

<sup>67</sup> The viewership demand function is then

$$N_i = \frac{\exp(-\gamma a_i / \mu)}{\sum_{j=1, \dots, n} \exp(-\gamma a_j / \mu) + \exp(V_0 / \mu)},$$

where  $V_0$  measures the attractivity of the outside option [see Anderson, Palma and Thisse (1992) for further details].

the transition from monopoly to duopoly. The intuition follows naturally since competition over viewers involves nuisance levels (where one would normally have direct prices) higher competition implies lower prices.

## References

- Anderson, S.P. (2000). "Advertising and the media". Mimeo. University of Virginia.
- Anderson, S.P. (2003). "Broadcast competition: Commercial and Pay TV". Mimeo. University of Virginia.
- Anderson, S.P., Coate, S. (2005). "Market provision of broadcasting: A welfare analysis". *Review of Economic Studies* 72 (4), 947–972.
- Anderson, S.P., Engers, M.P. (2001). "Pre-emptive entry in differentiated product markets". *Economic Theory* 17, 419–445.
- Anderson, S.P., Neven, D.J. (1989). "Market equilibrium with combinable products". *European Economic Review* 33, 707–719.
- Anderson, S.P., Palma, A., Thisse, J. (1992). *Discrete Choice Theory of Product Differentiation*. MIT Press, Cambridge, MA.
- Armstrong, M. (2004). "Competition in two-sided markets". Mimeo. University College, London.
- Bagwell, K. (2003). "The economic analysis of advertising". Mimeo. Columbia University.
- Barros, P.P., Kind, H.J., Nilssen, T., Sørgard, L. (2005). "Media competition on the Internet". *Topics in Economic Analysis and Policy* 4 (1), article 32 (18 pp.).
- Becker, G., Murphy, K. (1993). "A simple theory of advertising as a good or bad". *Quarterly Journal of Economics* 107, 942–964.
- Beebe, J. (1977). "Institutional structure and program choices in television markets". *Quarterly Journal of Economics* 91, 15–37.
- Berry, S., Waldfogel, J. (1999a). "Public radio in the U.S.: Does it correct market failure or cannibalize commercial stations?". *Journal of Public Economics* 71 (2), 189–211.
- Berry, S., Waldfogel, J. (1999b). "Free entry and social inefficiency in radio broadcasting". *RAND Journal of Economics* 30, 397–420.
- Berry, S., Waldfogel, J. (2001). "Do mergers increase product variety? Evidence from radio broadcasting". *Quarterly Journal of Economics* 116 (3), 969–1007.
- Blair, R.D., Romano, R.E. (1993). "Pricing decisions of the newspaper monopolist". *Southern Economic Journal* 59 (4), 723–732.
- Boulding, K. (1955). *Economic Analysis*. Volume I: Microeconomics, third ed. Harper & Row, New York.
- Brown, K., Alexander, P.J. (2005). "Market structure, viewer welfare, and advertising rates in local broadcast television markets". *Economics Letters* 86 (3), 331–337.
- Brown, T.J., Rothschild, M.L. (1993). "Reassessing the impact of television advertising clutter". *Journal of Consumer Research* 20, 138–146.
- Butters, G. (1977). "Equilibrium distributions of sales and advertising prices". *Review of Economic Studies* 44, 465–491.
- Cabizza, M.M. (2004). "Minority programming in television markets". Mimeo. University of York (UK).
- Caillaud, B., Jullien, B. (2001). "Competing cybermediaries". *European Economic Review* 45, 797–808.
- Caillaud, B., Jullien, B. (2003). "Chicken & egg: Competition among intermediation service providers". *RAND Journal of Economics* 34 (2), 309–328.
- Cancian, M., Angela, B., Bergstrom, T. (1995). "Hotelling location problems with directional constraints: An application to television news scheduling". *Journal of Industrial Economics* 43, 121–124.
- Choi, J.P. (2003). "Broadcast competition and advertising with free entry". Mimeo. Michigan State University.
- Corneo, G. (2001). "Work and television". IZA Discussion Paper 376.
- Crampes, C., Haritchabalet, C., Jullien, B. (2004). "Competing with advertising resources". Mimeo. IDEI, Université de Toulouse.

- Cunningham, B.M., Alexander, P.J. (2004). "A model of broadcast media and commercial advertising". *Journal of Public Economic Theory* 6 (4), 557–575.
- d'Aspremont, C., Gabszewicz, J.J., Thisse, J. (1979). "On Hotelling's "Stability in Competition"". *Econometrica* 47 (5), 1145–1150.
- Danaher, P.J. (1995). "What happens to television ratings during commercial breaks". *Journal of Advertising Research* 37, 37–47.
- Deneckere, R., Davidson, C. (1985). "Incentives to form coalitions with Bertrand competition". *RAND Journal of Economics* 16 (4), 473–486.
- Dertouzos, J.N., Trautman, W.B. (1990). "Economic effects of media concentration: Estimates from a model of the newspaper firm". *Journal of Industrial Economics* 39 (1), 1–14.
- Dixit, A., Norman, V. (1978). "Advertising and welfare". *Bell Journal of Economics* 9, 1–17.
- Dukes, A.J. (2004). "The advertising market in a product oligopoly". *Journal of Industrial Economics* 52 (3), 327–348.
- Dukes, A.J. (2005). "Media concentration and consumer product prices". Mimeo. Aarhus University.
- Dukes, A.J., Gal-Or, E. (2003). "Negotiations and exclusivity contracts for advertising". *Marketing Science* 22 (2), 222–245.
- Eaton, B.C., Lipsey, R.G. (1975). "The principle of minimum differentiation reconsidered: Some new developments in the theory of spatial competition". *Review of Economic Studies* 42, 27–49.
- Eaton, B.C., Wooders, M. (1985). "Sophisticated entry in a model of spatial competition". *RAND Journal of Economics* 16 (2), 282–297.
- Economides, N. (1996). "The economics of networks". *International Journal of Industrial Organization* 14, 673–699.
- Engwall, L. (1981). "Newspaper competition: A case for theories of oligopoly". *Scandinavian Economic History Review* 29 (2), 145–154.
- Ferrando, J., Gabszewicz, J., Laussel, D., Sonnac, N. (2004). "Two-sided network effects and competition: An application to media industries". *Lucarnes bleues 2004/09, CREST/LEI*.
- Fox, R.F. (2004). *Harvesting Minds: How TV Commercials Control Kids*. Praeger.
- Furhoff, L. (1973). "Some reflections on newspaper concentration". *Scandinavian Economic History Review* 1, 1–27.
- Gabszewicz, J.J., Thisse, J.-F. (1979). "Price competition, quality and incomes disparities". *Journal of Economic Theory* 20 (3), 340–359.
- Gabszewicz, J.J., Wauthy, X. (2004). "Two-sided markets and price competition with multi-homing". CORE Discussion Paper 2004 (30). Louvain-la-Neuve, Belgium.
- Gabszewicz, J.J., Laussel, D., Sonnac, N. (1999). "TV broadcasting competition and advertising". *Lucarnes bleues # 9972 CREST/LEI, CORE Discussion Paper # 2000(6)*.
- Gabszewicz, J.J., Laussel, D., Sonnac, N. (2001). "Press advertising and the ascent of the pensee unique?" *European Economic Review* 45, 645–651.
- Gabszewicz, J.J., Laussel, D., Sonnac, N. (2002). "Press advertising and the political differentiation of newspapers". *Journal of Public Economic Theory* 4 (3), 249–259.
- Gabszewicz, J.J., Laussel, D., Sonnac, N. (2004). "Programming and advertising competition in the broadcasting industry". *Journal of Economics and Management Strategy* 13 (4), 657–669.
- Gal-Or, E., Dukes, A.J. (2003). "Minimum differentiation in commercial media markets". *Journal of Economics and Management Strategy* 12 (3), 291–325.
- Gal-Or, E., Dukes, A.J. (2006). "On the profitability of media mergers". *Journal of Business* 79 (2).
- Gange, M. (2004). "Making a killing on our kids". *Review of harvesting minds: How TV commercials control kids*. By Roy F. Fox. <http://medialit.med.sc.edu/fox.htm>
- Goettler, R. (1999). "Advertising rates, audience composition, and competition in the network television industry". Mimeo. Carnegie-Mellon University.
- Goettler, R., Shachar, R. (2001). "Spatial competition in the network television industry". *RAND Journal of Economics* 32, 624–656.
- Grossman, G., Shapiro, C. (1984). "Informative advertising with differentiated products". *Review of Economic Studies* 51, 63–81.

- Gustafsson, K.E. (1978). "The circulation spiral and the principle of household coverage". *Scandinavian Economic History Review* 26 (1), 1–14.
- Hamilton, J. (2004). *All the News That's Fit to Sell*. Princeton University Press.
- Hansen, C., Kyhl, S. (2001). "Pay-per-view broadcasting of outstanding events: Consequences of a ban". *International Journal of Industrial Organization* 19, 589–609.
- Hazlett, T. (2004). "Requiem for the V-chip: A relic of the last battle over indecency on TV". <http://slate.msn.com/id/2095396/> Friday, February 13, 2004.
- Hotelling, H. (1929). "Stability in competition". *Economic Journal* 39, 41–57.
- Kaiser, U., Wright, J. (2004). "Price structure in two-sided markets: Evidence from the magazine industry". Mimeo. University of Southern Denmark.
- Kaitatzi-Whitlock, S. (1996). "Pluralism and media concentration in Europe". *European Journal of Communication* 11 (4), 453–483.
- Masson, R., Ram, M., Reynolds, R. (1990). "Oligopoly in advertiser-supported media". *Quarterly Review of Economics and Business* 30 (2), 3–16.
- Milgrom, P., Roberts, J. (1986). "Prices and advertising signals of product quality". *Journal of Political Economy* 94, 796–821.
- Musnick, I. (1999). "Le Cœur de Cible ne Porte pas la Publicite Dans son Coeur". *CB News* 585, 11–17. October 8–10.
- Mussa, M., Rosen, S. (1978). "Monopoly and product quality". *Journal of Economic Theory* 18, 301–317.
- Nilssen, T., Sjørgard, L. (1998). "Time schedule and programme profile: TV news in Norway and Denmark". *Journal of Economics and Management Strategy* 7 (2), 209–235.
- Nilssen, T., Sjørgard, L. (2003). "The TV industry: Advertising and programming". Mimeo. Bergen.
- Peitz, M., Valletti, T. (2005). "Content and advertising in the media: Pay-tv versus free-to-air". Mimeo. Imperial College, London. CEPR Working Paper #4771.
- Reimer, E. (1992). "The effects of monopolization on newspaper advertising rates". *American Economist* 36 (1), 65–69.
- Rochet, J.-C., Tirole, J. (2003). "Platform competition in two-sided markets". *Journal of the European Economics Association* 1 (4), 990–1029.
- Rochet, J.-C., Tirole, J. (2004). "Two-sided markets: An overview". Mimeo. IDEI, Université de Toulouse.
- Rogerson, R. (2005). "Structural transformation and the deterioration of European labor market outcomes". Mimeo. Arizona State University.
- Rosse, J.N. (1970). "Daily newspapers, monopolistic competition, and economies of scale". *American Economic Review* 57 (2), 522–533.
- Rosse, J.N. (1980). "The decline of direct newspaper competition". *Journal of Competition* 30, 65–71.
- Rothenberg, J. (1962). "Consumer sovereignty and the economics of television programming". *Studies in Public Communication* 4, 45–54.
- Rysman, M. (2004). "Competition between networks: A study of the market for yellow pages". *Review of Economic Studies* 71 (2), 483–512.
- Salop, S. (1979). "Monopolistic competition with outside goods". *Bell Journal of Economics* 10, 141–156.
- Samuelson, P. (1964). "Public goods and subscription TV: Correction of the record". *Journal of Law and Economics* 7, 81–83.
- Shaked, A., Sutton, J. (1982). "Relaxing price competition through product differentiation". *Review of Economic Studies* 49, 3–14.
- Shields, M. (2004). "Ad clutter reaches new high, encroaches on program content". *MediaPost Daily News*, Tuesday, April 13.
- Sonnac, N. (2000). "Readers attitudes towards press advertising: Are they ad-lovers or ad-averse?" *Journal of Media Economics* 13 (4), 249–259.
- Spence, A.M. (1976). "Product selection, fixed costs and monopolistic competition". *Review of Economic Studies* 43, 217–235.
- Spence, M., Owen, B. (1977). "Television programming, monopolistic competition and welfare". *Quarterly Journal of Economics* 91, 103–126.

- Stegeman, M. (1991). "Advertising in competitive markets". *American Economic Review* 81 (1), 210–223.
- Steiner, P.O. (1952). "Program patterns and preferences, and workability of competition in radio broadcasting". *Quarterly Journal of Economics* 66, 194–223.
- Stigler, G.J., Becker, G.S. (1977). "De gustibus non est disputandum". *American Economic Review* 67, 76–90.
- Sweeting, A. (2004). "Coordination games, multiple equilibria and the timing of radio commercials". Mimeo. Massachusetts Institute of Technology.
- Vickrey, W.S. (1964). *Microstatics*. Harcourt/World, Brace/New York. Selection republished as Vickrey, W.S., Anderson, S.P., Braid, R.M. "Spatial competition, monopolistic competition, and optimum product diversity". *International Journal of Industrial Organization* 17, 953–963.
- Wilbur, K.C. (2004a). "Modeling the effects of advertisement-avoidance technology on advertisement-supported media: The case of digital video recorders". Mimeo. University of Virginia.
- Wilbur, K.C. (2004b). "Not all eyeballs are created equal". Mimeo. University of Virginia.
- Wildman, S.S., McCulloch, R., Kieshnick, B.D. (2004). "A new look at the economics for television ad time". Mimeo. Michigan State University.
- Wiles, P. (1963). "Pilkington and the theory of value". *Economic Journal* 73, 183–200.
- Wright, J. (2002). "Access pricing under competition: An application to cellular networks". *Journal of Industrial Economics* 50, 289–315.
- Wright, J. (2003). "One-sided logic in two-sided markets". *Brookings Working Paper* 03-10.

## Further reading

- American Association of Advertising Agencies and the Association of National Advertisers, Inc. (2001). 2001 Television Commercial Monitoring Report.
- Anderson, S.P., Coate, S. (2000). "Market provision of public goods: The case of broadcasting". NBER Working Paper 7513.
- Berry, S., Waldfogel, J. (2003). "Product quality and market size". NBER Working Paper 9675.
- Brown, A., Cave, M. (1992). "The economics of television regulation: A survey with application to Australia". *The Economic Record* 68, 377–394.
- Busse, M., Rysman, M. (2005). "Competition and price discrimination in yellow pages advertising". *RAND Journal of Economics* 36, 378–390.
- Caplin, A., Nalebuff, B. (1991). "Aggregation and imperfect competition: On the existence of equilibrium". *Econometrica* 59, 25–59.
- Chakravorti, S. (2003). "Theory of credit card networks: A survey of the literature". *Review of Network Economics* 2 (2), 50–68.
- Coase, R. (1966). "The economics of broadcasting and government policy". *American Economic Review* 56, 440–447.
- Cornes, R., Sandler, T. (1996). *The Theory of Externalities, Public Goods and Club Goods*, second ed. Cambridge University Press, Cambridge.
- Crawford, G.S. (2000). "The impact of the 1992 Cable Act on household demand and welfare". *RAND Journal of Economics* 31 (3), 422–449.
- Davies, G. et al. (1999). "The future funding of the BBC". Report of the Independent Review Panel, Department of Culture, Media and Sport.
- Doyle, C. (1998). "Programming in a competitive broadcasting market: Entry, welfare and regulation". *Information Economics and Policy* 10, 23–39.
- Ekelund, R., George, F., Jackson, J. (1999). "Is radio advertising a distinct local market? An empirical analysis". *Review of Industrial Organization* 14, 239–256.
- Evans, D. (2003). "The antitrust economics of multi-sided platform markets". *Yale Journal on Regulation* 20 (2), 325–382.

- Gabszewicz, J.J., Laussel, D., Sonnac, N. (2006). "Competition in the media and advertising markets". *Manchester School* 74 (1), 1–22.
- George, L., Waldfogel, J. (2003). "Who benefits whom in daily newspaper markets?" *Journal of Political Economy* 111 (4), 765–785.
- George, L., Waldfogel, J. (2004). "Does the New York Times spread ignorance and apathy?" Mimeo. Michigan State University.
- Hackner, J., Nyberg, S. (2001). "Advertising and media market concentration". Mimeo. Department of Economics, Stockholm University.
- Holden, S. (1993). "Network or pay-per-view?" *Economics Letters* 43, 59–64.
- Kind, H.J., Nilssen, T., Sjørgard, L. (2004). "Advertising on TV: Over- or underprovision?" Mimeo. Bergen.
- Klein, B., Leffler, K.B. (1981). "The role of market forces in assuring contractual performance". *Journal of Political Economy* 89, 615–641.
- Minasian, J. (1964). "Television pricing and the theory of public goods". *Journal of Law and Economics* 7, 71–80.
- Motta, M., Polo, M. (1997). "Concentration and public policies in the broadcasting industry: The future of television". *Economic Policy* 25, 294–334.
- Nelson, P. (1974). "Advertising as information". *Journal of Political Economy* 82, 729–754.
- Noam, E.M. (1987). "A public and private-choice model of broadcasting". *Public Choice* 55, 163–187.
- Owen, B.M., Wildman, S.S. (1992). *Video Economics*. Harvard University Press, Cambridge, MA.
- Papandrea, F. (1997). "Modelling television programming choices". *Information Economics and Policy* 9, 203–218.
- Picard, R. (Ed.) (1988). *Press Concentration and Monopoly: News Perspectives on Newspaper Ownership and Operation*. Alex Publishing Corporation, Norwood, NJ.
- Samuelson, P. (1958). "Aspects of public expenditure theories". *Review of Economics and Statistics* 40, 332–338.
- Schwartz, M., Vincent, D. (2001). "Same price, cash or credit: Vertical control in electronic payments networks". Mimeo. Georgetown University.
- Shachar, R., Anand, B. (1998). "The effectiveness and targeting of television advertising". *Journal of Economics & Management Strategy* 7, 363–396.
- Shapiro, C. (1980). "Advertising and welfare: Comment". *Bell Journal of Economics* 11, 749–752.
- Sonnac, N. (2002). "Monopole televisuel et publiphobie". *Recherches Economiques de Louvain* 68 (4), 513–526.
- Stegeman, M. (2004). "A model of Internet advertising". Mimeo. Virginia Polytechnic Institute.
- 1999 Television Commercial Monitoring Report. American Association of Advertising Agencies.
- Tirole, J. (1988). *The theory of Industrial Organization*. MIT Press, Cambridge, MA.
- Vaglio, A. (1995). "A model of the audience for TV broadcasting: Implications for advertising competition and regulation". *Rivista Internazionale di Scienze Economiche e Commerciali* 42 (1), 33–56.
- Wildman, S.S., Bruce, O. (1985). "Program competition, diversity, and multichannel bundling in the new video industry". In: Noam, E. (Ed.), *Video Media Competition: Regulation, Economics and Technology*. Columbia University Press, New York.
- Wright, D. (1994). "Television advertising regulation and program quality". *Economic Record* 70, 361–367.

## THE MOVIES

ARTHUR DE VANY\*

*Department of Economics, University of California, Irvine, USA*

### Contents

Abstract	617
Keywords	617
1. Introduction	618
2. The essential economics of motion pictures	623
2.1. Nobody knows	623
2.2. Discovering demand	624
2.3. Adapting supply	624
2.4. Pricing films	625
2.5. The rental contract	627
2.6. Pricing hits	627
2.7. Clearances	628
2.8. The distribution fee	629
3. A brief history of motion pictures	630
3.1. The feature motion picture	630
3.2. The emergence of vertical integration	632
3.3. The Paramount litigation	632
3.4. Contracts and practices	633
Admission price fixing and price discrimination	633
Runs and clearances	634
Formula deals, master agreements, and franchises	634
3.5. The legacy of Paramount	635
4. Graduate movie economics: The stable Paretian hypothesis	637
4.1. Estimating stable Paretian parameters	639
4.2. Laws of the box office	641
“Winner takes all”	641
Hits have long runs	642

\* Without implicating him for my oversights and errors, I wish to thank Victor Ginsburgh for his superb editing. My thanks also to Simon Anderson for his helpful comments. The Private Enterprise Research Center of Texas A&M University supported my research.

*Handbook of the Economics of Art and Culture, Volume 1*

*Edited by Victor A. Ginsburgh and David Throsby*

*Copyright © 2006 Elsevier B.V. All rights reserved*

*DOI: 10.1016/S1574-0676(06)01019-2*

Non-linear information flow	642
Opening gross does not predict total gross	642
Momentum and chaos	642
Champions, births and deaths	643
Extreme events dominate	644
Volatility	644
Self-similarity	645
Non-finite moments	645
“Nobody knows”	647
4.3. Laws of production and cost	647
Sequential production	647
Production problems	648
End stage	648
Hold ups	648
Solutions	649
Contracts and roles	649
Final authority	650
Stable Paretian costs	650
Returns on budgets	651
The angel’s nightmare	652
A strange law of production	652
Artist productivity	652
4.4. Laws of profit	653
Dominance of extreme events	654
Stable profit laws	655
Returns	656
4.5. Stochastic market structure	656
4.6. Pricing talent	657
Extreme events and contracting in the tail	657
Stars, profits and the curse of the superstar	658
Bidding wars and the producer’s curse	658
Is an actor really worth \$20 million?	658
4.7. Decisions: Greenlighting, bias, and sure things	659
Ratings and genre	659
Aggregation bias	659
Ambiguous probabilities and sure things	660
Inside and outside odds	660
5. Conclusion	660
Appendix A	661
References	664



**Abstract**

This chapter is an overview of a new kind of economics of the movies; it also is my attempt to lay a new foundation of the economics of art and culture. The essence of cultural goods is that they are creative goods that have no natural limit on their consumption or dissemination; they are information goods. And they are wildly uncertain. I show how this vision may be implemented in a rigorous and insightful way in the study of the movies. A centerpiece of the analysis is the stable Paretian hypothesis and its usefulness as a model of motion picture revenues, costs, and returns. The industry's organization, contracts, pricing, and compensation deals are also seen as rational adaptations to the uncertainty captured by the stable Paretian probability model.

The essence of the stable Paretian model is that the probabilities of motion picture outcomes are far from Normal. The tails of the stable Paretian distribution are "heavy" and large-scale events are far more probable in a Paretian than in a Gaussian world. The large events far out on the probability tails dominate sample statistics. The variance is infinite and, for some variables, even the mean does not exist. Movie box office revenues, therefore, have no natural size or scale and there is no typical or average movie; each is unique unto itself. Revenue and cost dynamics are complex and expectations of cost or revenue at level  $X$  are proportional to  $X$ .

**Keywords**

uncertainty, contracts, adaptive optimization, movies, extreme events

*JEL classification:* C1, D0, D8, G0, G3, K2, Z1

## 1. Introduction

The movie business is hard to understand. It defies conventional economic models on many levels. But, this is what makes it so interesting. I see the movie business as a laboratory for developing a new kind of economics of art and culture. It is an extraordinarily well-documented business, so there is a rich data base to be exploited by researchers. Where else can you follow a product from birth to death in a matter of weeks or months? Where else can you study the evolution of its revenues and market penetration in such detail? And for so many products.

Studios get overnight reports of box office revenues. Weekly reports are available in industry trade publications, on internet sites, and on the evening news. There are at least three proprietary data sources where the researcher can follow statistics on a weekly or annual basis.<sup>1</sup> And these data sources go back for years.

Detailed filmographies exist of casts, producers, and directors for virtually every film ever released in North America. Because screen credits are so important in the movies, filmographies are highly accurate. Their accuracy is insured because craft unions have elaborate criteria for awarding screen credits. Where else can you find so widely distributed such a careful documentation of creative work and rigorous delineation of creative credits?

The industry requires all this information about box office returns and reputation in order to solve the many vexing problems the industry faces. I think the first order of business for research is to recognize that the movies are an information industry, arguably the first of the twentieth century. The industry produces information, for all that a film really is, and it lives on it. Without this elaborate reporting of film revenues and reputations, the industry could not function. The reporting is required because nobody knows what a film will gross so the industry has to have in place the instruments to adapt flexibly adjusting prices and supply to demand when the audience reveals it. Reputation is crucial when artists and creative inputs are hired for short periods of time and each can become the critical input at some stage of production.

Movies are one-off products that only go around once. You do not know if you will like a movie until you see it. If you like it, you tell your friends about it. This simple act of sharing information leaves no less of the movie for you to enjoy and may even increase your pleasure. This means information can be multiplied.

If information can be multiplied, motion pictures will have increasing returns. Movies that have many viewers will grow more rapidly than movies that have few viewers. The supply of theater seats is perfectly elastic because the run can be extended to increase supply until it equals demand. This means box office revenue is just a number, with no natural limit or scale. It turns out that scale-free distributions capture the statistics of

<sup>1</sup> A.C. Nielsen/EDI (my primary source), Exhibitor Relations, and Paul Kagen are three proprietary data sources. Industry publications, such as *Variety*, *Hollywood Reporter*, and *Box Office*, report on the industry prolifically. The Internet Movie Data Base is a good source for filmographies and various data as well.

movies with high fidelity. Because information does not obey the usual laws of production, the statistics of the movies are rather strange: distributions of variables are sharply peaked and skewed, with a long tail to the right. They are scale-free and high on kurtosis.

The need to manage and adapt to information leads to interesting business practices and contracts. The industry is filled with colorful phrases and unusual contracting practices like “four-walling”, “road shows”, “pay or play”, “house nut”, “legs”, “blind-selling”, and “block-booking”. It is an extremely hierarchical business, with a few movies, stars and directors standing far above the rest in box office grosses. It is frighteningly risky, yet the rate of return is low. Most movies lose money and a handful make nearly all the profit the industry earns.

Movies are produced in a novel way too. Producers assemble temporary teams from a large pool of artists and craftsmen to make movies. These production teams live only for the duration of the project and disband when the movie is finished. At any point in time, just a few people in the industry are actually working. But, a few well-connected actors and directors are very busy and they make most of the movies that dominate box office revenues.

Most movies take several years to produce. There are many pitfalls along the way and the final product is nothing more than a few reels of celluloid. All its production costs are sunk by the time a movie is finished and ready to be released. They may not be recoverable and a finished movie might never get a chance to be seen on a theater screen. Cost is notoriously difficult to forecast and some movies famously go far over budget. A studio can go broke if it goes just one year without a hit. And movies are financed in unusual ways.

These are hard problems, but the hardest one of all is that nobody really knows how much a movie will gross at the box office. If people in the industry did know, there would be no need for its strange practices and they would disappear. It is because “nobody knows” that the movie industry is so fascinating and so different from dentistry or warehousing. It does have analogues though in other creative activities that deal in information such as publishing, music, stage, and patents. The emphasis on complex dynamics and their statistical basin of attraction that has been the basis of a good deal of the work on the movies reviewed in this chapter would appear to have a profitable use in other industries.<sup>2</sup>

“Nobody knows” is the core problem of the movie business. This is the problem almost all of the industry’s colorful and strange practices, contracts and business organization are designed to solve. It is the source of its preoccupation with rank and power and relationships. It elevates “the deal” to its mythical status in the industry.

<sup>2</sup> For example, F.M. Scherer (2000) has used the Bose–Einstein process David Walls and I used to capture motion picture dynamics and the heavy-tailed distributions to which they are drawn to explain the dynamics and size distribution of patent royalties. D. Sornette and D. Zajdenweber (1999) have shown that the innovation time series contains the same ragged profile typical of the movies and other Levy Stable motions.

In this chapter, I shall try to use the ideas of sequential contracting and uncertainty in cost and revenue to explain the movie business. In the process, I will review some of the literature on the movies, but my intention is to reveal what light these concepts shed on these issues rather than to survey the literature. There are three bodies of research on the movie business; the old, the new, and what I call the new–new.

The old research came out of the decade of the antitrust litigation that eventually led to the dissolution of the great studios (and to much antitrust involvement in the movie business in other countries beyond the United States). The approach was to apply conventional antitrust economics to the industry, seeing virtually all its hard-to-fathom practices as attempts to exploit market power. This research is exemplified by Michael Conant's (1960) book, by the testimony in the *Paramount* litigation [Federal Trade Commission (1965) or the more recent book by Litman (1990)]. The bulk of this research did not see the movie business as an information industry and by incorrectly applying models suited to industrial products to this new industry they failed to see the efficiency of the industry's contracts and institutions. This research often confuses demand with supply, as when Seagrave (1997) says that Hollywood "dominates" the world's movie screens. Large openings may temporarily dominate theater screens, but that is supply, not demand. If demand does not fill seats, theaters quickly drop a film for something more promising.

The old motion picture research has been tested and found wanting. In a detailed analysis of the *Paramount* charges, De Vany and Eckert (1991) showed that the industry's organization and contracts were clever and necessary adaptations to the nobody knows principle. De Vany and McMillan (2004) used stock market prices to show that the forced divestment of theaters by the studios was a major financial blow to the industry and that the pattern of losses among firms was inconsistent with the idea that the industry had functioned as an implicit price-fixing cartel (the primary basis of the court's judgment). De Vany and Lee (2001) have shown that market shares are wildly volatile and unstable and that the industry is concentrated, as all information industries are, but the players at the top turnover at a high rate.

The new motion picture research is a modern variant of the old. A number of researchers have been drawn to study the industry (which is good), but have brought with them tools that are, essentially, game-theoretic versions of the old models. Papers have appeared that model the choice of opening date as a strategic game among studios,<sup>3</sup> entry and exit decisions of theaters are modeled as a location game, and rental price differentials among theaters have been described as (centralized) price discrimination (when they are determined in a completely decentralized fashion by the audience under the terms of the exhibition license). There is, of course, an element of truth in each of these points of view; studios do try to avoid opening against a blockbuster. But, no one knows half of what the models posit of the agents, which is more a criticism of the large

<sup>3</sup> See Krider and Weinberg (1998). Liran Einav (2004) has made some progress in empirically modeling motion picture release timing.

information requirements of standard models than of the research. Say, for example, you move your opening two weeks later to avoid going against *Titanic* only to find that *Titanic* is just getting its legs and completely flattens your movie. The models (at least those I have seen) do not correctly portray the uncertainty the agents face, and it can be shown that models that are faithful to the complexity of the problem are computationally intractable. I think the nobody knows principle cuts sharply against this line of research. But, this is good for economics and will eventually lead to progress.

The motion picture industry challenges equilibrium models and existing mathematics in fundamental ways that will require a new kind of modeling. The highly decentralized and adaptive mechanisms employed by the industry are difficult to model in the centralized, strategic mindset typical of optimizing models. I believe capturing the “wild” uncertainty and adaptive decision processes of creative industries provides a more rigorous and insightful basis for a new-new economics. By bringing the “wild” uncertainty of creative enterprises to the forefront, the work I report on here lays a foundation for a New Economics of Art and Culture

A hint of the formidable challenge of understanding “wild” uncertainty in motion pictures and the way the movie business relies on adaptive and decentralized mechanisms is developed in work I and colleagues have been doing. This work is a continuation of ideas originally developed by Mandelbrot (1963a, 1963b) and Simon (1955). The key ideas are that production is sequential and that there is a “wild” kind of uncertainty in cost. Motion picture revenue (nobody knows) unfolds during the run in a complex and recursive dynamic that can “go anywhere”, but the dynamics converge on a statistical attractor, the Lévy stable distributions with “heavy” (non-Gaussian) tails. These properties set the information discovery and contracting problems that are at the core of the industry’s structure, contracts, and organization.

This research is statistical in nature, emphasizing the complexity of motion picture revenue and cost dynamics and their basins of attraction. It is focused on a characterization of the sources and nature of the uncertainty which is then used to discover the problems to which the industry’s organization, contracts, and practices must be adapted. This line of inquiry was originated by De Vany and Eckert (1991) and pinned down by De Vany and Walls (1996) and famously named the “nobody knows” principle by Richard Caves (2000) in his *Creative Industries* [after screenwriter William Goldman’s line “With all due respect, nobody knows anything” in his *Adventures in the Screen Trade* (1983)]. Cassey Lee’s doctoral dissertation [Lee (1999)] was one of the first explorations of the stable Paretian hypothesis in motion pictures, which lies at the core of this line of work. Other papers have appeared in this vein recently by Chris Hand (2001), Collins, Hand and Snell (2003), and Jordi McKensie (2003). John Sedgewick’s (2001) book on British filmgoing during the 1930’s is also very much in this vein.<sup>4</sup> This is the

<sup>4</sup> In David Walls’ (2002) review of Sedgewick (2001) he shows that the distribution of Sedgewick’s index of film popularity is a Pareto law of the form found for the distribution of motion picture revenue by the above cited authors. I will show below that this is an example of the self-similarity of statistical distributions implied by the stable Paretian hypothesis.

line of research I shall emphasize in this chapter because I believe it is foundational to the development of a New Economics of Art and Culture.

In Section 2 I begin with a discussion of the essential economics of the business. This is Movie Economics 101 and its aim is to identify the main issues that confront the business and show how they are solved; you will see that the movie business shares many features and practices of other creative industries for they are all information industries. Following this, in Section 3, I apply these ideas to a brief historical analysis of the movie business. The purpose here is to show how the industry's methods of production, its structure and regional distribution, and its exotic contracts and practices evolved historically as solutions to many of its more vexing problems. You will see that the judges and lawyers relied on concepts more suited to toasters and bricks than to one of the first information industries – the movies. In this section, I also cover the Paramount antitrust cases and show how the courts reshaped this nascent information industry in the mode of the old economics of antitrust.

Section 4 covers the graduate level course in movie economics. It begins with the laws of the box office and with a discussion of the Lévy stable distribution. (Since we will often be interested in the upper or lower tails of this distribution and these tails are Pareto distributions, I shall freely interchange the terms stable Paretian distribution and Lévy stable distribution, letting the context indicate if the tail or whole distribution is at issue.) The laws of the box office are those of a winner-takes-all contest. The laws are driven by a non-linear information dynamic that takes movies that are close to one another and propels them apart at exponential speed. Under the influence of these recursive and non-linear processes, movie revenues can “go anywhere”. They may even be chaotic. The extreme skew of the revenue distribution, the influence of extreme events, and the unstable and non-finite moments provide a rigorous basis for the nobody knows principle. I show that the stable Paretian model captures with high fidelity all the essential features of the statistics of the business. I use the concept of statistical self-similarity to show the importance of extreme events in all facets of the industry from film revenues to budgets, productivity, pay, and careers. The influence of marketing and stars and the opening are discussed here in terms of the non-linearity of the revenue dynamics of motion picture information. In this section I show how pay keys on extreme events and I reveal the laws of stardom.

Other topics taken up in the graduate course are the structure of the industry implied by the stable Paretian hypothesis and the evidence for it, the pricing of talent, and the many decision problems that are engendered by the wild statistics of the movies. I introduce the concept of stochastic industry structure and use it to show that concentration measures are not well-founded in this industry. I show the industry is stochastically concentrated, but intensely competitive. Leaders turn over at a high rate and the Hershman–Hirfindahl index, a gold standard for antitrust, does not even exist. Then I discuss the difficulties of making decisions in a business where the probabilities are non-Gaussian and confound expectations. The stable Paretian probabilities lie somewhere between risk and uncertainty and they lead to many decision errors. I explain the “curse of the superstar”, the “angel’s nightmare”, the “sure thing principle”, and the

“producer’s curse” as examples of the many decision paradoxes faced by movie makers. In Section 5, I reach some conclusions.

Naturally, this chapter is a somewhat personal view of the business and what I think is interesting and important about it. Much of the material draws on my research in many articles and from my book *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry* [De Vany (2003b)]. This is not a review of research. Rather it is an exposition of a modeling strategy and a research agenda that I have found to be productive and informative. In my own work on this industry I have found that all the complexity of economics can be found there and there is much to explain that stretches the boundaries of current economic theory.

Much of economics focuses on averages, expectations, and standard deviations. Economists focus so intently on these measures that they often toss out “outliers” to eliminate their influence. This might be right for toasters or trousers, but the movies are not like that. The statistics of the movies are “wild” and utterly non-Gaussian. The outliers are the main event, not something to be tossed out, and averages are just numbers signifying nothing. Motion picture revenues are just a number – pure information – and have no natural scale or limit. Expectations and higher moments of the distribution need not even exist. One of the key lessons of trying to do science in such a non-standard industry is that we have to take the focus away from explaining particular events (why movie A did well and B did not) to understanding process and pattern; the ultimate object of a science of the movies is the probability distribution and the dynamic process that produces it.

## 2. The essential economics of motion pictures

This should go quickly as I have already alerted you to many of the key issues. On the other hand, there are some that have to wait for the graduate version of movie economics to come later.

### 2.1. *Nobody knows*

I still like the way the fundamental property of the movies was put by a wise District Court Judge (whose smart decision was overturned by the Supreme Court).

Moving picture films are a fluctuating and uncertain product. Until a film has been exhibited no one knows or can accurately estimate its value as a box-office attraction, either as a first-run exhibition or a subsequent run exhibition.<sup>5</sup>

<sup>5</sup> *United States v. Griffith Amusement Co.*, 68 F. Supp. 180, 196 [W.D. Okla (1946), cited in De Vany and Eckert (1991)].

Economists would put this by saying that movies are experience rather than inspection goods. You have to see a movie to know if you like it or not. What this means is that demand is revealed by the audience when the film opens. Audiences discover what they like, they do not reveal preferences they already have. Once demand is discovered, supply and pricing must adapt.

## 2.2. *Discovering demand*

To get the discovery process underway, films have to be exposed on screens. Throughout a film's run, audiences discover and exchange information about movies. Box office reports are signals to distributors and exhibitors about the audience's likes and dislikes.

A studio/distributor can sample this information in a variety of ways through the design of its release. A wide release on many screens simultaneously gathers a large sample of information, but leaves few degrees of freedom in responding to it. It leaves less time for word of mouth and other information to influence viewers. The film may disappear too quickly for word to spread and for audiences to find it. In a smaller release on fewer screens, a studio is sampling sequentially through the audience, discovering its demand as information flows over a longer period of time.

## 2.3. *Adapting supply*

Once demand is discovered, prices and supply adapt. There are three prices relevant to a movie: the admission price, the rental rate, and the distribution fee. The admission price does not vary much, though it may differ among theaters. The rental rate is highly variable. So is the distribution fee. These will be discussed below.

The main supply response is in extending the run, keeping the film longer in theaters. This is particularly true of a wide release. In a smaller release, prints may be ready or can be made quickly enough to expand the number of theatrical engagements; thus, both the length of the run and the number of theaters can be changed to respond to demand. The later engagements have the benefit of the demand information revealed in the earlier bookings, so the sequential sampling, smaller release reduces risk for the exhibitors who show the film later in the run. The large, wide release is more risky for exhibitors as they have no prior information to go on. The big release is likely to be more risky for the studio too because the number of prints to be made has to be decided before demand is known and because supplying more prints means more is spent before demand is known.<sup>6</sup>

So, demand is discovered and supply adapts, primarily by extending the length of the run. This goes on in real time using overnight and weekly box office reports. The decision to extend the run is a matter of contract which is delegated to the individual theater

<sup>6</sup> A print is a role of film and can cost thousands of dollars. Print and advertising costs are all expended up front and may be from 25 to 40% of a movie's costs.



under the holdover clause. The holdover clause is a contingent term of the exhibition license that specifies an amount of box office revenue which, if equaled or exceeded in the prior week, extends the run another week. The holdover amount is unique to each theater and film, negotiated as part of the license. The clause decentralizes the decision to extend the run because it relies only on local information of time and place, i.e. conditions at the theater where it is running. If box office revenue exceeds the holdover, the distributor is obligated to leave the film another week and the theater is obligated to run it another week. There is no limit to the number of times a film may be held over so a theater might run a movie for twenty or more weeks if it is grossing well.

#### 2.4. *Pricing films*

Motion pictures have three different prices. Most people think of the “price” of a movie as what they pay at the box office. But the price the theater pays for showing a movie is more important to the motion picture distributor, for this is their primary source of revenue from the film; this price is the rental rate or percentage of box office revenues paid to the distributor for the right to exhibit the film. The third “price” is what the producer of the movie pays to the distributor for placing the film in theaters and collecting film rentals. These prices are interrelated: the admission price is the source of box office revenues to the theater which pays a rental rate to the distributor, who deducts from these rental receipts a distribution fee.

Since demand must be discovered in a way that lets supply and prices adapt, three problems arise with respect to setting all three prices of the film. First, the demand signal shouldn't be noisy. Second, the demand signal should convey information about demand at the point where price is to be adjusted. Third, prices should respond in some reasonably optimal way to the demand signal. I have come over the years to admire the industry's solution to the film pricing problem.

The admission price is fixed over the duration of the run, with passes and discounts being restricted in the earliest weeks of the run. With a fixed admission price, any variations in box office revenue are due entirely to the number of admissions. The studio, therefore, receives a pure quantity signal of demand, uncontaminated by variations in the admission price. If a minimum admission price were not specified, the exhibitor could, literally, cut the price to zero and make it up on concession sales, particularly for Disney animated movies and movies that appeal to a young audience. As a protection against this, exhibition licenses prior to the *Paramount* decision usually specified either a minimum admission price, or, deferring to the local knowledge of the exhibitor, specified a “usual, or customary” price. The courts ruled this practice to be unlawful price fixing and such price clauses were expunged from exhibition contracts. Virtually all theaters in the post-*Paramount* era became first-run theaters, so the almost ancient practice of charging more for the first than for second and subsequent runs became extinct. So, we find that admission price does not vary over the course of a motion picture's run, though some theaters may charge different prices than others.

The stability of admission price among different movies and over the course of the run has puzzled economists and observers of the motion picture business. This is such a durable practice that it must have some enduring properties; if this pricing arrangement were not maximizing revenue, then it would change, according to economic principles. But, aside from its historical durability, there are economic reasons to think the practice is rational.

Consider first how one would decide how to price a studio's motion pictures. Would a Warner Brothers movie be priced differently from one produced by Twentieth Century, Paramount or Columbia? Since studios no longer produce a certain style of motion picture, there is no brand identity for differentiating their products. At one time, MGM had a brand name and its movies could command a premium price. But, that era is gone and all the studios produce a wide range of films with no evident or predictable studio style or genre; nor do they have particular stars under contract. No studio has a lock on blockbusters and they all make bombs. Moreover, every studio would strongly object to having its movies priced more cheaply than others; it would be seen as a signal of inferior quality.

But, then why not price the best movies more dearly? If only you knew ... The "nobody knows" principle, developed at some length later, says that you don't know which movies will be hits and which will be misses. You only know after a movie runs that demand is so high that a premium price could be charged. Setting a higher price before a movie is released gambles with the information dynamics. Choking off demand early in the run curtails the non-linear building of demand from week to week. You will see later that hits have distinctly different runs from conventional movies and they build demand over time. Charging a higher price in the early weeks of the run may prevent that dynamic from getting off the ground. A studio really only knows that it has a hit on its hands four or more weeks into the run, and then it is too late to change the admission price, and illegal for it to direct the exhibitor to do so. Once it is understood that supply can be expanded in a perfectly elastic way, then one sees that high demand can be fully captured by extending the run, without raising the admission price. A side benefit of an inflexible admission price is that the exhibitor cannot game the rental rate by varying his admission price.

Now that *Paramount's* restrictions have made virtually all theaters first run theaters, it is no longer possible to price the first run higher than the second and subsequent runs. Nor is it likely that limiting the film's initial run to a few theaters to give them a measure of exclusivity and pricing power would survive a legal test under *Paramount's* holding that excessively wide clearances are unreasonable restraints.

The simple answer to the question of why do prices not adjust to high demand is that they do, you just have to look at the prices that really count for the studio and the producer. The rental price adjusts almost perfectly to demand and the distribution fee does as well. The rental price depends directly on the grosses at each theater where

the film is playing.<sup>7</sup> This provision keys the price to local conditions, because the local rental rate depends wholly on the local theater's admissions. So, price is low where local conditions warrant and high otherwise; this gives a close-to-optimal spatial variation in prices that extracts more revenue from each film.

The solution to the problem of varying the rental price in accordance with demand is achieved through the contingent pricing clauses of the rental contract. The exhibition license cleverly specifies a two-part pricing rule that is adaptive and sufficient to capture most of the value of the film to the exhibitor. Note, this variable pricing component too is decentralized as the license sets a unique price to each exhibitor conditional only on her theater's grosses. The rental is determined by a formula in the contract. Let us consider the sophisticated pricing instruments contained in the motion picture rental contract and how the rental price is adjusted in accordance with demand.

### 2.5. *The rental contract*

First, there is a guarantee or fixed amount which the exhibitor must pay no matter what the film grosses. This is individually negotiated or, if the distributor awards licenses by taking bids, it will be part of the exhibitor's bid. For many films, the guarantee will be small or zero. For a number of heavily demanded films, it may be rather high (above \$100,000). We know a two-part price is sufficient to capture all the exhibitor's surplus; the guarantee is the fixed part of the two-part price.

But, neither the distributor nor the exhibitor knows what the film will gross, so the guarantee cannot be set optimally in advance. Adjusting price when demand is revealed is the task left to the variable terms of the rental contract. The rental is a percentage of the exhibitor's gross. It is reset each week according to a declining schedule. The exhibitor pays a minimum percentage of her gross each week, with the minimum declining over the course of the run. A typical example would be a four week minimum run (part of the contract) where the minimum percentages might be 70, 70, 60, 60 with any weeks beyond at the flat rate of 40 or 35 percent. This declining schedule compensates the exhibitor for the rising probability that the film will die next week [De Vany and Walls (1997)] as the run progresses.

### 2.6. *Pricing hits*

There is more. Suppose the film takes off in its fifth week and begins to gross major money. By that week, the rental rate is only 35 or 40 percent and the exhibitor is making a killing. If a film gets "legs" it is likely to occur around the fourth or fifth week [De Vany and Walls (2003b)]. And it may run 20 to 40 weeks if it is a *Star Wars* or *Forrest Gump*. Runaway hits such as these generate the bulk of revenue and profit, so the studio

<sup>7</sup> At one time there were contracts that called for the theater to pay a rental based on a national gross rather than the theater's gross. See the discussion of these "formula deals" in the historical section below.

must extract a high rent from them; otherwise profits will be left with exhibitors who had nothing to do with producing the hit.

To guard against leaving money in the exhibitor's hands, the rental contract contains a rental escalation clause. This is the 90 percent on gross above the house nut rule:  $\text{Rent} = 0.90 \times (\text{Box Office Gross} - \text{House Nut})$ . The rent will be the minimum applicable percent of gross or 90 percent of gross minus the house nut, whichever is higher. The house nut is a fixed, negotiated amount; years ago, the house nut was an approximation to the exhibitor's weekly cost of operating his theater.

How important is this rental escalation? The sort of data required to answer this question are private, but I have seen data to indicate that the clause is very important in pricing breakaway movies. An example is *Tootsie's* 23 week run in a theater. The minimum weekly percentages for the contracted six week minimum run were 70, 70, 60, 60, 50, 50, with holdover weeks at 35 percent. The house nut was \$3000. The data for this theater show that *Tootsie's* effective rental rate was 70 percent the first week. Then it jumped to 90 percent for 10 straight weeks (less the house nut), dropped to the minimum 35 percent for 5 weeks and then jumped back in the 16th and 17th weeks to 90 percent after which it ran 5 more weeks at 35 percent.

Overall, *Tootsie's* rental rate in this theater was 54.29 percent. Had the 90 percent over the house nut clause not been in the contract, *Tootsie* would have earned a rental rate of just 45.78 because it would have run 17 weeks of its 23 week run at 35 percent.

How much did the 90 percent clause add to *Tootsie's* rentals? We don't know for sure, but we can calculate an illustrative amount from the data for this theater. *Tootsie* played on 1222 screens. Its US gross was \$177.2 million. Suppose we take the difference in rental rates for this theater and apply it to *Tootsie's* national total box office revenue. Without the clause, the rental rate of 45.78 percent would have yielded total rentals of \$81.1 million. With the clause, the rental rate of 54.29 percent would have yielded total rentals of \$96.1 million, an increase of nearly 19 percent or \$15 million.

Lesson learned. *Tootsie* was the second-highest grossing film of 1982. It ran 23 weeks in the theater for which we have data and longer in other theaters. The laws of the box office, which shall see later, say that hit movies have long runs and they dominate total theatrical grosses and rentals. A distributor cannot afford to underprice the hits or leave them in theaters for long runs at the minimum rental rate. The 90 percent clause keeps this from happening, even when no one knows which movies will be hits.

## 2.7. Clearances

Distributors must give exhibitors some protection against competition. Otherwise, exhibitors will be reluctant to screen the movie. In addition, the distributor does not want the theater's box office take to be diluted among many theaters for this would lower rentals and shorten the film's run for its take would fall below the holdover. Finally, distributors would be reluctant to have exhibitors competing in admission price for that would reduce their rentals which, we have seen, depend directly on box office revenue.

The long-standing solution to most of these problems is the clearance clause of the rental contract.

A clearance is an agreement with the exhibitor that gives him a measure of exclusivity in the showing. The distributor agrees not to license the movie to specifically named competitors of the licensed exhibitor. The clearance is determined through bid or negotiation. A theater might include the clearance it wants in its bid. The bid letter might suggest a clearance. Alternatively, the clearance might be negotiated. Over time, clearances may tend to become more or less customary.

Broad or stable clearances tend to limit competition. For that reason, clearance arrangements are subject to court scrutiny. Granting a theater an exclusive franchise can be interpreted as a grant of a long-run clearance to the franchised theater against its competitors. When a distributor owns a theater the effect is to clear the territory of competing exhibitors. This is a controversial area because the Supreme Court found clearances that were too wide or routinized to be violations of the antitrust law.<sup>8</sup>

### 2.8. *The distribution fee*

This is the third of the prices we are concerned with. Here too, the nobody knows principle applies: since nobody knows how much a film will gross, it is impossible to set an optimal fee for distribution service. The distributor, a studio usually, books the film with theatrical booking agents, advertises, distributes prints to theaters and collects them after the showing for distribution to other theaters. Most importantly, the distributor collects rental payments and monitors exhibitors to verify they are showing the movies booked and at their proper times and that rental payments are properly correlated with box office revenues.<sup>9</sup>

If a movie has a long run, the distributor is continuously performing these duties over an extended time frame. No one knows how long a movie will run or what it will gross. So, it is impossible to gauge the distributor's cost or the value of the distributor's productive efforts. That is, until the film has run; after that, all these things become clear.

The industry's long-standing solution to these, and other issues, is to charge a conditional distribution fee that depends on the outcome. The typical arrangement is a fee around 30%, but the rate is negotiable. This fee not only addresses the nobody knows problem, it also better aligns the distributor's interests with those of the producer, who is really the residual claimant or owner of the movie. A proportional distribution fee gives the distributor an incentive to maximize rentals and, since rentals are the source of revenue ultimately to the producer, their interests are made more compatible. If the

<sup>8</sup> *United States v. Paramount Pictures, Inc.*, 334 U.S. 131 (1948).

<sup>9</sup> There are detective firms for hire to detect theater cheating because the theater pays rentals based on what they report their box office revenues to be. There had been in the past a good deal of underreporting, but this appears to have lessened with modern ticket machines and theater chains who have long-term relationships to preserve [see De Vany and Eckert (1991)].

distributor received a fixed fee, the incentive to maximize rental collections would be muted.

Because it is a percentage of rentals, the distribution fee varies directly with movie's success. So, the key prices are quite flexible and their levels are conditioned on the state of demand as demand is revealed over the course of the run.

### 3. A brief history of motion pictures

We have spent some time on the arrangements between distributors and theaters because they are at the heart of the industry's problem: how do you discover demand, adjust supply, and price movies when nobody knows their value? This problem is as old as the feature motion picture. And the solution to it has not changed much since the movies began. In fact, many features of the industry today resemble those at the inception of the industry. It is useful to look back briefly to see how little has changed, including nearly all the terms of the exhibition license, for it reveals those features of the industry that are deep and immutable and those that are superficial.

In the 1890s and 1900s, producers sold copies or prints of motion pictures outright to theater owners.<sup>10</sup> Exhibitors resold their prints when they no longer wished to show them. By 1903, organized film exchanges supported print trading among exhibitors and eventually the exchanges became rental agencies to reduce the cost of transferring prints. Most of these pictures were low-budget one-reelers: 10-minute comedies, 15-minute dramas, 5-minute scenic films, and 20-minute westerns. Four to eight of them comprised the typical two-hour program for which patrons paid a nickel. By 1911, about 150 exchanges supplied 11,500 theaters with enough reels to provide almost daily program changes.

Copying and bootlegging were common. The flat price that the producer received for selling a print to an exchange did not correlate with the picture's success in the market place. Revenues were too small to justify big productions. Releases had to be changed frequently to stimulate attendance. Without marketing, however, the public had scant information about which pictures to see. There were no great stars or studios in this period, and memorable pictures were rare. Exhibitors faced an unreliable stream of indifferent motion pictures by unrecognized producers and artists.

#### 3.1. *The feature motion picture*

The multi-reel feature picture, produced as early as 1907, became popular and widespread by 1913. It fundamentally changed the industry. Producers and exhibitors saw the value in moving from daily program changes to two or three features per week.

<sup>10</sup> In this section I rely on De Vany and Eckert (1991).

Demand would rise through word-of-mouth advertising, theaters could become more lavish and charge higher admission prices, and distributors could choose among better theaters. The major studios, great stars, and memorable features developed as the industry moved away from one-reelers to feature motion pictures.

Features cost more to make than one-reelers, so a nationwide distribution system was formed to lower the cost of distributing them. Higher cost productions required larger financial commitments from producers. To finance a picture, a producer had to show that a distributor would handle it and exhibitors play it. The ability to contract “forward” for the distribution and exhibition of a feature before it was produced was essential to financing. Forward contracting was also the genesis of vertical integration.

At first producers sold exclusive rights to their productions by territory. These were called “states’ rights” deals because they encompassed one or more states, although they were often for only part of a state. Copies of the motion picture were sold to a states’ rights distributor for a flat fee that was based on the territory’s population. As they gained experience, however, distributors learned that population was an unreliable estimator of demand and they began to charge a rental fee, or royalty, that was based on a percentage of the exhibitor’s box office revenue. This arrangement was also acceptable to the exhibitor since it was less risky than paying a flat rental for the feature no matter how well it did at the box office. However, for the royalty to correspond with demand, the exhibitor’s box office report had to be truthful. Anticipating vertical integration, the typical area or states’ rights distributors were exhibitors who would show the picture in their own theaters, and then lease it to other exhibitors in the allowed territory for later runs. These later runs were for designated periods and sometimes were exclusive.

By 1913–1914, producers licensed blocks of a full season’s production (perhaps thirty pictures) to a distributor in a single agreement. Paramount Pictures Corporation, formed in May 1914 by the merger of five exchanges and regional distributors, secured the franchise to distribute all the features of three of the leading production companies. The producers agreed to make a number of features each season for Paramount to distribute. In return, Paramount agreed to pay the producers 65% of the gross revenue it received from theater rentals and sub-licensing to other distributors. Paramount retained 35% as its distribution fee, an arrangement that is retained to the present day. Paramount guaranteed minimum returns to the producers and advanced cash to them on each features release date, against which film rental payments received from exhibitors were credited.

To secure exhibitors for the features they financed, distributors granted franchises to exhibitors. These franchises gradually replaced states’ rights deals. By taking a distributor’s franchise, an exhibitor agreed to show the distributor’s entire line of motion pictures. As *Cassady (1959)* noted, “this early version of block-booking developed out of the need of distributors for a more efficient method of selling films”. The franchising exhibitor also bought features “blind”, before they were seen, or before they had even been made.

### 3.2. *The emergence of vertical integration*

By 1916, features predominated over short films, and movie “palaces” had been built. But, producers and distributors wanted stable outlets and exhibitors wanted assured supplies. The contractual arrangements on which they relied were a more uncertain instrument to that end than vertical integration. Vertical integration of distribution with exhibition offered outlets for distributors, who could then guarantee play time to the producers with whom they contracted. With guaranteed distribution and exhibition, producers were better able to finance their features.

The impetus for integration came from all links of the production, distribution, and exhibition chain: distributors merged upstream and downstream; theater chains merged upstream through distribution to production; producers merged downstream through distribution to exhibition. Paramount the distributor became a producer and also built theaters. Fox and Loew’s entered production from the exhibition side of the industry. The producers Warner and Goldwyn formed theater chains and also developed distribution facilities. By 1931, Paramount owned nearly 1000 houses and its own studios, and the six largest circuits owned 2437 theaters (one-eighth of all theaters).<sup>11</sup>

The driving force behind these changes was the introduction of the high-cost feature motion picture. Within fifteen years of its introduction, the structure and controversial practices that the government would later attack had emerged: exclusive territorial licensing, a sequence of runs, block-booking, franchising, price stipulations, and vertical integration and become the industry norm.

### 3.3. *The Paramount litigation*

The five major defendants in *United States v. Paramount Pictures, Inc.*<sup>12</sup> were Loew’s (renamed MGM), Paramount, RKO, Twentieth Century-Fox, and Warner Brothers. Each of the major studios was vertically integrated from production through distribution and exhibition. The three minor studio-defendants were not fully integrated: Columbia and Universal produced and distributed motion pictures; United Artists distributed only. The court found that the defendants conspired to fix minimum theater admission prices; engaged in intertemporal price discrimination by charging higher admission prices for first-run than for later showings; conspired to fix patterns of exclusive clearances and runs for neighborhood theaters; and restrained trade by licensing motion pictures in blocks. It found that the five major defendants also operated theaters monopolistically through joint ownership or pools.

In its 1948 *Paramount* antitrust decision, the Supreme Court relied on simple measures of box office revenue concentration among the major motion picture distributors to make its finding that the studios, acting together, monopolized the market for “first

<sup>11</sup> Conant (1960).

<sup>12</sup> 66 F. Supp. 323 (S.D.N.Y. 1946).



run” motion pictures. The Court’s decision indisputably altered the studio system, forcing the divestment of studio-owned theaters and altering the contracts through which motion pictures were licensed to theaters [De Vany and Eckert (1991)]. Unfortunately, the courts got it wrong.

A central piece of evidence the courts relied on to infer market power of the defendants was the combined market shares.<sup>13</sup> It is interesting to note that during the half century following the Paramount Consent Decrees, all but one of the distributors–defendants (RKO, who exited in 1957) continue to survive. The original distributors–defendants were Loews, Paramount, RKO, Twentieth Century-Fox, Warner Brothers, Columbia, Universal and United Artists. Loews and United Artists are now part of MGM while Columbia is now part of Sony Pictures. The new entry into this upper echelon is the Disney-owned Buena Vista.

The current market share of the top five distributors (at 72 percent) is similar to that of the market share of the five major distributors–defendants during the Paramount litigation (at 73 percent in 1943–1944). The current Eight Firm Concentration level at 95 percent is the same as during the period of the litigation.<sup>14</sup> So, little has changed in terms of revenue concentration in the business, a property related to self-similarity which we discuss in the next section.

### 3.4. *Contracts and practices*

The courts challenged five practices of the defendants; each of these practices highlights the challenges of discovering demand, adapting supply, pricing movies, and structuring incentives which remain central issues in today’s motion picture industry. I will briefly discuss each the challenges to show where the court went wrong.

#### *Admission price fixing and price discrimination*

The courts found that the defendant distributors were guilty of price fixing and price discrimination. Admission price fixing was inferred from the practice of clearing a zone of competing exhibitions and a clause in the contract that stipulated the exhibitor’s admission price. Film rentals were then, as well as now, determined by box office receipts. An exclusive run limits rivalry so that the theater’s profit maximizing price will be closer to the price that will maximize rentals. The license exploited the exhibitor’s informational advantage by stipulating that the usual or customary price was to be the minimum admission price, leaving the exhibitor the (seldom exploited now, but historically important) option of raising the price for motion pictures the exhibitor thought would be in high demand.

<sup>13</sup> *United States v. Paramount Pictures, Inc.*, 85F. Supp. 881, 894 (S.D.N.Y. 1949). See Conant (1960) and De Vany and Eckert (1991). During the 1943–1944 season, the eight defendants had received 95 percent of domestic film rentals (Westerns excluded).

<sup>14</sup> Values of CR5 and CR8 for 1943–1944 are computed from *United States v. Loew’s Inc.*, 334 U.S. 131. Final Finding of Fact 100, February 8, 1950. See Conant (1960, Table 13, p. 46).

Intertemporal price discrimination may have been tempting for distributors to try and there is little to object to about it, since if it increases film rentals it will expand the supply of motion pictures. But it required information about audience preferences that could only be learned as the feature played. The best a distributor seeking to extract additional revenue from a film could do given this imperfect information would be to put it in a high-priced theater first and allow movie goers with high preference to “self-select” for that theater. Exclusive, first-run engagements followed by broader release to neighborhood theaters enabled distributors to create a system of priority among film goers that would extract more film rental if the film played well. The limited number of prints of a first-run release had to be rationed sequentially in any event.

### *Runs and clearances*

The District Court and the Supreme Court argued that the clearances of a competitive system would not be stable and might even change from one motion picture to the next. But this is not necessarily so. The exhibitors willing to pay the highest royalty on a motion picture may change little from picture to picture since this depends on the theater’s location, quality, and size which do not change from picture to picture. The courts also neglected the fact that runs are variable even within a fixed run and clearance system. Pictures that play poorly do not run the expected length of time and those that are more successful run longer than expected. The actual run and clearance period depends on how the film performs.

The stability of runs and clearances were reasonable restraints on trade because they enabled distributors and exhibitors to build trust and reputation. Longer term relations with first-run exhibitors supported demand-revealing experimentation and promoted stable expectations of product and play time on both sides. The exhibitors whose runs followed the riskier first run had more information on which to base their booking decision and could select films more carefully. Their rewards were potentially smaller, but so was their risk and the potential gross of their house. The modern analogue to this practice is following a theatrical run with videos and TV. The latter are more easily priced because of the information revealed in the former.

### *Formula deals, master agreements, and franchises*

What we now know as theater chains were at the time called circuits. Some of the most interesting deals the courts found wanting were designed to address issues involving the theater circuits that were beginning to become a major force on the exhibitor side of the industry. They can be seen as precursors to dealing with multiplex theaters and chains.<sup>15</sup>

<sup>15</sup> De Vany and Eckert (1991) argue that these practices were deemed objectionable because they were seen as favoring the circuits over the independent exhibitors. Today, even independent exhibitors are best seen as small chains.

Formula deals determined the rental price of a given feature for the circuit as a specified percentage of the feature's national box office gross receipts. All the circuit's theaters were included in the deal, and the circuit could allocate playing time and film rentals among them at its discretion. This flexibility is surely useful in today's multi-screened theater environment.

In a formula deal, a circuit paid royalties on the nationwide gross box office receipts rather than its own. Because the royalty was independent of the pattern of play chosen by the circuit, it could move a film among its theaters so as to maximize its earnings. The circuit could not decrease its rental payments by manipulating a film's play pattern among its theaters since its film rentals were tied to the national box office revenues [see [De Vany and Eckert \(1991\)](#) for a more complete discussion of this exotic deal]. A modern theorist would see this clever scheme as an asymptotically incentive compatible mechanism. Since a circuit's rentals would be a negligible portion of a movie's national gross, there is little to gain from falsely reporting its theater grosses; truthful revelation is then an asymptotically dominant strategy. This remarkable device was invented more than 50 years before [Groves \(1973\)](#) discovered his famous, truth-revealing mechanism and well before the full theoretical development of incentive-compatible mechanisms. Fact is stranger than theory.

Master agreements (or "blanket deals") licensed a feature in more than one theater (usually a circuit). In an age of multiscreen theaters operated in chains, a master agreement is simply a license that covers several screens and theaters. This seems to have an advantage over writing a contract for every screen independently. An  $n$ -screen contract is more efficient than  $n$  contracts for the same screens. One can see that this transactions-saving device presages [Coase's \(1937\)](#) theory of the firm. Again, the constraints of solving real problems in the movie business led to the development of mechanisms which economic theory had not yet imagined. The theoretical imagination pales in comparison to the "wild" uncertainty and rich complexity of the motion picture environment. Fact can inform theory in ways that are hard to imagine (see below my discussion of contracting when "nobody knows anything").

Franchises covered all of a distributor's releases for more than one season. The Department of Justice has dropped nearly all its objections to franchising and this practice has recently begun to grow.

"Moveover" clauses allowed a licensee to exhibit a given picture in a second theater as a continuation of a run in the first theater. This clearly is the precursor to permitting multiplex theaters the flexibility to move movies among screens in accordance with demand.

### 3.5. *The legacy of Paramount*

The intended beneficiary of *Paramount* was the independent exhibitor. The independent exhibitors issued a heavy volume of complaints to the Department of Justice regarding the defendants and their actions. Independent exhibitors were highly organized and resolutions adopted at their national meetings were directly communicated to the De-

partment of Justice. Not only is there evidence that the Department of Justice heard from the independent exhibitors, but there is evidence that the Department listened to them. Most Department-instigated changes favored the independent exhibitor over the theater chains and the studios.

*Paramount's* remedies did not have their intended effect, however, and they eventually harmed the independent exhibitor. They promoted wider competition for film exhibition rights and reduced production, which raised motion picture rental rates and admission prices. The higher film rental rates lowered exhibitor net profits. Distributors stiffened the terms of exhibition licenses as the less flexible licenses and licensing methods called for by *Paramount* increased their vulnerability to the decline of exhibition caused by the increased competition from television.

Perhaps the definitive test of whether *Paramount* improved the lot of the independent exhibitor is to look at their own actions. Just a few years after the studios sold their theaters, exhibitors, fearing television, called on the Department of Justice to let the studios reenter exhibition so that they might have stronger incentives to produce motion pictures. Exhibitors contended that competitive bidding did not solve any problems and increased film rentals [Federal Trade Commission (1965)]. The methods for allocating motion pictures that were devised in the *Paramount* decrees fostered an explosion of exhibitor antitrust cases against the distributors [Cassady and Cassady (1974)].

The integrated studios – Loews, RKO, Fox, Warner Brothers – lost market value in the range of from 4 to 12 percent when the Supreme Court handed down its 1948 decision [De Vany and McMillan (2004)]. But, Columbia and Universal, non-theater owning defendants, also lost market value on the order of 7 to 9 percent. This suggests the decision was more far-reaching than merely banning producer/distributors from owning theaters; it also barred franchising and other film licensing practices. These would have been more valuable to the non-integrated studio/distributors than to the integrated ones for they were the means through which they maintained vertical controls over theaters.

The evidence does not suggest that the District Court's self-acknowledged "experiment" in altering the long-standing and highly developed structure and practices in this new information industry was "ruinous". A fall in value of 5 to 12 percent is damaging, but not ruinous. But, it did not help and, in the opinion of the stock market, the Court appears to have done harm to the motion picture industry and the value of its assets. The pattern of harm is broad: the integrated and non-integrated studios were harmed and so was their major competitor who was not a defendant. Nor did the litigation seem to end the long-standing hostilities between the studios and independent exhibitors that had been the source of the litigation. The number of lawsuits brought by exhibitors over distribution practices grew rapidly after *Paramount*. Other evidence shows that neither the independent exhibitors nor the film-going public benefited from the fall in feature film production and rise in theater prices and film rental rates that followed *Paramount*.

Current antitrust decisions and policy with respect to the licensing of films and ownership of theaters by distributors have implications for what sorts of films are produced and how they are exhibited. The courts have required films to be individually licensed, theater by theater, and solely on the merits of the film and theater. This requirement has

stood in the way of ownership, franchising, or other forms of long-term contracting between exhibitors and distributors. It also has been interpreted to restrict multiple-picture licensing.

In practice, what this has meant is that it is not possible for a theater to agree with a distributor to exhibit more than one film at a time. No contracts can be made for the whole season of a distributor's releases, nor for any portion of them. This makes it impossible for a theater to be the outlet for a single distributor such as Twentieth Century-Fox or New Line Cinema or Buena Vista. Nor is it possible to license a series of films to theaters as a means of financing their production. Moreover, the inability to contract for portfolios of motion pictures restricts the means by which distributors, producers and theaters manage risk and uncertainty.

At one time, when the studios owned their own theaters and when they could contract for many productions with theaters for a period of years, the theaters were among the primary means of financing production. This was labeled "block booking" and halted by the Paramount Decrees. The licensing and ownership restrictions of the Paramount antitrust litigation have made it impossible for exhibitors to finance the production of motion pictures for release in their theaters. As a result, motion picture financing falls almost entirely on the distributors who rely on internally generated funds, the capital market, or presales of distribution rights in other countries. This, in turn, has probably been a major factor in the emergence of the concept of a "bankable" star whose participation in a project can assure its financing because the star will get it on theater screens where it has an opportunity to earn revenue.

A final irony of the *Paramount* legacy is that only the original signatories to the Decrees are bound by its restrictions. This leaves Sony and other new entrants free to own theaters while Warner Brothers, Paramount, and Twentieth Century Fox cannot.

#### **4. Graduate movie economics: The stable Paretian hypothesis**

Think about it; motion picture revenues are just information, a pure number that has no upper bound. Supply can be extended indefinitely and at constant cost. Consumption does not use up the movie, there is still as much for others to see. And, consumption may lead to an exchange of information with others, expanding the class of users. So there is no limit to what a film might earn, nor is there a natural scale to which revenues converge. Information variables can be magnified by non-linear feedback and, as a result, they can grow explosively or shrink disastrously. Thus, differences among movies can expand exponentially during the course of their runs and motion picture revenues can "go anywhere".

The ability of any model to capture these features seems unlikely, but a deeper look at the nature of the dynamics offers more hope. What we would like to find is the statistical attractor to which motion picture revenues are drawn. There is a class of distributions to which motion picture revenues are drawn; it is the stable class. Stable distributions are the attractors of dynamical processes that retain a self-similar structure

under choice, aggregation, and mixture. The generalized central limit theorem says that sums of independently distributed random variables converge in distribution to the class of stable distributions. The normal or Gaussian distribution is stable, but it is the only stable distribution with finite variance. Other distributions in the stable class are the Pareto, the Cauchy, and the Lévy distributions some un-named ones (because they do not have closed forms). See [Appendix A](#) for a more detailed discussion of the stable distribution.

A stable distribution  $S(\alpha, \beta, \gamma, \delta)$  is a 4-parameter distribution. Distributions in the stable class are characterized by their tail weights. If the tail weight,  $\alpha$ , is 2, the distribution is Gaussian and the tail decays exponentially. If  $\alpha$  is between 0 and 2, then the distribution is some other non-Gaussian member of the stable class and the tail is said to be “heavy” because it decays at less than exponential speed. Non-Gaussian stable distributions have Pareto or power law tails and their tail weight parameter lies in the interval  $(0 < \alpha < 2)$ . When  $\alpha < 2$ , the stable distributions do not have finite variances. And they may not, for  $\alpha < 1$ , have finite means.

Stable distributions, particularly their Paretian tails, give us the laws of the box office. Because they are the attractors of dynamics that are stable in form under choice (maximization), mixture (grouping) and aggregation (adding), they are the natural candidates for our laws of the box office. Motion picture revenues are the result of choices of individual movie fans, grouping of fans of diverse interests in theaters, and an adding up of revenues in individual theaters and over many theaters during the run. Motion picture revenue dynamics, therefore, should converge to stable distributions, but they needn’t be Gaussian (and are not).

With its characteristic high peak, “heavy” Paretian tails and skew, the stable distribution is a statistical monstrosity suffering from a condition called leptokurtosis. But, leptokurtosis is not something that requires mouth wash or deodorant; it is a thing of beauty. It is what makes the movies a business of the extraordinary; a kurtocracy in [Nassim Taleb’s \(2001\)](#) world of randomness. The stable distribution captures the winner-take-all nature of the movie business as well as the dominant influence of extreme events. As it is the limiting distribution of recursive, stable dynamical processes, it is relatively insensitive to initial conditions. But, the naturalness of these properties is insufficient reason to establish that the stable distributions are good descriptions of motion picture revenues. They must fit the data.

And they fit like a glove. The stable distributions and the dynamics that produce them give us the laws of the box office. There is a large and growing literature on this. To conserve space, I will just note these works here together; they are [Sornette and Zajdenweber \(1999\)](#), [De Vany and Walls \(1996, 1999, 2002\)](#), [Lee \(1999\)](#), [Walls \(1997\)](#), [Hand \(2001\)](#), [Ghosh \(2000\)](#), [McKensie \(2003\)](#), [De Vany \(2003a\)](#).<sup>16</sup>

<sup>16</sup> I must mention Art Murphy, a long-time reporter on the business with *Variety* and *The Hollywood Reporter*, who drew log-log plots of movie revenues and frequencies years ago and coined Murphy’s Law; 20% of movies earn 80% of the grosses. He discovered the Pareto law, though he did not know it as such, and his law has been pretty durable (a signature of self-similarity).

#### 4.1. Estimating stable Paretian parameters

A brief word about estimating the parameters of stable distributions. Lee (1999) in his doctoral dissertation covers five techniques for estimating the tail weight parameter  $\alpha$ : the Hill estimator, Nolan's Stable program, McCulloch's quantile estimator, and two techniques adapted from Mandelbrot's early research on cotton future's prices [Mandelbrot (1963a, 1963b)].

Mandelbrot plotted data to find linear portions in the upper tails of probability versus magnitude graphs, just as Art Murphy did in the *Hollywood Reporter*. If such portions existed, then log linear regressions of log survival probability versus log revenue could be used to estimate  $\alpha$ . The slope of this linear regression  $\text{Log}(\text{Prob}) = \text{Log}(\text{constant}) - \alpha \text{Log}(\text{Revenue})$  is the estimate.

An alternative method of estimating  $\alpha$  is to rank movies according to their grosses, from high to low. Assign rank 1 to the first, 2 to the second and on down the line. The ordered revenues are the so-called order statistics. Take the logarithms of these variables and plot them. The left-hand portion of the graph, corresponding to the top grossing movies, will be linear and well-fitted by  $\text{Log}(\text{Revenue}) = \text{Log}(\text{constant}) - \beta \text{Log}(\text{Rank})$ . The inverse of  $\beta$  is an estimate of the tail weight parameter  $\alpha$ .<sup>17</sup>

Nolan's Stable program (available at his web site) estimates the parameters of the stable distribution using maximum likelihood methods. The four parameters of the stable distribution  $\{\alpha, \beta, \gamma, \delta\}$  are interpreted as follows: The characteristic exponent  $\alpha$  is a measure of the probability weight in the upper and lower tails of the distribution; it has a range of  $0 < \alpha \leq 2$  and the variance of the stable distribution is infinite when  $\alpha < 2$ . The skewness coefficient  $\beta$  is a measure of the asymmetry of the distribution; it has a range of  $-1 \leq \beta \leq 1$ , where the sign indicates the direction of skewness. The scale parameter  $\gamma$  must be positive. It expands or contracts the distribution in a non-linear way around the location parameter  $\delta$  which indicates the center of the distribution; the scale parameter is, roughly, the "width" of the central part of the distribution.

We can illustrate these features of stable distributions in the following figures. Figure 1 from De Vany and Walls (2004) shows the stable distribution of motion picture profit as estimated by Nolan's Stable program; it also shows a normal distribution as well as the empirical data. Note how closely the stable distribution fits the data and how poorly the normal fits. The difference is even greater in the tails, but the figure cannot reveal this. A look at the upper tail of box office revenue is shown in a log-log plot of box office gross versus rank, where the top grossing film is rank 1, the next rank 2, and so on. Figure 2 from De Vany and Walls (1999) illustrates the tail linearity for star and non-star movies. The slope of the regression of the logarithm of revenue on the logarithm of rank, as stated above, is an estimate of  $1/\alpha$ . Finally, if the distributions can be shown to be a member of the stable class, De Vany and Walls (2003a) show that estimates of  $\alpha$  can be derived from the mapping of future on past revenues (see below).

<sup>17</sup> Essentially, the left-hand tail in the order statistics plot is the right hand tail in the frequency plot. Because the graphs are flipped over, the slopes are the inverses of one another.

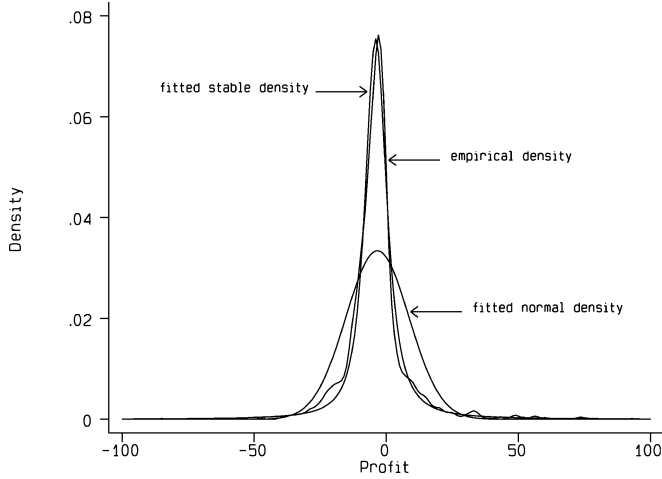


Figure 1. The stable, normal, and empirical distributions of profit.

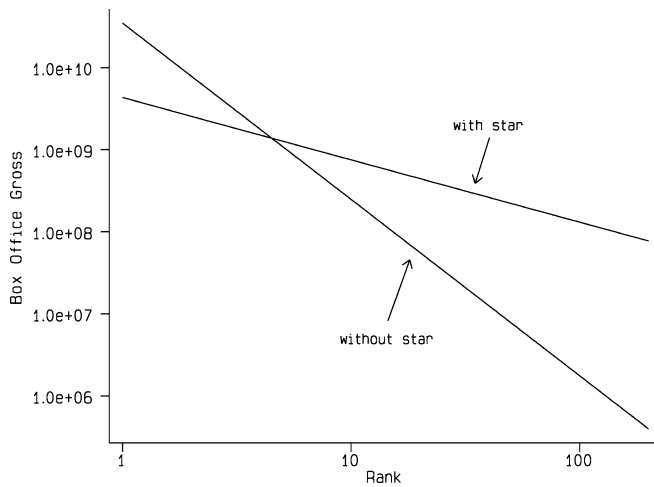


Figure 2. The linear upper region of the log rank versus log revenue distribution for star and non-star movies.

A further issue is the infinite variance. What does this do to regression analysis of the sort often done of the “determinants” of box office grosses or returns? The usual regression model assumes a normally distributed error term with finite variance. But if the variance is infinite, then the conditional expectation of a linear regression model has no precision – it may be useless or misleading. In their research on stars, [De Vany and Walls \(1999\)](#) use quantile estimators and probability point estimators instead of the usual linear regression. In recent work, [Walls \(2005\)](#) uses [McCulloch’s \(1998\)](#) estimator



of the stable regression model. He finds that the distribution of movie returns conditional on its attributes has infinite variance. In a related paper, he shows that the skew-t distribution is a good approximation to the more demanding stable distribution. Thus, without tackling the difficulty of stable regression, researchers can get close approximations to stable results using the skew-t distribution for the error term.

With this background, we can now turn to the laws of revenue and cost that are implied by their dynamics and their stable attractors.

#### 4.2. *Laws of the box office*

There are quite a few findings that have been replicated often enough to assume a tentative status of a law, at least of a statistical form. They relate to either the dynamics or the distribution of box office revenue. I shall be very brief in describing them since there are detailed descriptions in the literature cited above.

##### *“Winner takes all”*

Motion picture box office revenues are highly concentrated and unevenly distributed. The top 4 movies account for 20% of revenues and the top eight for nearly 30%. The Gini coefficient is 0.777, indicating a degree of inequality approaching the distribution of income in Third World countries, and just 20% of movies earn 80% of revenues.

One of the best measures of inequality is the kurtosis. This is a measure of departure from the shape of the Gaussian distribution. A leptokurtotic distribution is sharply peaked, asymmetric and skewed, just the kind of shape that tells us that movies are very different from one another. In the movies, kurtosis is high. A Gaussian distribution has a kurtosis of 3; in the movies we find kurtosis values ranging from 15 to over 100. The kurtosis of box office revenues is 45.

Another measure of inequality is the tail weight coefficient  $\alpha$ . Small values of  $\alpha$  are associated with large inequality because such values make extremely large grosses more likely. Many estimates of  $\alpha$  place it in a range of 1.3 to 1.7, with the most frequent estimate being 1.5. The movies are more unequal than a Gaussian world would ever be. This vast inequality makes Hollywood a kurtocracy. In new work [De Vany, Taleb and Spitznagle \(2005\)](#) have found that lifetime movie grosses of actors is so leptokurtotic and “wild” that the mean does not even exist ( $\alpha \simeq 0.84 < 1$ ). The kurtosis is on the order of 104, making it clear that the movies is a kurtocracy ruled by extraordinary artists and events. Yet, even these kurtocrats experience the “wild” uncertainty and cannot form even an expectation of the box office grosses of their future movies.

Finally, just in case you are an unrepentant Gaussian, you should know that *Forrest Gump* had a domestic gross revenue that was 10 standard deviations above the mean. *Titanic* was 20 standard deviations above the mean. These events would never occur in a billion reruns of history in a Gaussian world. The improbable happens in the movies and these are the main events.

*Hits have long runs*

It is impossible to find a movie that grossed \$50 million or \$100 million in real terms (relative to 1984 base) that did not have a long run. Over the past two decades only 5% of movies exceeded \$50 million and fewer than 2% exceeded \$100 million. A long run does not guarantee big revenues, but hits always have long runs. That is just what the dynamics of non-linear information flow would imply and leads to our next law.

*Non-linear information flow*

When you consume a movie you leave no less for others to see. Each viewer adds to the total without taking anything away because movies are pure information. New viewers are attracted to movies that already have many viewers; box office reports on the news and word of mouth news convey messages to prospective viewers proportional to the number who have seen the film. This non-linear growth rapidly propels film revenues apart from one another. When large revenues produce large revenue growth, the runs of the largest grossing films are extended and the result is long-running hits.<sup>18</sup>

*Opening gross does not predict total gross*

Suppose a movie dies after its first week. Its opening week gross would predict its total gross perfectly because they would be one and the same. On the other hand, opening week gross will not forecast total gross if a movie runs, say, 36 weeks. There is a high but slowly decaying auto-correlation between weekly revenues, but the correlation between opening week revenue and revenues in subsequent weeks decays rapidly. By the seventh or eighth week of the run, there is no correlation between a movie's opening revenue and weekly revenues or between its total and opening revenue. Because non-hit movies have short runs (4 to 6 weeks), their opening revenue is from 60 to 70% of their total revenue. So, for these movies, there is some correlation between opening and total revenue and it is stronger for movies whose revenues decay rapidly. Hit movies move on a different dynamic as we next will show.

*Momentum and chaos*

De Vany and Walls (2003b) offer an explanation for the sudden loss of correlation of revenues for hit movies around the fourth to fifth week of the run and this leads to another law of the box office, the law of momentum and chaos.

<sup>18</sup> Readers may think this information flow is a Gibrat process, but that is not the case. Growth rates in a pure Gibrat process are proportional to size. A Gibrat growth process predicts a linear increase of the variance with time; this is not found and the variance increases non-linearly over time. Second, the standard deviation of the growth rates does not decrease according to  $S^{1/2}$ . A Gibrat process converges on a log-Normal distribution and this is rejected in the dynamics (they do not settle down and show longer range correlations).

Momentum is a non-linear expansion or contraction of revenues, conditional on past revenues; the winners separate themselves from the losers. De Vany and Walls (2003b) show that the expectation of revenues going forward during a movie's run can be mapped on past revenues. Under the stable Paretian hypothesis, the conditional expectation forward revenue on past revenue is  $E[R_i^+ | R_i^+ \geq R_i^-] = R_i^- \alpha / (\alpha - 1)$ , where + means future and - means past. The expectation of future on past revenues is proportional to past revenue. The mapping is multifractal, meaning that  $\alpha$  changes value during the run (hence violating the Gibrat property). Though variable, a typical  $\alpha$  is about 1.5 which implies that  $\alpha / (\alpha - 1)$  is about 3.<sup>19</sup> This implies that, for a value  $\alpha = 1.5$ , a film's expected future earnings for its complete run in week  $i$  is about three times what it has earned to date. A movie that has grossed a large amount does not "use up" its revenue; it can continue to expand its gross. Big films grow faster and small differences between films can grow explosively. This is momentum.

The non-linearity of momentum opens the possibility of chaos. De Vany and Walls (2003b) did find evidence of a bifurcation in the mapping of future on past demand at the fourth or fifth week of the run; at this point hit movies rapidly diverge from the bombs and the mapping bifurcates into two branches, a hit and a non-hit branch. Bifurcation is known to be a route to chaos. De Vany and Walls do no formal tests for chaos because these tests require very long time series and most films do not run long enough to discern chaos. But, the evidence is there in the change in the form of the mapping – it becomes expansive for hits and contractive for bombs – and in the evidence of a rapid separation in the distance (measured as revenue) between hits and bombs. When two near points are propelled rapidly away from one another by a non-linear mapping, you have evidence of chaos.

### *Champions, births and deaths*

Films that attain high rank in the revenue tournament have longer runs than films that attain lesser rank. This and the following laws are from De Vany and Walls (1997) who model box office revenues as a dynamic rank tournament. Champions last longer in rank than contenders do in theirs and expected duration in rank is declining in rank. This is a law of champions that seems to hold in sports as well as in the movies.

Seen from another angle, we can say the hazard of death rises during a film's run and the hazard is everywhere greater for a film the lower is its revenue rank. The smaller hazard of successful films in successive weeks of their runs implies that they should have longer runs. And they do. The expected run life of a film is longer if, at some time during its run, it attains rank 1 and the higher is the highest rank it ever achieves during its life.

Another survival statistic that speaks to the unpredictability of films is that the distributions of deaths and new releases are Poisson. Births and deaths are pure counting

<sup>19</sup> The value of  $\alpha$  can be derived from the regression of the expectation mapping and, thus, is yet another method for estimating its value [see De Vany (2003b)].

processes and their weekly rates are equal. This implies that the intervals of time between deaths is exponentially distributed. The exponential distribution is said to be forgetful because the time to the next death is independent of the length of the interval.<sup>20</sup> So, a film in the  $n$ th week of its run has the same probability of death as one that just opened (if they are at equal rank). De Vany and Lee (2003) extend this model by estimating the Markovian transition probabilities between ranks. A descent in rank is more probable than a rise in rank, and small leaps are slightly more probable than large ones. Nonetheless, a film can transition from low to high rank – a humble beginning does not prevent a film from becoming a hit.

### *Extreme events dominate*

The sample average of motion picture revenues is strongly influenced by extreme events. Dropping just one film, *Titanic*, from a sample of 6289 movies covering twenty years of films released from 1982 through 2001 causes the mean box office revenue to fall by more than a million dollars and the maximum to fall from \$601 million to \$431 million. This single film accounts for just under 1% of the cumulative box office revenue of all the movies released in North America over the past twenty years. In the year it was released, 1997, *Titanic* earned 9% of all motion picture box office revenue that year.

The dominance of extreme events is revealed in many ways: by the rightward skew with the mean far above the mode and reaching the 75th or higher percentile. Figure 2 reveals the dominance in another way. Revenue declines rapidly with a decline in rank; a fall from rank 1 to rank 10 produces a change in revenue of nearly 2 orders of magnitude (from a power 10 to 8). The figure also reveals that extreme events are more dominant for non-star than for star movies (the flatter slope for non-star movies implies a higher value of the tail weight  $\alpha$ ). De Vany and Walls (1999, 2003a) show why this is so: stars increase the least revenue a film might earn, but do not increase the most it might earn.<sup>21</sup>

### *Volatility*

What follows from the dominance of extreme events is that aggregate box office revenues are volatile. So are the average and variance of revenues. One or a few big movies can move aggregate revenues by a lot and this pulls the average and variance along. When revenues obey a stable Paretian law, increments scale as  $(\Delta t)^{1/\alpha}$ . Put another way, the phase space is stable-distributed with parameters in the time space as  $S_\alpha(|\Delta t|^{1/\alpha}, 0, 0)$ , where  $\Delta t$  is the length of the time step.<sup>22</sup>

<sup>20</sup> So much for the motion picture release timing game discussed above. If time to death is independent of how long a film has been running, there is no way for a studio to predict when to release its film in order to avoid competition with some other film.

<sup>21</sup> This is confirmed in the stable parameter estimates in Table 1 where the estimate of  $\alpha$  is smaller for non-star than for star movies.

<sup>22</sup> This relation opens another method of estimating  $\alpha$  which has been coded in the DFA (detrended fluctuation analysis) algorithm by Peng.

Thus, if we observe the box office revenues over a period of 4 years, the maximum difference among revenues we will likely see in 4 years will be about  $4^{2/3} = 5.333$  larger than what we would see in 2 years. The largest leap or fall in revenue from one film to the next that occurs in 4 time years should be about 5 times the change that occurs in 2 years. Thus, we expect quite a bit of volatility in motion picture revenues when they are stable Paretian distributed. At times the bottom may appear to fall out and at others it may appear that the industry is on a rapid expansionary path. But, they are all part of the natural variation and do not indicate a fundamental change in market conditions or fundamentals. Things just happen.

### *Self-similarity*

Volatility is linked to the statistical self-similarity of motion picture revenues. This appears in many guises, but the operative point is this: a random variable  $X(t)$ ,  $t \in T$  is self-similar with index  $H$  if for every  $k > 0$

$$\{X(kt), t \in T\} \simeq \{k^H X(t), t \in T\}$$

where  $\simeq$  means equal in distribution. For a stable motion,  $H = 1/\alpha$  and since we know  $\alpha = 1.5$  it follows that  $H = 2/3$ . For Brownian motion,  $H = 1/2$ . Thus, movie revenues do not follow a Brownian motion. What follows is that there are longer-range correlations in motion picture revenues than in a Brownian motion. This reflects, among other things, the dependence of revenues on extreme events; such events linger in the data because of the powerful influence they exert on the statistics.

But, self-similarity may reflect other kinds of scaling beyond time. We shall see this later, but we can for now imagine that we could restrict movies according to some observable measure that, in effect, is a rescaling of the variable. What we should find is that we get back the same probability distribution once we properly rescale it. Thus, theater screens are another way to rescale revenues and we should find that a similar distribution holds for screens as for revenues, allowing for the difference in their magnitudes. And we do find this [De Vany and Lee (2001)]. Alternatively, if we group movies in budget categories, high, medium, low, we should get back a rescaled distribution of the same probability distribution in each of these classes. However we proceed to analyze the data, self-similarity suggests we may find that they are all windows of different scale onto the same underlying process. We shall encounter self-similarity again.

### *Non-finite moments*

Among the most counterintuitive properties of the stable distribution is that the variance may be infinite; even more mind-blowing is that the mean need not exist. All physical processes that can be realized in an observable world have to be finite, at least we feel that ought to be true. How then is it possible for a probability distribution that actually describes a real phenomenon to have an infinite variance or, worse, a non-finite mean?

Mathematically, this is not hard. All that is required is that probability in the outer tails, or upper tail if we are talking about box office revenue because its lower tail is bounded by zero, decays slowly. A probability distribution can be perfectly well-behaved and yet decay so slowly in the tail that the integrals of the variance or mean may not converge. For example, the variance is the expectation of  $x^2$ . In the Paretian upper tail of box office revenue, probability is of the order  $x^{-\alpha}$ . Since  $\alpha = 1.5$ , the term in the integral of the variance is, therefore, of the order  $x^2 x^{-1.5}$ . This goes to infinity as  $x$  goes to infinity and, so, the variance does not exist. If  $\alpha < 1$ , the same limit argument holds for the expected value, or mean.

There is nothing hard about that. What is hard is how to interpret a stochastic process that does not have a variance or may not have an expected value. I see these properties as a warning that our knowledge is limited and that further observation of the process will reveal novel outcomes. The process is not changing, but what it cranks out does change; some changes will appear to be so abrupt that we may feel that we are observing something entirely new. But, only the outcome of the process has revealed heretofore unseen events, the process remains stable.

Take the infinite variance. We know that our stochastic process takes leaps and falls on the order of  $\Delta t^{1/\alpha}$ . This means that, as time expands, the size of changes grows. Thus, the largest event so far observed will, eventually, be exceeded. In fact, the variance will grow as the sample grows. So, the infinite variance is just a warning that we can never see the full realization of the process which our observations can only reveal a glimpse of. The Lévy stable process unveils perpetual novelty, an attractive feature for a creative process like movie-making.

As to the possible non-existence of the expected value, a similar point is to be made. Now, most motion picture data do tend to have a finite sample mean; though there are some exceptions, notably market shares of the major distributors. But, even if the theoretical mean exists (is finite), we expect it to be variable. This follows from theory, as we know that a linear combination of stable-distributed random variables will also be stable-distributed, so the mean must be a random variable and is stable distributed (with suitable rescaling). But, it is the dependence of the mean on extreme values and the unfolding of the process generating new extreme values that makes the mean variable.

Aside from cautions against drawing inferences from finite observations of a process that can “go anywhere”, the infinite variance (and possibly infinite expected value) has a real advantage. We can forecast outside the sample; the largest box office revenue that we have so far seen does not put an upper bound on the possibilities. It would in a stochastic process that makes only finite leaps, for then the probability of an event larger than the largest event in the sample would be vanishingly small. We couldn’t forecast revenue beyond the largest in the sample. Yet, we know that the probability mass in the upper tail of an infinite variance stable-distributed variable does not vanish, even for events more than four or five standard deviations above the mean. If box office revenues were Gaussian, the chances of a movie reaching *Forrest Gump*’s or *Titanic*’s revenues would be zero since the former is 10 and the latter is 20 standard deviations

above the mean! But, box office is not Gaussian and the stable distribution tells us there are movies yet to be made that will exceed the largest grosses we have ever seen.

*“Nobody knows”*

The laws of the box office offer a foundation for Goldman’s nobody knows anything proposition [Goldman (1983)]. Sample averages and variances are unstable; they change with the sample and the variance grows with the size of the sample. The expected variance is infinite; the mean or expected value may not exist. The stochastic process unveils perpetual novelty. Leaps and falls are so large that the business may seem to have entered a new era. Forecasts of expected values have zero precision. And, yet bigger things lie out there. Put these together and you will see that nobody on Earth can forecast how much a movie will gross.

We know this has to be true anyway. If box office grosses could be forecast, then making movies would be easy. It is not and it should not be. The nobody knows principle elevates the creative process to the center of the motion picture universe. Nobody knows because movies are one-off, creative products, each one unique unto itself. And, they each play out their lives on the screens in equally unique ways. There is a hint of chaos in the dynamics which are richly complex and non-linear.

One suspects that Broadway shows, pharmaceuticals, patent royalties, records, books, and many other creative products share this property. It would be instructive to explore these industries using the models and tools that have proven so useful in motion pictures.

#### *4.3. Laws of production and cost*

Hollywood has its laws of production and cost, but I would not dream of writing down production or cost functions and estimating them. I would not know how to measure the output or the inputs. That is because movies are unique. And, they are produced in fairly unique ways. When a movie gets produced it is by a temporary assemblage of individuals, a temporary team. The producer puts the team together to make a movie and then the team members go back into the labor pool to wait for their next project. The movie can be storyboarded, scene by scene, and the requirements and shooting time can be estimated for each of them. But, shooting time does not translate linearly into frames and scenes of a finished picture. Linear feet of film do not add up to make a movie.

#### *Sequential production*

Producing a movie is a drawn-out process that unfolds in fits and starts over time. It is a recursive process with complex linkages among stages and this point of view is born out, not only by tales told by filmmakers such as Steven Bach (1986) or Mark Litwak (1986), but in the behavior of cost, as you will see. I will describe the stable law of production which says that a movie gets no closer to being finished as work progresses. This odd law of production is revealed in the behavior of expected cost in what I call the

angel's nightmare which says that the expected cost of finishing a movie is proportional to what has already been spent.

### *Production problems*

What I first wish to emphasize are the many problems that arise from the way films are produced and how they are dealt with to varying degrees of success. These problems make the transformation of inputs into output uncertain; hence the difficulty of writing down a function that represents that transformation. Broadly, there are hold up problems, sunk costs, and principal-agent problems among the many creative inputs and financial interests. And, they are solved in the usual ways, by crafting clever contracts and contingent payment schemes, by reputation, by careful monitoring, and by ultimatum.

A movie goes through many stages: the development stage where the story is crafted and tentative casting and financing are explored, to pre-production where the creative elements are cast and contracted and financing and distribution arrangements are made, to production where the scenes that make up the movie are shot, to post-production where editing and reshooting are done, to its eventual release. Money is spent at each stage so if the process fails to advance to the next stage, that money is spent to no effect. It is sunk and not recoverable. The hard lesson is that unless a film is finished and playing on theater screens, none of the money expended on it can be recovered. All the money spent on a movie is sunk at the time it is released.

### *End stage*

Naturally, this end of stage problem presents difficulties. A distributor might decide to renege on a distribution deal and leave the producer with all those sunk costs. At that point, unless there is competition for the film, the distributor might pick it up at a fraction of the cost knowing the producer has sunk his costs and has to sell the film to recover them. This end of stage problem was one of the reasons producers and distributors merged to form the studio system that governed production in the movies from about 1915 to around 1950. Now, it is settled by contract under which the distributor acquires an equity interest in the film. The distributor advances at least part of the production cost of the film in exchange for an interest in it along with the right and obligation to distribute it. With this arrangement, producer and distributor have a mutual interest in seeing the film finished and distributed and the end-stage problem is solved. We know that solving the end stage problem is crucial for, if the end stage cannot successfully be completed, then each preceding stage will come undone and the film will unravel from the end to the beginning and the movie won't get made.

### *Hold ups*

Another stage problem is the potential hold-up that any creative element can expose the whole production to. If an actor or director decides to change the terms of their deal or



insist on some change in the story or role or location or what have you, they threaten the whole project and may render sunk and non-recoverable everything that has been spent up to that point. It is very hard to fire an actor or director or cinematographer after scenes have been shot and are in the can. If they have tantrums or refuse to follow direction or conform to any of the many details that have been laid out in the story board or script, the project is in trouble. These days, when a studio only makes from five to a dozen movies, one movie can bankrupt a studio if it fails to be finished or runs far over budget (or is just expensive and lousy). Tantrums, creative conflicts, a script that falls apart as it is shot, bad weather, or a mechanical shark that breaks under the stress of being towed at high speed in the ocean, and a thousand other things can bring a movie's production to a halt.

### *Solutions*

So, these sorts of problems must be dealt with carefully and creatively and they are among the many sources of what appear to be strange business practices in the movie business. First, we know that reputation can make people act more responsibly and to more diligently pursue the common goal of all the agents involved in making a good movie that comes in on budget. Reputation is one solution. Hollywood may have the most carefully adjudicated system of reputation of any industry. You are in Hollywood what your screen credits say you are. They are out there for everyone to see. So are the box office statistics. Hollywood's many guilds carefully parcel out screen credits and they have developed detailed criteria for awarding director, screenwriter or producer credits.

Repeat dealing amplifies the effect of reputation. People who have worked together before tend to do so again, if their collaboration was a good one. One can trace networks of contacts, as in the Six Degrees of Kevin Bacon game played by devotees of the Internet Movie Data Base, and they are densely wired with a few key people at central nodes. Opportunities flow over this network and the structure of this network is the key to understanding patterns of work and productivity. The network structure of work and opportunity is known to be Pareto-distributed; a few central nodes have dense contacts while most nodes have few. This is another example of self-similarity in the movies; the contact network is Pareto-wired just as box office revenue is Pareto-distributed.

### *Contracts and roles*

Another method of reconciling the creative types and the money people is to write contracts or organize roles suitably. Thus, a director can be punished for going over budget; just reduce her fees in proportion to the amount the production is over budget. This is a common contractual remedy. An alternative is to blur the line between art and money by having talent participate in the profit a movie earns. Actors, directors, or screen writers can get a piece of the action, blunting their willingness to act contrary to the project's economic interest. Some directors or actors assume the role of producer, thus merging

the creative and money roles. Similarly, an actor or director who has agreed to make a movie must have assurances that the project will move ahead to production; his or her participation may be important to getting financing if the talent is “bankable”. So, there is a drop dead date at which the producer must be prepared to “pay or play”. If production does not commence by that date, the producer must pay the artist a fee.

Monitoring the artists is second nature to the money people in Hollywood. The story board tells scene by scene how the movie should progress and this progress is constantly monitored by the producer and a representative of the financial people. They watch the dailies, the film shot each day, and compare the dailies with the story board. If the shooting falls behind schedule, this is noticed and a work-around is discussed. If the dailies show too many reshoots another discussion takes place. And so on.

### *Final authority*

At some point, there must be a decision with whom to vest the authority to decide each problem that comes up. This is another place where reputation and authority crop up. No contract can contemplate all the contingencies that will arise or who should decide when there is disagreement. But, someone must decide when there are opposing views. Who should it be? It is here that contracts can become arcane and inscrutable to an outsider. It is in the best interests of all concerned to place the residual authority in the hands of the person best suited to choose the best action. But, this will differ; sometimes the director is the best one to decide, at other times it is the producer. A sloppy deal memo that vests all residual authority in a director may put the production in the hands of a mad genius and break the studio. If it puts creative authority in the hands of a money person, the cast may walk off.

Ultimately, someone must make the hard decisions that may be required to rein in a runaway production. The stage game is turned into an ultimatum game. That is the role of the completion guarantee, a bonding company that puts up the money to finish the movie. When a movie reaches this stage, the game is up and a third party steps in to make the final decisions. They may not be the best creative choices, but at least the movie will be completed, salvaging some of the costs now long sunk. The only hope of recouping these costs is to finish the movie, which the ones in charge have shown themselves to be incapable of doing. So, the completion guarantor steps in and finishes the job.

### *Stable Paretian costs*

Having described the tangible possibilities that a movie might go over budget or that budgets might be uncertain, a natural question is: How predictable are production budgets? There are many stories about movies famously over budget; Cleopatra, Heaven's Gate, and Titanic are examples that readily come to mind. We are asking what the probability distribution of motion picture production budgets is. There is a problem in finding

the answer to this question. We have production budgets for most movies that are produced and released. But, we don't have data on whether they are on- or over-budget. The best we can do is to estimate the distribution of production budgets, as reported by third party sources, to get a picture of the variety and possibilities. The production budget data will include movies that are over and on budget, giving us some idea of the possibilities.

In two papers, [De Vany and Walls \(2002, 2004\)](#) examine the distribution of production budgets. They find that budgets are Pareto-distributed, though they are less skewed than revenues. Mean budget is at the 62nd percentile and its standard deviation is approximately equal to the mean.

[De Vany and Walls \(2004\)](#) estimate the general stable parameter values of the profit distribution. The lower tail of the profit distribution is Paretian with a tail exponent of  $\alpha = -2.64$ . Because the estimated value of the tail coefficient exceeds 2, budgets have a finite variance. The lower tail of the profit distribution is dominated by cost and so this is a reasonable estimate of the tail weight of the cost distribution. In another paper, [De Vany and Walls \(2002\)](#) estimate the tail coefficient and test the Pareto distribution as a model of the cost data. It turns out to be is a very good model of the budget data. Interestingly, they find that the budgets of R-rated movies have heavier tails than do the distributions of the other ratings and, thus, have a higher probability of large budgets. G-rated movies have the least heavy tails, the inference being that they have a lower probability of reaching very large expenditure values.

### *Returns on budgets*

What does a higher budget buy? [De Vany and Walls \(2003a\)](#) estimate returns to production budgets. The elasticity of mean box-office gross revenue with respect to production budget is 0.54. Larger than zero, but it does show that you cannot buy box office revenue by throwing production dollars at a movie. You only get back fifty cents on the dollar. Hence, there are decreasing returns to production budgets with respect to the box office take.

The elasticity of opening box office revenue, the lowest decile, with respect to production budget is a bit higher at 0.636, but it still is not a good return. Ten percent more budget buys just six percent more opening revenue. Higher deciles show slightly lower budget elasticities: 0.56 for the median, 0.52 for the upper quartile, and just 0.44 for the top decile. To explain it in plainer terms, the top decile is the amount ninety percent of movies will fail to earn and only ten percent will equal or exceed. The estimates say you move this upper decile by just forty cents when you spend another dollar on production. Thus, you have increased the revenue by forty cents of an event that has a probability of just ten percent of occurring. You could say the expected value of the best that can happen is just four cents. Not a very good return for a dollar investment.

### *The angel's nightmare*

An “angel” in Hollywood or on Broadway is an investor who agrees to finance a production.<sup>23</sup> Having found budgets to be Pareto-distributed, we are in a position to see what sort of a nightmare is in store for an angel when a production goes over budget. The angel’s nightmare is tied to the conditional expectation of the Pareto-distributed budget. The Pareto distribution has a counter-intuitive conditional expectation: the expected budget, given that  $y_0$  has been spent, is  $y_0 \frac{\alpha}{(\alpha-1)}$ . Substituting  $\alpha = 2.64$  the expected budget, conditional on  $y_0$  having been spent, is  $y_0 \times 1.609$ . Suppose \$20 million has been spent on a movie budgeted to cost \$16 million. What is the expected cost now that it is already \$4 million over budget? It is  $20 \times 1.609$  or \$32 million. The movie is over the budget and there is no end in sight. This is the angel’s nightmare. It arises from the peculiar property of the conditional expectation of a Pareto-distributed random variable; the conditional expectation is linear in the value of the conditioning event.

### *A strange law of production*

We can infer something of the production process from the behavior of budgets. Time is money and we can invert the budget to obtain production time. This gives us a rather strange law of production. It is that the expected time to completion of a movie in production for  $t$  time periods is  $t \times 1.609$ . The time it will take to finish the project is proportional to how long the movie has been in production. This means that there really is no end in sight, which is why production schedules and completion guarantees are so important in the movie business.

### *Artist productivity*

The first thing one should know about productivity is that artists can differ in the extreme in productivity. The movies is not like washing dishes where two people cannot differ by much. Actors and directors, the two creative productive factors we will focus on, can differ enormously in their productivity, as we shall see. This is natural for several reasons. First, non-linearity can magnify small differences in talent or luck into extreme differences in box office revenues. Second, work opportunities flow over a non-linear network and densely connected artists – made so because they act or direct hit movies – get most of the opportunities and work. Third, the most successful artists tend to have long careers. These facts are implied by the high kurtosis of the stable Paretian distribution that describes the laws of production.

In spite of Rosen’s (1981) explanation of superstardom, there is not much empirical evidence about it. According to his model, small differences in talent translate into large differences in income. Hamlen (1994) presents evidence from popular music regarding

<sup>23</sup> This is drawn from De Vany and Walls (2004).

the distribution of hits. Chung and Cox (1994) show that the Yule distribution is a good model of hit movies of the top stars. The Yule distribution is a power law distribution with  $\alpha = 1$ , so their model is a variant of the stable Paretian model. Interestingly, in their model differences in success are due wholly to chance and talent need not be invoked to explain the variation in artist productivity. Ravid (1999) and De Vany and Walls (1999) examine the influence of stars on profits or revenues and find they are weak.

To shed more light on the productivity of artists, De Vany (2003a, 2003b) examined the productivity of directors and actors. Unsurprisingly, there is a high degree of self-similarity of the laws of productivity and the laws of the box office already presented. The distributions of the number of movies made by actors or by directors are leptokurtotic; kurtosis is 16 and 22, respectively.

Most actors or directors make just one movie: the mode of both distributions is one, the mean of both distributions is 2 and is located at the 75th percentile of each distribution. If we take box office revenue as a measure of productivity, things become a bit more wild. The mean cumulative gross of actors is \$32.6 million and the kurtosis is a breath-taking 104. The median is just \$1.01 million and the mean is up at the 85th percentile. In keeping with the dominance of extreme events, De Vany (2003a, 2003b) finds that the mean portion of cumulative gross due to just one picture is 22% for actors and 32% for directors.

The stable Paretian distribution fits the production data well and estimates of  $\alpha$  values are 1.5 for directors and 1.8 for actors. The form, fit, and values of the tail coefficient follow the stable Paretian property of self-similarity. We know that  $\alpha = 1.5$  for box office revenues; to find a similar value for actor and director productivity tells us that their productivity is statistically similar to box office revenues. In other words, the distribution of the number of films artists make is just another view of the distribution of box office revenues. In keeping with Chung and Cox (1994), it turns out that talent is hard to separate from luck when it comes to how many movies directors make. De Vany (2003a, 2003b) shows that it is only among the most productive directors that one finds a significant role for talent over luck in the odds ratios.

#### 4.4. *Laws of profit*

“Upfront and sunk” is the rule for costs. The rule for profits might be “way after and maybe”. You have to wait a long time for profits to accrue and it might never happen. Movies do lose money, most of them. Revenue streams in over a fairly long time period as a film makes its way through its many venues, domestic theaters, foreign markets, videos, television and so on.

And then there are portfolios to consider. Studios produce portfolios of films, most of which lose money and they have to pool profit over all the films in the portfolio. This is the so-called studio accounting problem. But, it isn’t trickery to allocate all the cost of the portfolio against all the revenue and then find that a few movies took in most of the revenue. The producer of a hit may cry “foul” but this is the way the studio has to be to manage the huge risks it is up against. All this makes accounting separately for

film profits problematic. And, on top of this, is the fact that these figures are not made public. So, one approaches the subject of film profitability with a lot of trepidation and caution.

But, economists want to know and there are ways to at least approximate film profitability in ways that get at the object of our interest; the probability distribution. If it is stable Paretian, we are in luck (it is), because bias will not change the form of the distribution, only its scale. So, our probabilistic inferences will not be that far off. Another way to get at this is to look at returns which can be approximated by the ratio of revenue to cost.

### *Dominance of extreme events*

Our first clue that profit is stable Paretian distributed is in the dominance we find of extreme events. Seventy-eight percent of movies lose money and 35% of profitable movies earn 80% of total profit. Losses are more evenly distributed, after all they are not subject to the non-linear information flow like revenues, and are generally finite; you can only lose what you spend. Fifty percent of the unprofitable movies accounted for 80% of total losses. Just 6.3% of movies earned 80% of profits earned in Hollywood over the past decade. Profits are skewed (but in surprising ways we shall see later) and the kurtosis is large, 15 for all movies in the sample, 23 for movies that do not feature a star, and a mere 5.6 for movies with stars. The mean profit is negative in each of these categories. Finally, the impact of extreme events is clearly evident in the stunning changes that occur over time in the running cumulative average profit shown in [Figure 3](#) taken from [De Vany and Walls \(2004\)](#). The rugged, fractal appearance of the graph is

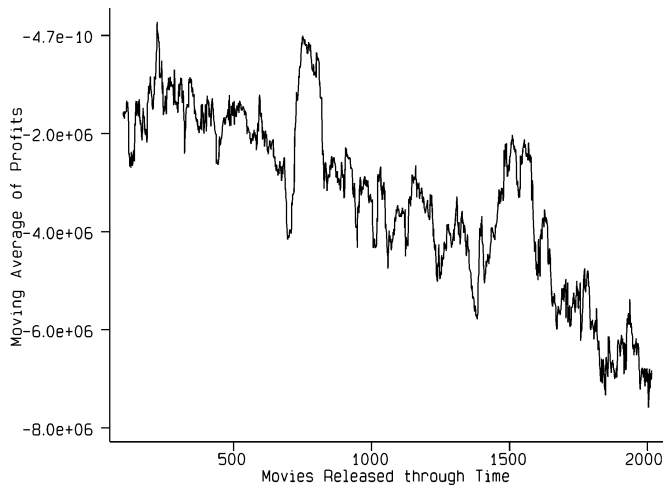


Figure 3. Average cumulative profit of films by date of release.

testimony to the impact of extreme events on the running average and of the stable law of motion of profit.

### Stable profit laws

Having mentioned the stable law of motion, we should now consider the evidence that profit is a stable distributed random variable. There is no doubt profit is random, as we have seen, so how can we say it is stable-distributed? Because the stable distribution fits motion picture profit like a glove and other distributions, like the Gaussian, can be rejected with high confidence. De Vany and Walls (2004) use John Nolan's Stable program to estimate the four parameters of the general stable distribution  $S_\alpha(\alpha, \delta, \gamma, \beta)$ ; I reproduce their results in Table 1.

Note that the Normal distribution, which requires tail index  $\alpha = 2$  and skewness  $\beta = 0$ , does not fit as well (by the Log-likelihood statistic) and normality can be directly rejected by tests. The distribution is not symmetric either, as the best fitting distribution is the general stable distribution with positive skew for movies featuring stars and negative for others. Only positive skew can be accepted with high confidence for star movies; the negative skew of non-star movies and of all movies are not significant. Notice that the locations of the distributions, given by  $\delta$ , roughly corresponding to the mean, differ; the stable distributions place more probability mass on larger losses than the Gaussian.

The scale parameter,  $\gamma$ , represents the width of the central part of the distribution. In all cases, this region is narrower than a Gaussian distribution. This reflects the greater weight in the tails of the stable distribution relative to the Gaussian; the lesson is that the location statistic is a brittle estimate of the mean because probability mass spills away

Table 1  
Maximum likelihood parameter estimates

	Index $\alpha$	Location $\delta$	Scale $\gamma$	Skewness $\beta$	Log-likelihood
<i>All movies</i>					
Normal	2	-3.351	8.442	0	-7855.37
Symmetric $\alpha$ -stable	1.268	-4.079	4.032	0	-7279.87
$\alpha$ -stable	1.259	-4.042	4.020	0.043	-7279.46
<i>Movies with stars</i>					
Normal	2	-2.083	14.186	0	-1439.69
Symmetric $\alpha$ -stable	1.582	-4.568	10.555	0	-1419.16
$\alpha$ -stable	1.624	-6.385	10.805	0.768	-1410.82
<i>Movies without stars</i>					
Normal	2	-3.595	6.789	0	-6216.46
Symmetric $\alpha$ -stable	1.358	-3.932	3.507	0	-5739.47
$\alpha$ -stable	1.335	-3.827	3.441	-0.122	-5737.95

from that area, flowing rapidly into the tails; this is kurtosis – a sharp peak with heavy tails.

The prime parameter is  $\alpha$ , the tail weight coefficient that measures the probability mass that lies in the tails.  $\alpha$  is between 1 and 2, so the expected value of profit exists, but the variance is infinite.  $\alpha$  is well below 2, the tail weight of a Gaussian tail; this is the sense of saying the stable distribution has heavy tails – its probability mass decays more slowly than a Gaussian rate, 1.26 for all movies rather than 2. The star distribution is asymmetric, with a heavier upper than lower tail (this leads to the curse of the superstar, discussed below). The non-star distribution is symmetric.

### *Returns*

Returns are a bit less difficult to measure than profit. One measure of motion picture returns is given by the ratio of box office revenue to production cost. These figures omit certain details, like rentals and print and advertising cost, but they are readily measured and non-controversial. Given reasonable proportionality of these measures, which are partly assured by the way the rental rate and distribution fees are determined, this ratio mimics returns.

Using the self-similarity argument, our expectation is that the ratio of revenue to cost is a Pareto-distributed random variable, at least asymptotically in the tails. The mean return is 0.069 and the standard deviation is a large 6, evidence of a dependence on extreme events. Further evidence is that the mean is well above the 75th percentile. Kurtosis is a monstrous 940, very far from a Gaussian kurtosis of 3.

Taking a lower bound on returns of 2, to move into the tail of the return distribution, De Vany and Walls (2002) estimate the tail exponent  $\alpha$  for G-rated movies at 1.35, PG at 1.56, PG-13 at 1.78 and R at 1.48. A Kolmogorov–Smirnov test accepts the hypothesis that the data follow a Pareto distribution in this upper tail where returns exceed 2. Returns have finite expected values and infinite variance. The authors go on to show that G-rated movies stochastically dominate other ratings in returns and that, with respect to profit, movies in every rating category stochastically dominate R-rated movies.

### *4.5. Stochastic market structure*

We have already seen that motion picture revenues are unequal in the extreme. What about the market shares of firms? De Vany and Walls (1997) show that inequality among motion picture distributors exceeds the inequality of revenues. De Vany and Lee (2003) show that the concept of stochastic market structure is relevant to this industry because shares are extremely volatile. They conclude that the motion picture industry is stochastically concentrated, but enormously competitive.

Market shares are driven by successful movies and follow a stochastic Pareto motion with  $\alpha = 1.3$ . When market structure is stochastic, a stationary measure of concentration, such as the Hershman–Herfindahl Index (HHI) is a misleading index of competitiveness. One must use a stochastic measure. One such measure is the expected value of



the HHI. De Vany and Lee (2003) show that the expected value of the HHI is the second moment (variance) of market shares. Because shares are Pareto-distributed with tail index less than 2, the HHI does not exist (because the variance is infinite). The movies are so competitive that this vaunted (and much over-rated) measure of competitiveness does not exist; the measure has no measure.

#### 4.6. *Pricing talent*

If you really don't know anything, you should wait until you do know. This principle applies to determining the rental rate, as we have seen. The question now is how to price talent when you don't know how much a movie will make. The same principle, the option principle of delaying the decision until you have more information, applies. But, it only applies to the most expensive talent. There we find that prices are determined through contingent clauses in the contract that trigger adjustments to the artist's price if the movie does well. This is called a participation contract in Hollywood, meaning that the artist participates in the movie's profit or revenue. We already know that the studio portfolio problem makes individual motion picture profit somewhat arbitrary (economics says the allocation of joint costs is arbitrary). Revenue is separable from other movies in the portfolio, so participation contracts more often are contingent on revenue rather than profit.

#### *Extreme events and contracting in the tail*

The stable Paretian hypothesis says that extreme events dominate, which implies that compensation should be conditioned on extreme events because that is where the real action is. De Vany and Walls (2004) call this contracting in the tails because the extreme events are located in the upper tail of the revenue distribution, where it is Paretian. Mark Weinstein (1998) studied the evolution of profit-sharing contracts. Darlene Chisholm (1997) employs a sample of contracts to show that contingent compensation is given only to established actors. In her 1995 study [Chisholm (1995)], she analyzes actor incentives in share contracts and looks at degrees of residual claims ("points" in Hollywood). De Vany (2003a, 2003b) makes the same point about directors. Using a sample of contracts of established directors, he shows that there are break points where the percentage of gross distributor revenues paid escalates as revenues rise. He then shows the Gaussian probabilities of these break point events is essentially zero, but significant under the Pareto distribution that actually fits the data. Conclusion: the parties seem to sense they are contracting in a stable Paretian, non-Gaussian world.

Taking this point a bit further, the contracts condition on such extreme events that there can be only a modest incentive effect here. De Vany (2003a) argues that the participation contract is a dynamic decision rule that pays the conditional expectation of future revenue at discrete breakpoints along the realized revenue stream. The contract pays when you do know and its escalating terms reflect the Pareto property of a rising conditional expectation of future revenues as revenues accumulate.

A related important finding is that total director pay is dominated by contingent compensation and that the latter is dominated by one or two movies. De Vany (2003a) shows self-similarity of the revenue and compensation probability distributions for these major directors; they both are Pareto-distributed and have equal tail weights of 1.5. Director careers share the same statistics. Director careers, productivity, pay, and the revenues of their movies all are windows onto the same underlying fundamental process, one that is dominated by extreme events.

### *Stars, profits and the curse of the superstar*

There are difficulties in pricing superstars, particularly if they command a fixed, upfront fee. De Vany and Walls (2004) call it the curse of the superstar. It too, contributes to a preference for paying superstars through participation deals, as shown above.

The curse of the superstar arises because the general stable distribution of profits of superstar movies is skewed and asymmetric; it has a heavier upper than lower tail (recall our discussion of the profit distribution above). This causes the expected profit, \$7.68 million, of superstar movies to be larger than the most likely profit,  $-\$7.50$  million. You can see what would happen if the star's agent extracts the expected profit as the star's fee. The movie almost surely would lose money. Why? A single movie is a small sample (of one). In small samples, the most probable event dominates and the expectation, being a large sample statistic, is improbable. Calculations show that the probability of a loss is 0.80 if the star is paid the expected profit of her movie. This may explain why Ravid (1999) finds that star movies are not more profitable. The curse is another good reason to pay stars contingent rather than fixed fees.

### *Bidding wars and the producer's curse*

There is another reason for producers to be cautious in bidding for stars. There are about 33 big budget movies produced each year. If each of their producers tries to hire an established star, a currently active actor who has three hits in his or her filmography, there will be 33 producers chasing 18 actors. Each producer places a value on each star, known only to the producer. The producer who outbids everyone had the highest value among the 33 producers bidding. That means that 32 producers had lower values, evidence that the winning producer bid too much. This is the producer's curse. It is known as the winner's curse in private value auctions; when you win against many other bidders bidding personal values, which are unknown to you, you have a lot of evidence stacked up against you that your value was too high. Rational bidding requires that you shade your bid downward when you bid against a large number of other bidders.

### *Is an actor really worth \$20 million?*

When you combine the superstar curse and the producer's curse with the evidence that star movies are not more profitable, you begin to see that an actor might not really be

worth \$20 million. But, they will get it anyway if Hollywood fails to guard against these curses. Contingent compensation is one way to avoid these hazards. In the Epilogue to his book, *Hollywood Economics* (2003b), De Vany shows that the sure thing principle adds another layer of complexity to the problem for it leads producers to overvalue the prospects of known entities relative to lesser known ones because the latter have more ambiguous probability distributions.

#### *4.7. Decisions: Greenlighting, bias, and sure things*

There is more to tell, now that we have opened the Pandora Box of decision problems in movie making. I shall be brief because we have set out the model in enough detail to see the issues. Hollywood's decision problems stem from the counterintuitive properties of the stable probabilities and the need to treat movies as prospects, probability distributions, not sure things. Of course, executives share our human failings too, and known biases in human decision making must share the blame.

#### *Ratings and genre*

Ratings and genre predict almost nothing. The odds do differ among films rated differently, but only fairly sophisticated tools can detect this. An R-rating does narrow the audience some, but even among a smaller group of potential viewers, revenue is essentially unlimited. Self-similarity reminds us that all these categories have the same Paretian probability distribution. And, it is stochastic anyway, so there is no way to predict a movie's outcome knowing its rating or genre. Genre is a particularly weak predictor.

#### *Aggregation bias*

Aggregates of variables share the same probability distribution as the variables themselves, the essential property of stable distributed random variables. So, the aggregate returns or revenues of a collection of small budget movies have the same odds or expectations as one large movie whose budget is the sum of the small movie budgets. But, it is much harder to get a studio to finance several small movies than to get one large one for the same money. Why is this so? It is aggregation bias.

Humans are prone to aggregation bias, but the stable probabilities in the movies compound the difficulties. I think this is because it is difficult to see that the odds that the sum of random variables will be equal to some amount are no different from the odds that a random variable will be equal to that sum. The odds of making two movies whose revenues sum to \$100 million are the same as the odds of making one movie that grosses \$100 million. Most people would see the former event as less likely than the latter. This is an aggregation bias, but it is only a bias because movie grosses are stable-distributed, which makes the odds of these two events the same.<sup>24</sup>

<sup>24</sup> See Appendix A.

*Ambiguous probabilities and sure things*

Another source of difficulties occurs when probabilities are ambiguous. They always are in the movies, but here is the issue. An executive must always make some assessment of the odds in choosing movies. This is harder to do when the movie or the actors or producers are unknown entities. A known entity gives more assurance that the odds can be estimated. At least there is some experience to rely on in estimating the probabilities so they seem to be less ambiguous and the known entity is more likely to be chosen.

But, what is the meaning of ambiguity when we have distributions with infinite variance and where the mean depends so heavily on extreme events in the upper tail? Since movie outcomes are so scattered over the whole space of possibilities, there is no typical movie or representative type. If the studio were to treat both prospects as having Pareto distributions, then more ambiguous probabilities of the lesser known entity would be captured by the spread between the least outcome and the best outcome. Pareto distributions reach an upper value of infinity, no matter what parameter values they have. So one Pareto distribution can only be more diffuse than another if it has a lower minimum value. This argument would suggest that ambiguous probability distributions might be evaluated as having worse worst outcomes (lower lower bounds in the support of the distribution). It is surely easier for an executive to conjecture as the worst outcome than to estimate the parameters of the Pareto distributions and they are likely to be unduly pessimistic when unknown artists are involved. This is where good statistical modeling might make a big difference in the quality of decisions.

*Inside and outside odds*

These decision paradoxes suggest that good statistical modeling might improve Hollywood's decision making. I think all I have said in this review of research points to a conclusion that Hollywood should compare the inside against the outside odds when they choose movies. Outside odds are objective odds given by careful statistical modeling of the data. Inside odds are part of the arguments, stories, and forecasts that people tell each other in meetings when they choose the studio's production slate. Overly optimistic forecasts, known artists, big projects, and "sure things" tend to carry the argument in this setting, even when they do not square with the real odds outside the studio.

**5. Conclusion**

How durable are the methods and results that I have described here? Will the rise of DVDs and the international market and the declining share of domestic theatrical revenue in total gross change matters? I think not because this research has looked for deeper patterns that have a universal order. We know that international revenues, and revenues in many foreign countries, follow the same Paretian pattern (even with the same tail weight). Thus, culture and local arrangements in specific countries do not

alter the universal pattern. In each country where a film plays, its revenues follow a similar dynamic where choice, mixture and aggregation are the stable operators. Necessarily, they will result in a stable distribution of box office revenues. The same points hold for DVD sales. The timing of these sales does help in resolving the nobody knows problem for the later markets. If foreign theatrical distribution is later than domestic release, then the foreign film buyers have prior information on which to condition their estimates. The same point holds for DVDs; the theatrical run reveals prior information not available at the initial opening. And, we should be able to do better in predicting revenues.<sup>25</sup>

Nothing I have said changes when we open the analysis to include these markets. All the results and techniques apply to them. But, prior information, garnered from the North American theatrical run, lets us to make better estimates of outcomes in the downstream markets (just as the second and third run theaters had prior information from the first run theaters in the days before the *Paramount* decision). This prior information can be incorporated by conditioning the probability estimates of the stable distributions of foreign and DVD markets on the North American grosses. Lastly, just to show that things never change, piracy is pushing the studios to release films simultaneously in all markets, domestic, foreign, and DVD. Thus, the prior information revealed in the North American run is gradually washing away in the battle against piracy. And we shall then once again be up against the full force of the nobody knows principle.

The research reported here shows clearly that anyone who thinks he knows how much a movie will gross doesn't know what he is doing. He would have to have a model and we have seen there isn't one. If there were a model, making movies would be easy and no one would ever lose money. Movies are prospects, probability distributions as ephemeral as the images on the screen. It has to be this way because movies are creative products – they are pure information and the revenue they might earn is just a number – pure information – with no natural scale or limit.

I hope I have shown there are powerful ideas and tools that we can use to reveal the deeper order of this seemingly chaotic, challenging, and endlessly fascinating business of the movies. I believe these tools can be a constructive basis of a new economics of art and culture.

## Appendix A

What is the role of stability in the movies? Theoretically, if the random revenues of movies aggregated over choices, theaters, weeks, and movies are stable, then the distributions of these sums all will be similar. The concept of stability means that the distribution of sums of random variables is similar to the distribution of the random variables (this is defined more precisely below). This implies that each observed sequence

<sup>25</sup> I am told that the price charged for a foreign release is conditioned on its domestic revenues, meaning the latter are a good predictor of foreign revenue.

of random variables or sums of random variables of a stream of box office revenues has the same distribution and is only a different view of the underlying process. The concept of similarity, or self-similarity, is used extensively in this chapter and leads to deep insights about the movie business.

The other important property for motion pictures of stable or self-similar processes is that they have limiting distributions. This means that the distributional dynamics will converge on a stable distribution and the distribution will be relatively unaffected by the initial conditions. Despite all the uncertainty about where motion picture revenues may go, they will converge on a statistical attractor. Consequently, the statistical attractors and the underlying dynamics have a deep property: they are stable and convergent. Finally, self-similar, stable processes need not have a finite mean or variance, a very important property of the movies.

Taken together, the properties of stable aggregation, non-finite variance, and self-similarity give us a way of coming to grips with the “wild” uncertainty in the movies. They make the probability distributions of outcomes central to the study of this industry. Let me briefly discuss stable distributions and their properties. I highly recommend [G. Samorodnitsky and M.S. Taqqu \(1994\)](#) as an authoritative text on stable processes and stable distributions.

Most readers know that the Central Limit Theorem states that the sum of a large number of iid random variables with finite variance converges to a Normal distribution. What they may not know is that if the finite variance assumption does not hold true of the random variables, then the Generalized Central Limit Theorem states that the limiting distribution must be a member of the stable class. Informally, the term ‘stable’ is used because when iid members of a stable family are added the shape of the distribution does not change. The distribution obtained by adding random variables is equivalent to convolving their probabilities. If the “shape” of the convolved distribution of sums is similar to the distributions of the random variables, the distribution is stable. A random variable is stable if  $\sum_{i=1}^n Y_i \simeq^D Y$  for any  $n$  where the symbol  $\simeq^D$  means equal in distribution. The distribution of stable random variables and their sums is similar. Remarkably, a random variable is stable as soon as the equality is true for  $n = 2$  and  $3$ . Surprisingly, this implies that the sum of revenues of two or three small budget movies have the same probability of hitting a revenue of, say,  $X$  as a large budget movie. An equivalent way to define stability is in linear transformations of the probability distribution. Two random variables are similar in distribution if there exist constants  $a$  and  $b > 0$  such that  $Y \simeq^D a + bX$ . That is,  $Y$  and  $X$  are similar in distribution if the distribution of  $Y$  is a linear transformation of the distribution of  $X$ .

Stability means that the dynamics lead to a stable attractor since the sequence of sums of a stable random variable is stable or self-similar. A stronger statement is this: the distribution function  $F$  possesses a domain of attraction if and only if it is stable. The limiting distribution function  $F_X$  of independent r.v.’s  $X_k$  belongs to the domain of attraction of  $F$  if there exist normalizing constants  $a_n, b_n > 0$  such that the distribution of  $(\sum_{i=1}^n X_i - a_n)/b_n \rightarrow F$ . The Normal (Gaussian) distribution is stable, but it is the only stable distribution with a finite variance. The Pareto, Lévy and Cauchy

distributions are the other (named) stable distributions. But, there are many other members of the stable class, whose functional forms cannot be given. Lévy characterized the class of stable distributions through their characteristic function. A stable distribution  $X \sim \mathbf{S}(\alpha, \beta, \gamma, \delta)$  is a four-parameter distribution with characteristic function given by

$$C(t) = \begin{cases} \exp(-\gamma^\alpha |t|^\alpha [1 + i\beta (\tan \frac{\pi\alpha}{2}) (\text{sign } t) ((\gamma|t|)^{1-\alpha} - 1)]) & \text{if } \alpha \neq 1, \\ \exp(-\gamma |t| [1 + i\beta \frac{2}{\pi} (\text{sign } t) (\ln |t| + \ln \gamma)]) & \text{if } \alpha = 1. \end{cases}$$

The exponent  $\alpha$  is a measure of the probability weight in the upper and lower tails of the distribution; it has a range of  $0 < \alpha \leq 2$  and the variance of the stable distribution is infinite when  $\alpha < 2$ . The basin of attraction is characterized by the tail weight of the distribution ( $\alpha$ ). This remarkable feature tells us that the weight assigned to extreme events is the key distinguishing property of a stable probability distribution. The skewness coefficient  $-1 \leq \beta \leq 1$  is a measure of the asymmetry of the distribution. Stable distributions need not be symmetric; they may be skewed more in their upper tail than in their lower tail. The scale parameter  $\gamma$  must be positive. It expands or contracts the distribution in a non-linear way about the location parameter  $\delta$  which is the center of the distribution.

The tails of a stable distribution are Paretian and moments of order  $\geq 2$  do not exist when  $\alpha < 2$ . This is typical of motion picture revenues. Its mean need not exist for values of  $\alpha < 1$ . This is true of the cumulative grosses of movies made by actors.

The  $\alpha$ -stable distribution becomes the Cauchy distribution when  $\alpha = 1$  and  $\beta = 0$ . The stable distribution is the Lévy distribution when  $\alpha = 0.5$  and  $\beta = \pm 1$ . When  $\alpha = 2$  the skew parameter ceases to have any impact and the stable distribution has only two parameters ( $\delta, \gamma$ ) the mean and variance of the symmetric Gaussian distribution. The general stable distribution contains the Lévy, Pareto, Cauchy and Normal distributions for specific values of the parameters. It need not be symmetrical; thus, the upper and lower tails may differ. This is precisely what we would expect to find for motion picture profit because the lower tail is limited by expenditures while the upper tail is essentially unlimited.

Aside from these theoretical properties of stable-distributed random variables, the sample statistics are unusual and they befuddle standard analysis that is too often based on the Normal distribution. For one thing, the sample variance (not the theoretical variance, which is infinite) increases with the size of the sample. For another thing, the sample average is volatile. Just one movie can move the average in a dramatic way. Indeed, the sample average  $(X_1 + X_2 + \dots + X_n)/n$  has the same distribution as  $X_k n^{1/\alpha-1}$  where  $X_k$  is any term in the sum. If  $\alpha < 1$  then the maximum value contributes heavily to the average. This is true, though to a lesser extent, when  $1 < \alpha < 2$ . This is the Titanic effect wherein a single movie may contribute a large portion of the sum of movie revenues in a year. Since the average is a linear combination of stable random variables, it too is stable-distributed. This means that sample averages are unreliable and that the perils of forecasting future events from past averages are large. Forecasts have an infinite variance as well.

## References

- Bach, S. (1986). *Final Cut: Dreams and Disaster in the Making of Heaven's Gate*. New American Library, New York.
- Cassady, R. (1959). "Monopoly in motion picture production and distribution: 1908–1915". *Southern California Law Review* 32, 325–390.
- Cassady, R. Jr., Cassady, R. III (1974). "The private antitrust suit in American business competition". Working Paper. Bureau of Business and Economic Research, University of California, Los Angeles.
- Caves, R.E. (2000). *Creative Industries: Contracts between Art and Commerce*. Harvard University Press, Cambridge.
- Chisholm, D. (1995). "Actor fixed effects and incentives in motion-pictures share contracts". Technical Report. Lehigh University.
- Chisholm, D.C. (1997). "Profit-sharing versus fixed-payment contracts: Evidence from the motion pictures industry". *Journal of Law, Economics and Organization* 13 (1), 169–201.
- Chung, K.H., Cox, R.A.K. (1994). "A stochastic model of superstardom: An application of the Yule distribution". *Review of Economics and Statistics* 76 (4), 771–775.
- Coase, R.H. (1937). "The nature of the firm". *Economica* 4, 386–405.
- Collins, A., Hand, C., Snell, M. (2003). "What makes a blockbuster? Economic analysis of film success in the United Kingdom." *Journal of Decision Analysis*.
- Conant, M. (1960). *Antitrust in the Motion Picture Industry: Economic and Legal Analysis*. University of California Press, Berkeley.
- De Vany, A. (2003a). "Contracting with artists when nobody knows anything". In: *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry*. Routledge, London, pp. 431–478. Chapter 11.
- De Vany, A. (2003b). *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry*. Routledge, London.
- De Vany, A., Eckert, R. (1991). "Motion picture antitrust: The Paramount cases revisited". *Research in Law and Economics* 14, 51–112.
- De Vany, A., Lee, C. (2001). "Quality signals in information cascades and the dynamics of the distribution of motion picture box office revenues". *Journal of Economic Dynamics and Control* 25, 593–614.
- De Vany, A., Lee, C. (2003). "Stochastic market structure: Concentration measures and motion picture antitrust". In: *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry*. Routledge, London, pp. 351–371. Chapter 9.
- De Vany, A., McMillan, H. (2004). "Was the antitrust action that broke up the studios good for the movies? Evidence from the stock market". *American Law and Economics Review* 6 (4), 135–152.
- De Vany, A., Taleb, N., Spitznagle, M. (2005). "Can we shield actors from "wild" uncertainty". Working Paper. *Ars Analytica and Empirica*, LLC.
- De Vany, A., Walls, W.D. (1996). "Bose–Einstein dynamics and adaptive contracting in the motion picture industry". *The Economic Journal* 439 (106), 1493–1514.
- De Vany, A., Walls, W.D. (1997). "The market for motion pictures: Rank, revenue and survival". *Economic Inquiry* 4 (35), 783–797.
- De Vany, A., Walls, W.D. (1999). "Uncertainty in the movie industry: Does star power reduce the terror of the box office?" *Journal of Cultural Economics* 23 (4), 285–318.
- De Vany, A., Walls, W.D. (2002). "Does Hollywood make too many R-rated movies? Risk, stochastic dominance, and the illusion of expectation". *Journal of Business* 75 (3), 425–451.
- De Vany, A., Walls, W.D. (2003a). "Big budgets, big openings, and legs: Analysis of the blockbuster strategy". In: *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry*. Routledge, London, pp. 223–252. Chapter 6.
- De Vany, A., Walls, W.D. (2003b). "The breakdown of herding". In: *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry*. Routledge, London, pp. 89–116. Chapter 3.
- De Vany, A., Walls, W.D. (2004). "Motion picture profit, the stable Paretian hypothesis, and the curse of the superstar". *Journal of Economic Dynamics and Control* 28 (6), 1035–1057.



- Einav L. (2004). "Not all rivals look alike: Estimating an equilibrium model of the release date timing game". Technical Report. Stanford Department of Economics, Stanford, CA.
- Federal Trade Commission (1965). "Paramount Pictures Inc., et al. Consent Judgments and Decrees Investigation". Report of Federal Trade Commission. Washington, DC.
- Ghosh, A. (2000). "The size distribution of international box office revenues and the stable Paretian hypothesis". Working Paper. Economics department, University of California at Irvine.
- Goldman, W. (1983). *Adventures in the Screen Trade*. Warner Books, New York.
- Groves, T. (1973). "Incentives in teams". *Econometrica* 41, 617–631.
- Hamlen, W.A.J. (1994). "Variety and superstardom in popular music". *Economic Inquiry* 32 (3), 395–406.
- Hand, C. (2001). "Increasing returns to information: Further evidence from the UK film market". *Applied Economics Letters*, 419–421.
- Krider, R., Weinberg, C. (1998). "Competitive dynamics and the introduction of new products: The motion picture timing game". *Journal of Marketing Research* 35 (1), 1–15.
- Lee, C. (1999). "Competitive dynamics and the distribution of box office revenues in motion pictures: The stable-Paretian hypothesis". Doctoral Dissertation. Department of Economics, University of California, Irvine.
- Litman, B. (1990). "The motion picture entertainment industry". In: Adams, W. (Ed.), *The Structure of American Industry*. Macmillan, New York, pp. 183–216.
- Litwak, M. (1986). *Reel Power: The Struggle for Influence and Success in the New Hollywood*. William Morrow, New York.
- Mandelbrot, B. (1963a). "New methods in statistical economics". *Journal of Political Economy* 71, 421–440.
- Mandelbrot, B. (1963b). "The variation of certain speculative prices". *Journal of Business* 36, 394–419.
- McCulloch, J.H. (1998). "Linear regression with stable disturbances". In: Adler, R.J., Feldman, R.E., Taqqu, M.S. (Eds.), *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Birkhäuser, Berlin, pp. 359–378.
- McKensie, J. (2003). "Box office revenue distributions, Paretian tails, and survival analysis". Draft Thesis. University of Sydney.
- Ravid, S.A. (1999). "Information, blockbusters and stars: A study of the film industry". *Journal of Business* 72, 463–486.
- Rosen, S. (1981). "The economics of superstars". *American Economic Review* 71, 167–183.
- Samorodnitsky, G., Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes*. Chapman and Hall, New York.
- Scherer, F.M., Harhoff, D., Kukies, J. (2000). "Uncertainty and the size distribution of rewards from innovation". *Evolutionary Economics* 10, 175–200.
- Seagrave, K. (1997). *American Films Abroad: Hollywood's Domination of the World's Movie Screens from the 1890s to the Present*. McFarland & Co., Jefferson, NC.
- Sedgewick, J. (2001). *Popular Filmgoing in 1930s Britain: A Choice of Pleasures*. Exeter University Press, Exeter.
- Simon, H.A. (1955). "On a class of skew distribution functions". *Biometrika* 52, 425–440.
- Sornette, D., Zaidenweber, D. (1999). "The economic return of research: The Pareto law and its implications". *European Physical Journal B* 8 (4), 653–664.
- Taleb, N. (2001). *Foiled by Randomness*. Texere, New York.
- Walls, W.D. (1997). "Increasing returns to information: Evidence from the Hong Kong movie market". *Applied Economics Letters* 4 (5), 187–190.
- Walls, W.D. (2002). "Review of John Sedgewick, *Popular Filmgoing in 1930s Britain: A Choice of Pleasures*". *Journal of Cultural Economics*, 298–301.
- Walls, W.D. (2005). "Modeling movie success when 'nobody knows anything': Conditional stable-distribution analysis of film returns". *Journal of Cultural Economics* 29 (3), 177–190.
- Weinstein, M. (1998). "Profit-sharing contracts in Hollywood: Evolution and analysis". *Journal of Legal Studies* 27 (1), 67–112.

## ROCKONOMICS: THE ECONOMICS OF POPULAR MUSIC\*

MARIE CONNOLLY

*Princeton University, USA*

ALAN B. KRUEGER

*Princeton University and NBER, USA*

### Contents

Abstract	668
Keywords	668
1. Introduction	669
2. The players	670
2.1. Contracts	673
3. Some theoretical issues regarding concert pricing	675
4. Concert industry trends	676
4.1. Trends in prices	677
4.1.1. More on price indices	679
4.2. Shows, sales and revenues	682
4.3. Distribution of revenues	684
4.4. Explanations: Baumol and Bowen's disease; cartelization; Bowie theory	685
5. Ticket distribution and scalping	688
5.1. Scalping: Theoretical issues	688
5.2. Evidence on scalping	689
5.3. Scalping and price trends	690
6. Rankings	691
7. Superstar effects	695
8. The world of radio broadcasting	697
8.1. Royalties from performing rights	697
8.2. Music publishing in the US	701
8.3. Foreign markets	701
8.4. Payola	702
8.5. Digital recordings in the Internet era	706

\* We thank Art De Vany and Bentley MacLeod for helpful comments and Gary Bongiovanni for data, and absolve them from any mistakes we may have made.

*Handbook of the Economics of Art and Culture, Volume 1*

*Edited by Victor A. Ginsburgh and David Throsby*

*Copyright © 2006 Elsevier B.V. All rights reserved*

*DOI: 10.1016/S1574-0676(06)01020-9*

9. File sharing and other new technologies	707
9.1. Intellectual property issues	708
9.2. Does file sharing lower CD sales? Preliminary considerations	709
9.3. Empirical studies	711
9.4. Searching for a new business model	714
9.5. New business practices	715
10. Conclusion	715
References	716

### **Abstract**

This chapter considers economic issues and trends in the rock and roll industry, broadly defined. The analysis focuses on concert revenues, the main source of performers' income. Issues considered include: price measurement; concert price acceleration in the 1990s; the increased concentration of revenue among performers; reasons for the secondary ticket market; methods for ranking performers; copyright protection; and technological change.

### **Keywords**

rockonomics, concerts, superstars

*JEL classification:* Z1, L82, O34

*The fact of the matter is that popular music is one of the industries of the country. It's all completely tied up with capitalism. It's stupid to separate it.*

Paul Simon

## 1. Introduction

As was highlighted by a much ridiculed box in the 2004 *Economic Report of the President* that questioned whether fast food restaurants should be classified in the manufacturing sector, defining an industry necessarily entails some arbitrariness. We seek to survey the economics of the popular music industry, a subfield of economics that we euphemistically call *Rockonomics*. But what is popular music? Where does one draw the lines? Here, we will define popular music as music that has a wide following, is produced by contemporary artists and composers, and does not require public subsidy to survive. This definition rules out classical music and publicly supported orchestras. It includes rock and roll, pop, rap, bebop, jazz, blues and many other genres. What about Pavarotti? Well, we warned you that the border of the definition can be fuzzy. If the three tenors attract a large following and are financially viable, we would include them in the popular music industry as well.

Why is popular music worthy of a *Handbook* chapter? There are several responses. First, Paul Simon's sentiment in the epigraph notwithstanding, for many fans popular music transcends usual market economics and raises spirits and aspirations. In this vein, for example, Bruce Springsteen once commented, "In some fashion, I help people hold on to their own humanity, if I'm doing my job right." Dewey Finn, the character played by Jack Black in the hit movie, *School of Rock*, went even further, immodestly claiming, "One great rock show can change the world." The rock and roll industry arguably started as a social movement intended to bring about political, economic and cultural change, as much as it did as a business. Certainly, popular music is an important cultural industry.

Second, precisely because emotion and non-traditional economic concerns loom large in popular music, the industry can be a breeding ground for new insights into economics. Social considerations are important in transactions outside the music industry; they are just magnified when it comes to a rock and roll concert.

Third, the popular music industry provides a testing ground for some important economic theories. For example, popular music is a classic superstar industry, where rewards are highly skewed. Can economic models explain the distribution of rewards? Also, despite the non-economic forces that affect the popular music industry, can basic economic factors, such as supply and demand, still provide a good explanation of many of the important developments in the industry?

Fourth, the industry is profoundly affected by technological change, such as the advent of radio, TV, record albums, cassette tapes, CDs, MP3 players, the Internet, etc. Thus, popular music provides an unusual setting to understand how rapid technological change affects an industry.

Fifth, and finally, the popular music industry is, by definition, popular. As a consequence, students are particularly motivated to learn about the industry, and examples drawn from the industry thus provide good material for teaching economics.

To help guide our coverage, [Table 1](#) provides a summary of the main income sources for the top 35 popular music performers who toured in 2002, ranked by income. The figures, which are taken from *Rolling Stone* magazine, should be viewed as rough estimates. Another *caveat* to bear in mind is that some sources of income – such as revenue from merchandise sales, movies, commercials and (don't laugh) cell phone jingles – are not itemized in the table, but included in the total. These other sources of revenues can be substantial. The Osbournes, for example, had a huge success with their reality TV show that aired on MTV. Nevertheless, the table provides an indication of the relative importance of live concerts, record sales, and publication royalties in performers' income. Although the concert figures are somewhat inflated because artists do not tour every year (and our sample conditions on having toured), it is clear that concerts provide a larger source of income for performers than record sales or publishing royalties. Only four of the top 35 income-earners made more money from recordings than from live concerts, and much of the record revenue for these artists probably represented an advance on a new album, not on-going royalties from CD sales. For the top 35 artists as a whole, income from touring exceeded income from record sales by a ratio of 7.5 to 1 in 2002. Royalties from publishing music was slightly less than income from recordings. Consequently, we will devote much attention to live concerts in this chapter.

The remainder of the chapter is organized as follows. The next section describes the organization of the music industry, devoting particular attention to live performances. Section 3 discusses theoretical issues in the pricing of concerts. Section 4 considers major developments in the popular music concert industry, with particular emphasis on prices, ticket sales, revenue, and concentration among promoters. Section 5 considers the important role played by scalpers. Section 6 provides a method for ranking performers based on economic data. Section 7 considers the role of the superstar model in the rock and roll industry. Section 8 discusses the role of radio and royalties, and Section 9 considers related issues involving file sharing. Section 10 concludes by highlighting important questions for further research.

## 2. The players

The market for popular music has many players and complex contracts. [Figure 1](#) provides a schematic diagram of the organization of key elements of the popular music industry. First and foremost, of course, are the musicians, who form a band. The band may write its own music and lyrics, or it may purchase music from an outside composer. In [Figure 1](#), we have illustrated a situation for a band that writes its own music. The bands have managers who represent them and take a share of their earnings in exchange for their managerial services. On behalf of the bands, managers make contracts with promoters to promote live concerts. The promoter secures a venue, advertises the

Table 1  
 Estimated pre-tax gross income by source for 35 top artists who toured in 2002 (millions of US dollars)

Rank	Artist	Live concerts	Recordings	Publishing	Total income
1	Paul McCartney	64.9	2.2	2.2	72.1
2	The Rolling Stones	39.6	0.9	2.2	44.0
3	Dave Matthews Band	27.9	0.0	2.5	31.3
4	Celine Dion	22.4	3.1	0.9	31.1
5	Eminem	5.5	10.4	3.8	28.9
6	Cher	26.2	0.5	0.0	26.7
7	Bruce Springsteen	17.9	2.2	4.5	24.8
8	Jay-Z	0.7	12.7	0.7	22.7
9	Ozzy Osbourne/The Osbournes	3.8	0.2	0.5	22.5
10	Elton John	20.2	0.9	1.3	22.4
11	The Eagles	15.1	0.7	1.4	17.6
12	Jimmy Buffett	13.7	0.2	0.5	17.6
13	Billy Joel	16.0	0.0	1.0	17.0
14	Neil Diamond	16.5	0.0	0.3	16.8
15	Aerosmith	11.6	1.0	0.8	16.5
16	Crosby, Stills, Nash & Young	15.7	0.0	0.3	16.0
17	Creed	10.9	1.1	1.6	13.4
18	Rush	13.4	0.0	0.0	13.4
19	Linkin Park	1.7	4.7	6.3	13.1
20	The Who	12.6	0.0	0.0	12.6
21	Red Hot Chili Peppers	6.1	3.4	2.7	12.1
22	Brian "Baby" Williams	0.2	2.7	0.9	11.8
23	Nsync	7.7	0.5	0.9	9.4
24	Barry Manilow	8.0	1.2	0.0	9.2
25	Britney Spears	5.5	1.8	1.0	9.1
26	Alan Jackson	4.6	3.0	1.4	9.0
27	Rod Stewart	6.6	1.4	0.8	8.8
28	Andrea Bocelli	8.1	0.2	0.4	8.7
29	Brooks and Dunn	6.7	0.4	1.4	8.1
30	Enrique Iglesias	4.4	1.5	1.7	7.6
31	Tom Petty	6.6	0.2	0.7	7.5
32	Tool	7.3	0.0	0.0	7.4
33	Kid Rock	3.4	0.8	1.3	7.0
34	Kenny Chesney	5.8	1.1	0.1	7.0
35	Santana	6.0	0.0	0.7	6.9
	Average	12.7	1.7	1.3	17.4

Notes: Figures are estimates of pre-tax gross income in 2002. The total income may exceed the sum of the first three columns because of TV, movie, merchandise and other potential sources of income.  
 Source: LaFranco (2003).

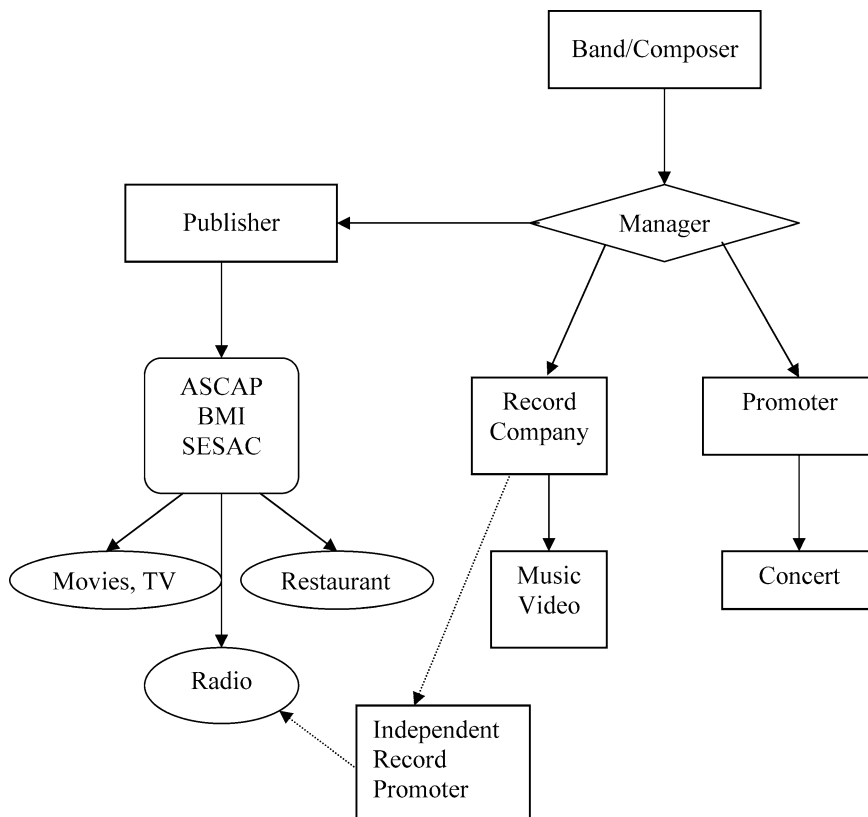


Figure 1. Organization of the popular music industry.

event, and takes care of other arrangements. Successful bands also have contracts with recording companies to produce and market CDs. Record companies are occasionally involved in promoting concert tours, but they typically play only a peripheral role in concerts, when they are involved at all.

If a band composed its own music, it will also contract with a publisher to copyright the music. The publisher will contract with a performing rights organization, which licenses the music for radio stations, television and other users, monitors the use of the music, and collects royalties. The publisher usually takes half the royalties, and the composer receives the other half (some of which goes to the manager). The performing rights organizations also coordinate with performing rights organizations in other countries to collect and distribute fees for music played abroad (see Section 8). Costs are not deducted from the publishing royalties the band receives.

As is clear from Table 1, bands receive relatively little of their income from recording companies. Indeed, only the very top bands are likely to receive any income other than

the advance they receive from the company, because expenses – and there are many – are charged against the band’s advance before royalties are paid out. In 2003 the total value of recording sales (including CDs, singles, LPs, etc.) in the US was \$11.8 billion according to IFPI (2004), while the total value of concert ticket sales was \$2.1 billion according to our tabulations. Thus, from the consumers’ perspective, recordings are a much larger market, but from the artists’ perspective, concerts represent a much more important income source. This point was made by Scott Welch, manager of Alanis Morissette and LeAnn Rimes: “The top 10 percent of artists make money selling records, the rest go on tour.”<sup>1</sup>

### 2.1. *Contracts*

Contractual arrangements between bands, promoters and record labels are heterogeneous, but the typical contract resembles a book contract, with an initial advance and then royalties if sales exceed a certain level. The typical contract between a band and a concert promoter is most easily illustrated with a hypothetical example. Consider an agreement covering a single concert.<sup>2</sup> The band receives a “guaranteed advance” – e.g., equal to the first \$100,000 of ticket sales, and then, before additional revenue is distributed, the promoter recovers his expenses and a “guaranteed profit” – say \$50,000 for expenses and \$22,500 for profit. The expenses could include advertising, rent for the venue, costs of unloading the equipment, etc. The band also has expenses (e.g., travel), which it pays for out of its income. The promoter and the band split any ticket revenue above the guarantee plus expenses and minimum profit (above \$172,500 in this case), usually with the band receiving 85 percent and the promoter receiving 15 percent of these revenues.<sup>3</sup> The band’s guaranteed advance and percent of revenue after expenses is higher for bands with greater bargaining power.

In its negotiation with the promoter, the band (or its manager on the band’s behalf) agrees to the concert price, which naturally affects the amount of revenue collected. In addition, the band usually receives 100 percent of merchandise sales (e.g., T-shirts) that take place at the concert.<sup>4</sup> The venue usually receives the beer and parking revenue. An interesting economic question is why the contracts for concerts take this form. Because the parties receive revenue from the sources for which they are most responsible – the band and promoter from ticket sales, the band from merchandise sales, and the venue for parking and food – it is possible that this division of revenue streams provides optimal incentives for efficient provision.

<sup>1</sup> Quoted in Kafka and Powers (2003).

<sup>2</sup> It is interesting to note that as promoters have become more consolidated, more bands have signed nationwide tours with a single promoter.

<sup>3</sup> These hypothetical figures were used by the head of a major management firm to illustrate a “typical” contract.

<sup>4</sup> In some cases, the band will be required to give a proportion (e.g., 30 percent) of the merchandise sales to the venue for the right to sell there, however.



Promoters contract with a ticket distributor to distribute tickets. Tickets may also be distributed directly by the venue box office and by the band to its fan club. By far the largest ticket distributor is Ticketmaster. Ticketmaster also has exclusive arrangements to distribute tickets for some venues. Ticketmaster fees are usually around 10 percent of the list price. Unknown to the consumer, in some cases the venue, promoter or performers receive a portion of this fee, depending on their contract.

Record companies tend to sign long-term agreements with bands that specify an advance on royalties and a royalty rate. The typical new band has very little negotiating power with record labels, and the advance rarely covers the recording and promotion costs, which are usually charged to the band. Because fixed recording costs vary little with band quality, only the most popular artists earn substantial revenue from record sales.

In the following passage from his book, *So You Wanna Be a Rock & Roll Star*, Jacob Slichter (2004), the drummer for *Semisonic* (and grandson of former AEA President Sumner Slichter), describes a typical recording contract:

Thus, armed with an attorney and a manager, we began our negotiations with Elektra. Dan [the lead singer] would relay the developments of those negotiations after our evening rehearsals, when we went out for drinks. I leaned back in my chair, sipped merlot, and listened as Dan and John tutored me in the basics of record contracts.

Elektra would lend us money, called an *advance*, so we could pay for the recording costs of making an album. As I already knew, those costs would be high – studio rental could run \$2000 per day and recording could take months. Producers' and engineers' fees might add another \$100,000, not to mention mastering, flights, hotels, rental cars – we could easily spend \$250,000. If there were anything left over, we'd get to keep it, but it wouldn't amount to much.

In return, we would grant Elektra the exclusive rights to our recordings. As money from the sales of records came in, we would be allotted a percentage of the proceeds, known as *points*. In a typical deal, the band gets thirteen or fourteen percentage points. We'd have to give a few of our own points (four perhaps) to the producer of our record (producers typically get a fee *and* points). Then we'd be down to ten points. Before calculating the value of those ten points, however, Elektra would subtract a large percentage of the gross sales to account for *free goods*, records given away for promotional and other purposes. Thus, the amount on which our 10 percent was calculated would be reduced by 20 to 25 percent. So we'd be down even further, perhaps 10 percent on 75 percent of the wholesale album revenue. If our CD was sold in stores for fifteen dollars, the band's share of the revenue might be something between fifty cents and a dollar per CD. Would we get to keep it? No! Elektra would add up all of the expenses of recording and promoting our album – rock videos, radio promotion, touring costs, and so on. The total of those costs, which could run into the millions, would be our *recoupable debt* to the record company. Our share of each CD sold would be swallowed up by that debt

... When it came time to record and release future albums, any unpaid debt from our past albums would carry forward. In fact, even if we sold millions of records (in which case the size of our share would increase), we might never recoup. As one friend of mine joked, we'd be rock-and-roll sharecroppers. (pp. 34–36)

Caves (2000) analyzes the contractual arrangements in the music industry in terms of the efficient division of risk, incentives and rewards. He emphasizes that reputation and the prospect of repeated contracts are essential for contract enforcement. Eliot (1993) provides many colorful examples of malfeasance in music contracts. For example, the Beatles accused Capitol Records of failing to pay royalties on 19 million albums and singles. An audit revealed more than 20 separate areas where Capitol/EMI had “wrongfully accounted” for costs or revenue concerning promotion, manufacture and sales, resulting in \$19 million of unpaid royalties due the Beatles from 1969–1979. Caves prosaically notes that, “From the artist’s viewpoint, a problem of moral hazard arises because the label keeps the books that determine the earnings remitted to the artist” (p. 65).

An analogous problem arises with live concerts. The following remark by Sharon Osbourne (2002, p. 56) underscores the difficulty of contract enforcement in the concert industry: “My husband’s whole career, people stole from him. They walk off with thousands of dollars that’s yours. So the only way, unfortunately, for me is to get nasty and to get violent.” She described the following disagreement with John Scher, a legendary New York promoter, who claimed advertising expenses for ads placed long after a concert had sold out: “[H]e would not give in, and he was threatening that ‘Ozzy will never work in the New York area again.’ All this crap. So I got up and nuted him with my head, and then I kicked him in the . . .” Caves notes that contract enforcement in this industry relies heavily on repeated transactions among parties who value their reputations. The Osbourne method is apparently another contract enforcement mechanism.

### 3. Some theoretical issues regarding concert pricing

Here we consider some of the main theoretical issues in concert ticket pricing, the main source of performers’ incomes. As an economic good, concerts are distinguished by five important characteristics:

- (1) although not as extreme as movies or records, from a production standpoint concerts have high fixed costs and low marginal costs;
- (2) concerts are an *experience good*, whose quality is only known after it is consumed;
- (3) the value of a concert ticket is zero after the concert is performed;
- (4) concert seats vary in quality;
- (5) bands sell complementary products, such as merchandise and records.

Rosen and Rosenfield (1997) provide a thorough treatment of ticket pricing, devoting particular attention to price discrimination, the practice of charging different prices to

different customers.<sup>5</sup> Price discrimination tends to occur when marginal costs are below average costs. Because fixed costs for a concert are high relative to variable costs, and because high- and low-elasticity demanders can be sorted by seat location, price discrimination is possible. Furthermore, bands are likely to have monopoly power, deriving from the fact that they produce differentiated products and have loyal fans.

Rosen and Rosenfield consider a case where there are two types of seats, high quality and low quality. Buyers prefer high quality to low quality. The seller chooses the total number of seats and the quantity of each class of seat, and a pricing policy for complementary goods, such as merchandise. Buyers have reserve prices for high- and low-quality seats, conditional on the seat quality and prices of complementary goods. The seller knows the distribution of reserve prices, but cannot identify customers with high and low reservation prices; ticket quality is used to sort buyers. Rosen and Rosenfield show that the seller would solve the pricing problem in two steps: "First, given the quantities and quality of the two classes of seats and the price of complements, the seller chooses ticket prices to maximize revenue . . . Second, given the optimum pricing policy, the seller decides on the quantity and quality of seats and on the price of complements" (pp. 353–354).

The price of a concert ticket is set lower than it would be in the absence of complementary goods, because a larger audience increases sales of complements and raises revenue.

One puzzle in actual pricing is that price discrimination is surprisingly rare, as we will see in the next section. Another puzzle is that pricing results in excess demand for many concert performances, which leads to scalping; scalping is addressed in Section 5.

#### 4. Concert industry trends

This section, which draws heavily from Krueger (2005), makes extensive use of *Pollstar's* Box Office Report database to describe developments in the concert industry from 1981 to 2003. *Pollstar* is the trade magazine of the concert industry, and a widely recognized authority on concerts. Since 1981, the magazine has collected and published data on concert revenue, venue capacity, ticket sales and prices. The data are provided by venue managers, who have an incentive to report their data because *Pollstar* disseminates it to potential clients. Managers report data on a wide range of musical concerts, and occasionally on other entertainment events, such as comedians, professional wrestling matches and traveling Broadway shows. The data are most complete for concerts, and we tried to exclude the non-concerts from the sample. Before restrictions, the database contains 260,081 box office reports. After eliminating non-concerts, benefit concerts (which we think of as charity events), and events that occurred outside the United States, the sample contains 232,911 reports, representing 270,679 separate performances.

<sup>5</sup> Also see Courty (2000) for a thoughtful summary of theoretical issues in ticket pricing.

Reporting of concerts to *Pollstar* increased substantially in the 1980s, so one potential problem is that the dataset may not be representative of the entire concert industry in all years. Major acts are more likely to be included in the dataset throughout. As a partial adjustment for changes in sample composition, in some of the analysis we restrict the sample to artists listed in *The Rolling Stone Encyclopedia of Rock & Roll*, hereafter called *Encyclopedia* bands.<sup>6</sup> This *Encyclopedia* contains information on 1786 artists, and 1275 of these artists performed at least one concert represented in the *Pollstar* database. The edition of the *Encyclopedia* we use was published in October 2001; two earlier editions were published in 1984 and 1995. Thus, the *Encyclopedia* contains something of a moving average of the leading bands in the period under study, which produces more of a consistent sample. Bands listed in the *Encyclopedia* are responsible for 75 percent of the dollar value of ticket sales in the *Pollstar* data from 1981 to 2003.

Two other limitations of the data should be noted. First, the ticket price and revenue pertain to the list price. Any service fees charged by the ticket distributor are excluded. Because service fees have been growing rapidly in recent years, this omission probably serves to understate the acceleration in ticket prices in recent years. Second, we do not have information on the secondary market, and it might be common for tickets to be resold in a scalper market. Nevertheless, the list price, not the resale price, is relevant from the standpoint of artists and promoters, as their ticket revenue is derived from tickets sold at the list price. Moreover, fragmentary evidence summarized in Section 5 suggests that scalping is a less common phenomenon than widely believed.

#### 4.1. Trends in prices

Figure 2 displays the average price of a concert ticket (total revenue divided by total tickets sold each year) for all concerts from 1981 to 2003, and the (ticket-weighted) average high and low price of a concert ticket. The figure also shows what the average price would have been had it grown in lockstep with the CPI-U. From 1981 to 1996, concert prices grew slightly faster than inflation: concert prices grew a compound 4.6 percent per year while overall consumer prices grew 3.7 percent per year. From 1996 to 2003, concert prices grew much faster than inflation: 8.9 percent a year versus 2.3 percent a year. And if the sample of concerts is limited to those performed by bands listed in the *Encyclopedia of Rock & Roll* in an attempt to hold constant changes in composition and quality, the acceleration in concert prices after 1996 is slightly greater: 11.1 percent a year growth from 1996 to 2003 versus 4.9 percent a year in the 1981–1996 period.

The cost of the highest priced ticket in the house has grown even faster than the average ticket (see the top dashed line in Figure 2). Weighted by total ticket sales, the average high price ticket grew by 10.7 percent per annum from 1996 to 2003, while the average of the lowest price ticket grew by 6.7 percent a year. Thus, price dispersion increased across seats for the same concert. (The rise in income dispersion among

<sup>6</sup> George-Warren, Romanowski and Pareles (2001).

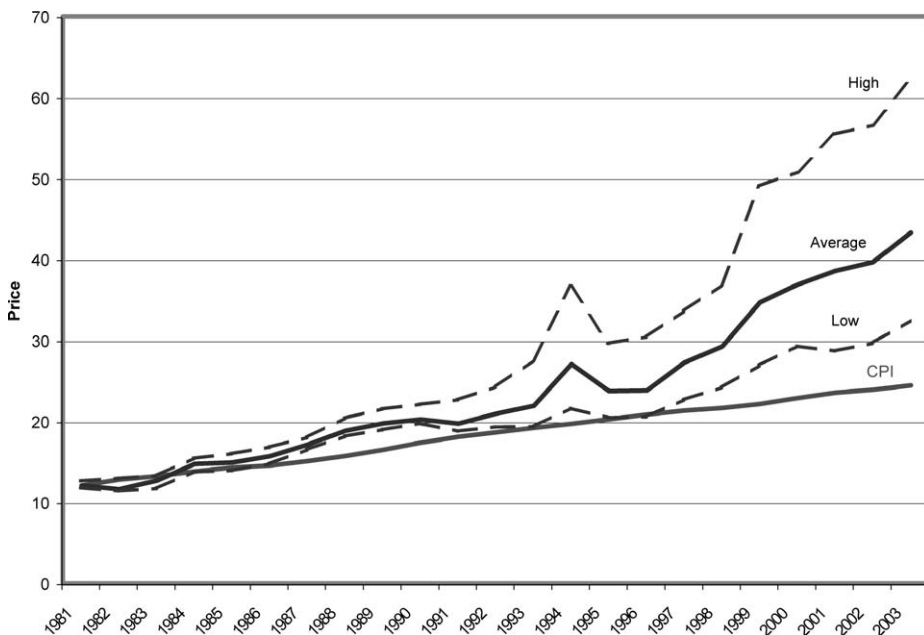


Figure 2. Average price per ticket, high and low price tickets, and overall inflation rate, 1981–2003. Source: Krueger (2005).

consumers may partially account for the rise in price differentiation; unfortunately, data on consumers is unavailable.) Nonetheless, in 43 percent of concerts in 2003, *all* seats in the house were priced the same, suggesting less price discrimination than might be expected from Rosen and Rosenfield (1997).<sup>7</sup> Even in venues with more than 25,000 seats, 26 percent of shows charged just one price for all seats in 2003. The amount of price differentiation has grown over time, however: in the 1980s, 73 percent of concerts with more than 25,000 seats charged just one price for all seats.

Instead of overall consumer price inflation rate, probably a more appropriate comparison for concerts is the price of other live entertainment events. Figure 3 reproduces Krueger's (2005) comparison of concert prices to the CPI-U sub-index for movies, sporting events and theater.<sup>8</sup> To make the data as comparable to the CPI as possible, a Laspeyres price index for concerts using the *venue* as the unit of observation was

<sup>7</sup> Larger concerts are more likely to vary prices. A quarter of all tickets in 2003 were for shows that had just one price, as compared to 43 percent of concerts.

<sup>8</sup> To be precise, the BLS produces a CPI for movies, sporting events, theater *and* concerts. A separate sub-index covering just movies, sporting events and theater is not available from BLS, so Krueger adjusted the index as follows. In November and December 2001, concerts accounted for 8.4 percent of price quotes for this sub-index (e-mail correspondence from Patrick Jackman, February 7, 2002). Consequently, Krueger netted out the concert component using his Laspeyres estimate of the concert CPI.

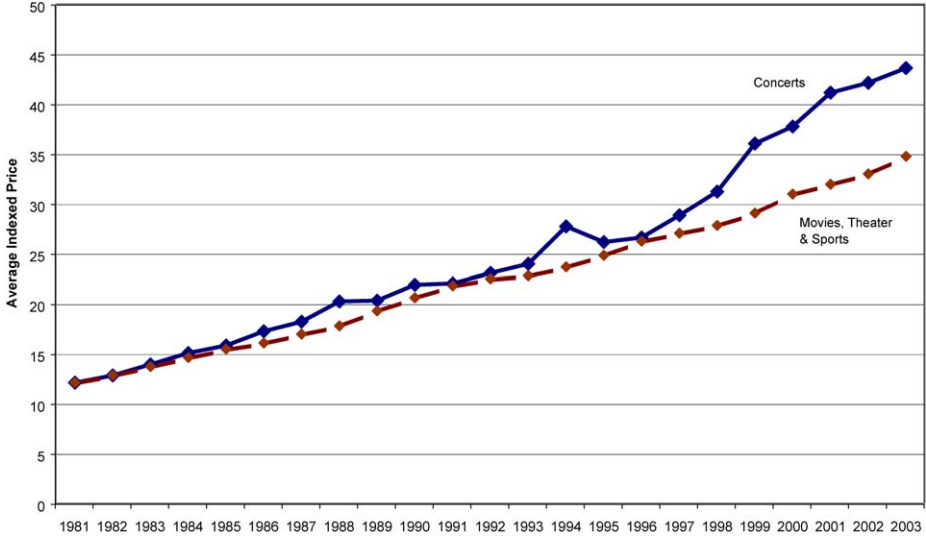


Figure 3. Concert prices tracked prices of movies, theater and sports tickets well until 1997. Laspeyres price index for concerts versus CPI-U for movies, theater and sports events. Source: Krueger (2005).

computed. It is clear that price growth for entertainment events exceeded overall price inflation throughout the period. Concert price growth tracked price growth for movies, theatre and sporting events remarkably well from 1981 to 1996, but beginning in 1997 the two series diverged. From 1997 to 2003, the concert Laspeyres index rose 64 percent, whereas the CPI for other entertainment events increased 32 percent.

4.1.1. More on price indices

Recall that the Laspeyres price index is defined as

$$L = \frac{\sum p_1 Q_0}{\sum p_0 Q_0}$$

and the Paasche index is defined as

$$P = \frac{\sum p_1 Q_1}{\sum p_0 Q_1}$$

where  $p$  is the price and  $Q$  is the quantity, and the subscript refers to either the base period (0) or the follow-up period (1). Intuitively, the Laspeyres index gives the proportionate increase in money needed to buy the exact same bundle of goods in the follow-up period as was purchased in the base period, and the Paasche index gives the proportionate difference in money if the bundle purchased in the follow-up period had been purchased in the base period at the base period prices. If tastes are constant – a strong

assumption for musical entertainment – and other assumptions are met, the Laspeyres Index is expected to overstate the cost of living, and the Paasche Index is expected to understate it. The Fisher Ideal index, which approximates a true cost of living index, is the geometric mean of these two indices:  $F = (LP)^{1/2}$ .

Left unstated in these formulas is the unit of observation. When the CPI is computed, the sum is taken over products within stores. For entertainment events, the venue is the unit of observation in the CPI. In essence, the BLS interviewers go to a venue and ask for the price this month, and compare it to last month's price, *regardless of what the performance is*. This could obviously create a good deal of noise in the price data, as the product being compared is not exactly the same. For example, in April 2004, Beyoncé, Alicia Keys, Missy Elliott and Tamia performed a concert at Madison Square Garden for an average price of \$81, and in May 2004, Yes performed there for an average price of \$61. The within-venue price index would record this as a decline in price, while it might more appropriately be viewed as a decline in quality. (As further support for this view, we note that Beyoncé et al. sold out, while Yes only sold 79 percent of the seats.)

An alternative to using the venue as the unit of observation is to use the performer as the unit of observation; that is, to follow the same movie or concert over time in different venues. Krueger (2005) computed a Fisher Ideal price index using the headline *band* as the unit of observation in an effort to hold composition constant. The artist was selected to more directly control for composition effects, although there are clearly problems with this approach as well: the venue could be larger or smaller, or in a more remote location, so the experience is different from concert to concert.

Thus, concerts by different performers in the same venue over time, or concerts by the same band in different venues over time are not the same products. It is therefore worthwhile to consider the impact of measurement error in prices on the various price indices. Suppose the baseline price is measured correctly, and the second period price is a noisy measure of the price of the same performance in the baseline. The simplest case is classical measurement error. Let  $p'_1 = p_1 + e$ , where  $p'_1$  is the observed price,  $p_1$  is the correct price (i.e., price for the same quality of performance), and  $e$  is a white noise, mean-zero measurement error. In this scenario, the Laspeyres and Paasche indices are still unbiased estimators, but the Fisher Ideal index will overstate the true rate of price inflation in the limit. The probability limit of the square of the Fisher Ideal index with the noisy price data in the second period ( $F'$ ) is:

$$p \lim[F'^2] = p \lim \left[ \frac{\sum P'_1 Q_0}{\sum P_0 Q_0} \cdot \frac{\sum P'_1 Q_1}{\sum P_0 Q_1} \right] = F^2 + \frac{\sigma^2 \sum Q_0 Q_1}{\sum P_0 Q_0 \sum P_0 Q_1}, \quad (1)$$

where  $\sigma^2$  is the (assumed constant) variance of  $e$ . Because the last term is positive, in expectation the Fisher index will overstate the value of the index if prices were measured without error. Intuitively, the reason the index is biased upward is because the error in follow-up period prices appears in the numerator of both the Laspeyres and Paasche indices.

If the error in prices were in the first period, the asymptotic bias would be in the opposite direction, because the variance of the errors would appear in the denominator.

Table 2

Various price indices for concert tickets and other entertainment events, using either the headline artist or venue as the unit of observation

Year	Artist			Venue			
	Laspeyres (1)	Paasche (2)	Fisher (3)	Laspeyres (4)	Paasche (5)	Fisher (6)	Movies, sports and theater (CPI) (7)
1981	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1982	112.8	108.9	110.8	106.0	106.3	106.2	106.1
1983	129.6	118.0	123.6	115.2	115.7	115.5	113.4
1984	143.8	126.0	134.6	124.7	127.0	125.8	120.8
1985	157.2	136.5	146.5	130.8	132.8	131.8	127.8
1986	166.5	144.9	155.4	142.7	142.1	142.4	133.0
1987	179.1	155.0	166.6	150.7	148.4	149.5	140.3
1988	199.6	171.6	185.0	167.4	165.2	166.3	147.1
1989	215.3	187.6	201.0	168.1	169.3	168.7	159.6
1990	236.0	200.3	217.5	181.2	185.1	183.1	170.5
1991	254.0	207.7	229.7	182.3	188.6	185.4	180.3
1992	273.9	214.3	242.3	190.8	198.8	194.7	186.0
1993	286.6	225.8	254.4	198.3	207.0	202.6	188.7
1994	310.0	209.5	254.9	229.3	235.1	232.2	195.7
1995	340.5	219.5	273.4	216.4	227.7	222.0	205.4
1996	398.5	234.6	305.8	220.1	225.2	222.6	217.1
1997	426.2	238.6	318.9	238.6	230.9	234.7	223.7
1998	518.0	273.9	376.7	258.2	251.5	254.8	230.1
1999	606.0	273.0	406.7	298.0	288.5	293.2	240.4
2000	671.7	300.9	449.6	312.2	304.7	308.4	256.1
2001	750.1	324.7	493.5	340.3	326.9	333.5	264.3
2002	802.1	334.9	518.3	348.5	336.1	342.2	272.9
2003	877.1	365.1	565.9	360.7	347.3	353.9	287.5
	Per annum percentage growth rate						
1981–1989	10.1	8.2	9.1	6.7	6.8	6.8	6.0
1989–1996	9.2	3.2	6.2	3.9	4.2	4.0	4.5
1996–2003	11.9	6.5	9.2	7.3	6.4	6.8	4.1

Notes: Authors' calculations based on Pollstar data and data from BLS. Index sets 1981 to 100. Weights are updated each year for columns 1–6.

It seems more natural, however, to think of the first period concerts as defining the quality standard.

Table 2 explores the effect of the unit of observation on the various price indices for the *Pollstar* concert data. The first three columns report the Laspeyres, Paasche and Fisher ideal indices, respectively, using the headline artist as the unit of observation. The next set of three columns report the same indices using the venue as the unit of observation. The seventh column reports the CPI for movies, sporting events and theater, based on BLS data, which also uses the venue as the unit of observation. The weights



used to compute the price indices for the concert data are updated each year, which is more frequent than the CPI.

Looking at [Table 2](#), it is immediately clear that the price growth is much greater if the artist is used as the unit of observation instead of the venue, especially for the Laspeyres index. This is probably a result of sample selection: only artists who perform in adjacent years can be used in the analysis if the artist is the unit of observation. These artists may not be representative of all artists, and their prices appear to be growing very rapidly, especially when base period quantities are used as weights. It is also interesting to note that when the venue is used as the unit of observation, the growth in the Paasche index exceeds that of the Laspeyres index in two of the three subperiods.

A final issue about price indices worth mentioning involves rationing. The price indices, which already have well known deficiencies as measures of the cost of living [see [Moulton \(1996\)](#)], are even more problematic if there is rationing. If a concert is sold out, there is likely some degree of rationing. In 2003, a third of tickets sold were to concerts that were sold out, down from 55 percent in the 1980s. These figures may overstate the amount of rationing, however, if artists perform multiple shows in the same city, and tickets are available for some shows.

#### 4.2. Shows, sales and revenues

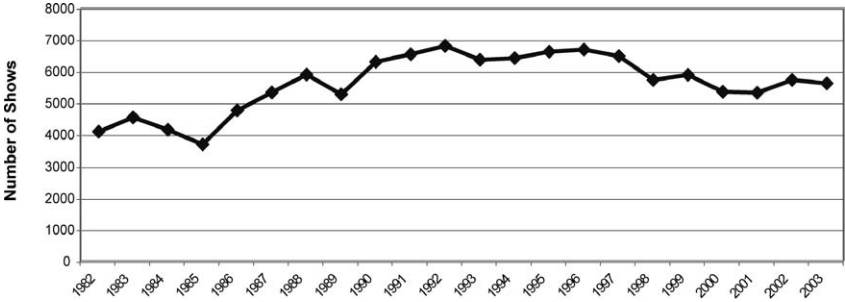
[Figures 4\(a\), 4\(b\) and 4\(c\)](#), taken from [Krueger \(2005\)](#), summarize trends in the number of shows performed, tickets sold, and revenue collected from 1981 to 2003. The figures restrict the sample to artists in the *Rolling Stone Encyclopedia* because coverage in the *Pollstar* database should be more consistent for these artists.<sup>9</sup>

Several trends are noteworthy. First, the number of shows performed rose in the 1980s, plateaued in the first half of the 1990s, and has declined by 16 percent from 1996 to 2003.

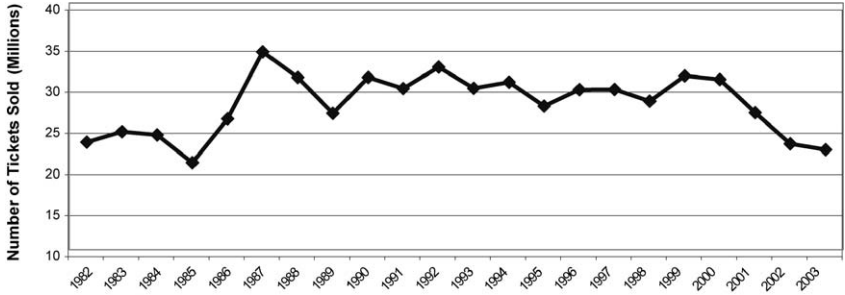
Second, the number of concert tickets sold by these bands fluctuated around 30 million per year from the late 1980s until 2000, and has dropped since 2000. In 2003, 22 million tickets were sold to concerts performed by these bands. The drop in ticket sales is also consistent with a Gallup poll, which found that the percentage of teenagers who said they attended a rock concert fell from 40 percent in 1976 to 31 percent in 2000. (By contrast, the percent of teens who said they attended a pro sports event rose from 43 percent to 63 percent over this period.)

Third, despite flat or declining tickets sales, total revenues (in 2003 dollars) trended upwards until 2000 because of price increases. Other things equal, these trends suggest the elasticity of demand was less than 1 before 2000. Since 2000, however, there has been a 10 percent drop in ticket revenue for these artists, suggesting that prices increases have been offset by a larger than proportional demand response.

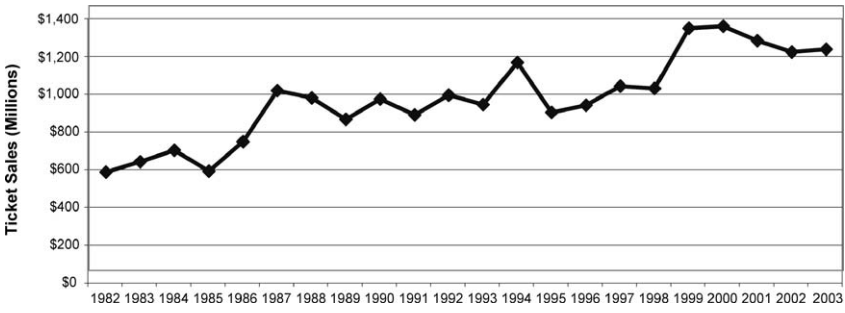
<sup>9</sup> The trend in capacity utilization for the full universe is similar to that for the *Encyclopedia* bands, but the number of shows and tickets sold has trended upwards if the larger sample is used.



(a)



(b)



(c)

Figure 4. *Rolling Stone Encyclopedia* Artists: (a) number of shows each year; (b) number of tickets sold each year; (c) total ticket revenue in 2003 dollars. Source: Krueger (2005).

Another trend worth noting is that the capacity utilization rate, or the fraction of available seats that are sold, has fallen over the last two decades. The fraction of tickets sold fell from around 90 percent in the late 1980s to just over 75 percent in 2003. Interestingly, the drop in the capacity utilization rate was much steeper for concerts held in larger venues.

One possible interpretation of these trends is that these artists are becoming less popular. But this view is hard to reconcile with the sharp increase in ticket prices for *Encyclopedia* bands. Instead, it seems that price growth is causing a movement up the demand curve for tickets.

#### 4.3. Distribution of revenues

As was documented, concert revenues increased in the 1980s and 1990s. Figure 5 displays the share of ticket revenue going to the top 1 percent and top 5 percent of all performers, ranked by their total annual concert revenue. Bear in mind that these are ticket revenues, and not income, but they still indicate how the fan dollars are allocated across the distribution of acts.

The figure indicates that concert revenues became markedly more skewed in the 1980s and 1990s. In 1982, the top 1 percent of artists took in 26 percent of concert revenue; in 2003 that figure was 56 percent. By contrast, the top 1 percent of income tax filers in the US garnered “just” 14.6 percent of adjusted gross income in 1998 [see Piketty and Saez (2003)]. The top 5 percent of revenue generators took in 62 percent of concert revenue in 1982 and 84 percent in 2003. Surely, this is a market where superstars receive the lion’s share of the income. We return to the issue of superstar effects in Section 7.

To further investigate the distribution of concert revenues, we followed De Vany’s (2006) Chapter 19 on movies in this *Handbook* and De Vany and Walls (2004), and

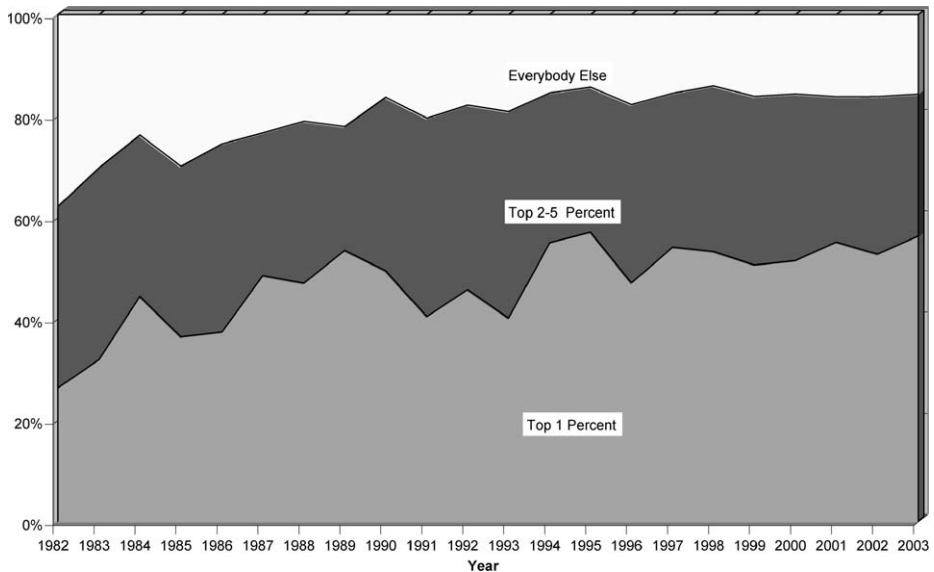


Figure 5. Share of total ticket revenue accruing to top performers, 1982–2003. Source: Krueger (2005).

fit a Pareto distribution to the revenue data. As is well known, a Pareto distribution is characterized by thick tails (on one or both sides), and thus provides a good fit for income distributions. The Pareto distribution is part of a more general class of stable distributions,  $S(\alpha, \beta, \gamma, \delta)$ , of which the Gaussian is also a special case. The parameter of interest here is  $\alpha$ , the tail weight, with  $0 < \alpha \leq 2$ . A tail weight of 2 implies a normal distribution. As  $\alpha$  approaches 0, greater weight is placed in the tail of the distribution. To estimate  $\alpha$ , we used a simple regression method.<sup>10</sup> We first assigned ranks to each artist's 2003 revenues, with rank 1 indicating the highest revenue, rank 2 the second highest, and so on. Then we regressed log revenue on the log of the ranks as follows:

$$\text{Log(Revenue)} = a - b \text{Log(Rank)}, \quad (2)$$

where the inverse of  $b$  is an estimate of  $\alpha$ . Note that in the class of stable distributions, the variance is infinite when  $\alpha$  is less than 2, and when  $\alpha$  is less than 1 the mean does not necessarily exist either.

We used this method to estimate  $\alpha$  for the distribution of artists' concert revenues, as well as for promoters' revenues, in 2003. We find a coefficient  $\alpha$  of 0.45 for artists' revenues, and 0.55 for promoters' revenues. In comparison, [De Vany \(2006\)](#) and [De Vany and Walls \(2004\)](#) estimate  $\alpha$  to be in the range 1.3 to 1.7 for motion picture box office revenues. This suggests that the concert performers' revenues are not only very far from being Gaussian, but they are also more skewed than movie revenues. Probably a more appropriate point of comparison for artists' revenues is actors' lifetime cumulative movie grosses, however, for which De Vany estimates an  $\alpha$  of 0.4 – very close to what we find for artists' revenues in 2003. Thus, the movie stars' lifetime revenues are positively skewed to about the same degree as concert performers' annual revenues.

Despite the infinite expected variance of revenues in the parametric distribution, in the finite sample of data we have the distribution of artists' revenues is fairly stable from year to year, with a correlation of 0.75 between revenues in 2002 and 2003. Promoters' revenues are even more stable, with a correlation of 0.98 between 2002 and 2003.

#### 4.4. *Explanations: Baumol and Bowen's disease; cartelization; Bowie theory*

[Krueger \(2005\)](#) examines several explanations for the coincidence of declining ticket sales and rising prices, which he notes is consistent with the market becoming more monopolized over time, and inconsistent with a downward shift in demand. We consider these in turn.

In some respects, popular music concerts are a slow productivity growth sector: it takes just as long and about as many people to perform a concert today as it did 20 years ago. As [Baumol and Bowen \(1966\)](#) point out, prices should rise faster than overall inflation in low-productivity growth sectors because of cost increases. Baumol and

<sup>10</sup> See [De Vany \(2006\)](#).

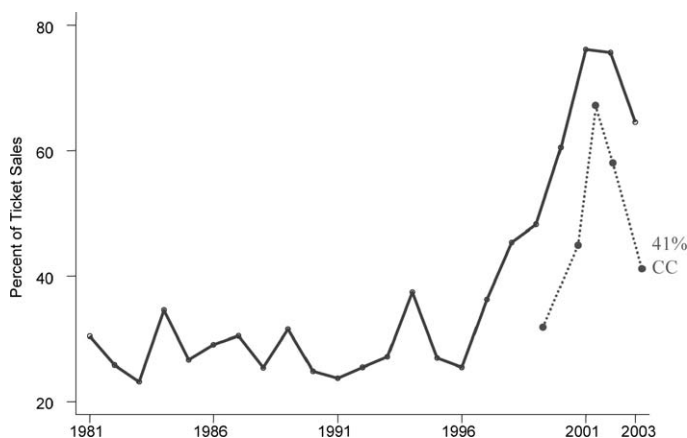


Figure 6. Percent of total revenue handled by biggest four promoters, nationwide and by Clear Channel Communications. Source: Krueger (2005) based on Pollstar data. Only concerts performed in the US are included in the analysis. Sample consists of artists listed in *Rolling Stone Encyclopedia*.

Bowen's disease may well account for the mildly faster price growth in live entertainment events than overall price inflation in the pre-1996 period. Yet it is unlikely that there was a discrete jump in costs in the concert industry compared to other industries – let alone other entertainment industries – after 1996. Indeed, reductions in the costs of audiovisual electronics equipment probably reduced the cost of concerts. Nevertheless, some concert promoters do point to an increase in production costs (e.g., pyrotechnics) and insurance costs as a rationale for the acceleration in prices.

Another popular explanation for the acceleration in concert prices is that the concert industry has become monopolized by Clear Channel Communications, the giant multimedia conglomerate. On the surface, there is an air of plausibility to this story. After the Telecommunications Act of 1996 relaxed constraints on radio station ownership, Clear Channel acquired nearly 1200 stations. It also owns amphitheaters, billboards and TV stations. Clear Channel entered the concert promotion business in a major way by acquiring SFX Entertainment in 2000. As shown in Figure 6, the share of concert revenue that Clear Channel promotes rose dramatically from 1999 to 2001 and then fell sharply in 2002 and 2003. Despite the recent dip, concentration in the industry is still high at the national level.

Many critics have accused Clear Channel of using its vertical and horizontal concentration to monopolize the concert industry. Although anecdotal evidence abounds, and some court cases have charged Clear Channel with anticompetitive practices, Krueger finds little evidence linking Clear Channel to the sharp growth in concert prices.

First, he finds that Clear Channel's share of listeners in the radio market in a city was unrelated to the share of ticket revenue for concerts promoted by Clear Channel in those markets in 2000 and 2001. Additionally, at either the city or state level, Clear Channel's share of concert promotion dollars was insignificantly or negatively related to

the growth in prices. It is possible that Clear Channel uses its muscle to sign up concerts for national or international tours, obscuring the city- and state-level correlations, but it is surprising that the company does not exercise its monopoly position at a regional level as well.

Another fact that casts doubt on Clear Channel's role is that ticket prices have also risen sharply in Canada and Europe since the mid-1990's. Although, to some extent, prices are arbitrated between countries because bands play across national borders, it is unlikely that deregulation of radio in the United States and the rise of Clear Channel can account for concert price growth worldwide.

Perhaps the most important strand of evidence against the concentration argument is that concert promotion has always been a highly concentrated business on a regional level. In the 24 largest cities, the four-firm concentration ratio within cities has hovered around 90 percent, on average, for the last two decades. The average within-city Herfindahl–Hirschman Index (HHI) for promoters actually fell from a lofty 4200 in 1986 to a still high but less lofty value of 2800 in 2001. (An industry with an HHI above 1800 is considered highly concentrated according to the Justice Department Merger Guidelines.) Thus, the concert industry has gone from having regional monopolies to having a large national firm, but within cities competition could quite possibly have increased.

Krueger's final hypothesis is that concert prices have accelerated because recording artists have seen a large decline in their income from record sales, a complementary product to concerts. Record sales slumped from 1999 to 2002, and were flat for 5 years before then, putting downward pressure on artists' royalties [see [Weinraub \(2002\)](#)]. As discussed in Section 9, it is quite possible that record sales are down because many potential customers frequently download music free from the Web or copy CD's, either legally or illegally.

Formally, the problem is one of a firm with two complementary outputs, concert seats and record albums, denoted good 1 and good 2, and monopoly power in both markets [see [Tirole \(1988\)](#) or [Rosen and Rosenfield \(1997\)](#)]. The demand curves for the band's products are denoted  $D_1(p_1, p_2)$  and  $D_2(p_1, p_2)$ , each of which depends on both prices. Costs are independent of each other and depend only on the quantity of the specific good produced,  $C_1(D_1)$  and  $C_2(D_2)$ . A profit maximizing band will set the proportionate markup of concert tickets over marginal cost so that:

$$\frac{p_1 - C'_1}{p_1} = \frac{1}{\varepsilon_{11}} + \frac{(p_2 - C'_2)D_2\varepsilon_{12}}{p_1 D_1 \varepsilon_{11}}, \quad (3)$$

where the  $\varepsilon_{ij}$ 's represent the value of the own- or cross-price elasticities of demand. Bands will keep the price of concerts below the single-market monopoly price if greater attendance raises record royalties, but if this is no longer the case because of file sharing or CD copying, the price of concerts will rise.

To some extent, this model was anticipated by the rock and roll singer David Bowie, who predicted that, "Music itself is going to become like running water or electricity"

and he advised performers, “You’d better be prepared for doing a lot of touring because that’s really the only unique situation that’s going to be left” [quoted from Pareles (2002)]. Hence the name *Bowie Theory*.

As support, Krueger (2005) notes that relative to album sales, jazz fans are much less likely to download music from the Web than are fans of rock and pop, and that from 1996 to 2003 concert prices increased by only 20 percent for jazz musicians, but by 99 percent for rock and pop performers.<sup>11</sup> The declining complementarities argument can also account for the price growth in Canada and Europe. Section 9 provides a detailed review of the direct evidence on the effect of file sharing on record sales, and concludes that the evidence is mixed. Thus, the reason for the sharp acceleration in concert prices remains something of an open question.

## 5. Ticket distribution and scalping

As was mentioned, promoters and venues utilize a variety of options for ticket distribution, including the box office, Ticketmaster, and direct sales to fan clubs. Tickets are almost always initially distributed at a fixed price, as opposed to a floating price determined by an auction or other mechanism. Ticketmaster and other distributors have recently begun experimenting with using auctions to sell tickets, however. We suspect that ticket auctions will be more prevalent in the future, and a worthy topic for research.

About a third of popular music concerts currently sell out. Tickets for the hottest concerts are often sold on a secondary market, through unregistered scalpers, over the web (e.g., eBay), or through ticket brokers (who can also be online). These distribution channels are often lumped together and viewed as a *scalper or secondary market*. Persistent pricing of tickets at a level that permits scalping is a puzzle for neoclassical economic models of concerts. Why don’t performers or promoters raise the price of tickets and capture some of the revenue from the secondary market for themselves?<sup>12</sup> Below we consider theoretical issues and available evidence on scalping.

### 5.1. Scalping: Theoretical issues

Various theories have been proposed for why a firm – restaurant, ski lift, or rock band – may price their services below the market level. None of them is entirely satisfactory. Becker (1991) presents a model in which eating at a popular restaurant (or going to a concert) is a social event, so customers’ demands are positively related. The bigger the

<sup>11</sup> Oberholzer and Strumpf (2004) note that jazz is the genre least downloaded on the Internet, but do not provide a reason. Perhaps it is that MP3 files are of lower quality than CDs, and that jazz enthusiasts value quality more than others.

<sup>12</sup> As an aside, we note that Warren Buffett recently came to this realization. Tickets for the annual meeting of Berkshire Hathaway were given to shareholders, and then often resold. Apparently, Mr. Buffett was distressed by this practice, and began selling tickets for \$5 a pair on eBay in 2004 to capture the secondary market.

audience, the more enjoyable the experience. Concert promoters and fans often do treat concerts like social events, lending some credence to this view. As Courty (2000) points out, however, although Becker's model "explains why a firm may not raise prices in the short run when capacity is fixed, it does not shed much light on the long run outcome that firms typically do not raise capacity to meet excess demand". Kahneman, Knetsch and Thaler (1986) argue that customers value being treated fairly, and the market clearing price may be considered unfair. Fairness is likely to be a more important consideration if attendance at a concert is viewed as a social event rather than an economic transaction.

Courty (2003) makes the insightful point that customers for live entertainment events have time-dependent demands. He presents a model in which there are two types of consumers: die-hard fans who want to see a concert and secure a ticket in advance, and others who are not sure if they will be free during the concert. As time elapses, the uncertainty is resolved for the latter group. The late-demanders have higher valuations of the event than the die-hard fans in his model. He further assumes that promoters cannot compete with ticket brokers or scalpers, and that die-hard fans outnumber the late deciders. With these assumptions, in equilibrium ticket brokers will buy tickets early at face value and resell them for a profit. Although the model requires many ancillary assumptions to prevent promoters from taking over the secondary market, the observation that some customers learn about their demand over time is undoubtedly an important feature of the ticket market.

## 5.2. Evidence on scalping

Because scalping is primarily an underground activity, little systematic empirical analysis has been done on secondary ticket markets. In an effort to make a small step toward closing that void, one of us (Krueger) conducted a survey of 858 fans at Bruce Springsteen and the E Street Band's concert at the First Union Center in Philadelphia on October 6, 2002, with the help of 12 Princeton students.<sup>13</sup> As was common in the past, every ticket in the house was originally sold for a single price, \$75 (plus service charge if distributed by Ticketmaster). The concert was part of the group's "The Rising" tour, and it quickly sold out when tickets were put on sale. Thus, the concert would be expected to have a high scalping rate.

Several results of the survey are worth noting. First, only 20 to 25 percent of the tickets were bought through a scalper or ticket broker or over the Web. Many industry analysts had expected a higher reselling rate prior to the survey. Scalping at the stadium was quite rare; it was much more common for the secondary market to clear through purchases from licensed ticket brokers or the web. Second, the average ticket that was resold went for around \$280, yet most fans paid the list price. Third, tickets for the

<sup>13</sup> The survey of fans was conducted shortly before the concert began. A stratified random sample of rows and sections was drawn. Weights were computed to make the sample representative of the entire venue. The response rate for the survey was very high. Ticketmaster and the First Union Center arranged for us to have access to the venue before the start of the show.



best seats were *less* likely to be resold than were seats in the upper deck, even though the consumer surplus was greater for the better seats. One interpretation of this finding is that serious fans queued for tickets (or applied to Ticketmaster early), and if they obtained a good seat they attended the concert and if they obtained a bad one they sold it. This finding, which was not anticipated, is consistent with how one would expect tickets to be allocated in a market: those who valued the best seats the most were the ones who sat in them. But one could argue that the distribution mechanism is inefficient (e.g., because of time wasted queuing and uncertainty), even if it mimics the market in terms of allocative efficiency.

Fourth, fans were asked when they purchased their tickets, in an effort to test Courty's model of scalping. The results yielded mixed support. On the one hand, tickets on the secondary market were purchased later than tickets sold by Ticketmaster or the box office, as expected. (Ninety percent of those who purchased their tickets from the box office or Ticketmaster bought their tickets more than a month before the concert, compared with 47 percent of those who bought from ticket brokers.) On the other hand, the price did *not* rise as the date of the concert approached, as Courty's model would seem to predict. Instead, prices on the secondary market fell as the day of the concert approached, consistent with the literature on the declining price anomaly in auctions [see Ashenfelter (1989)].

Fifth, the concert would have earned substantially more revenue if tickets were priced high enough to eliminate the secondary market. If the market price equaled \$280, the average price of a ticket in the secondary market, then \$4 million [=  $(\$280 - \$75) \times 19,738$  tickets] of additional revenue could have been collected by Springsteen and his band. Given that the actual revenues collected were \$1.5 million, this figure is staggering even if one allows for some error. The revenue foregone by the band in the secondary market alone was sizable, between \$1.1 and \$1.4 million, according to our estimates. These calculations suggest that, at least in the short run, performers sacrifice considerable income if they price their shows below the market rate.

An important cautionary note, however, is that these results pertain to just one concert, and it is unclear whether they generalize to concerts for other performers. The nature of the First Union Center, which is isolated alongside Interstate 95, may also have led to less on-the-street scalping than in other, more centrally located venues. But we would argue that replicating this type of survey in other concerts will yield valuable insights into secondary markets.

### 5.3. *Scalping and price trends*

An important question concerns the effect of the secondary market on the trends documented in Section 4. We have seen that the secondary market can be substantial, at least for some concerts. It is possible that the list price does not represent the price to consumers, because of widespread scalping. Perhaps the rise in ticket list prices has only cut out scalpers, and not affected the price to consumers.

Although we have no doubt that the secondary market is important in the popular music industry, the following three reasons lead us to doubt that a disconnect between the list price and price to consumers is responsible for the major trends documented in Section 4. First, the total number of tickets sold has declined. If concerts are no more expensive to consumers than before, then one would not expect to see attendance fall. Second, the decline in the capacity utilization rate also suggests that customers are finding concerts more costly. Moreover, even in the early 1990s, most concerts did not sell out, so it would have been possible to avoid the higher priced secondary market. Third, [Krueger \(2005\)](#) found that prices surged in the late 1990s even when he limited the sample to concerts that sold fewer than 90 percent of their tickets, events where scalping would have been unnecessary.

## 6. Rankings

It is common in the arts for various parties to devise schemes for ranking artists. Music is no exception. For example, Billboard provides numerous “music charts” based on record album sales and radio airplay. Being ranked high on the charts is important to artists because future sales and recording contracts are related to their placement on the charts. Evidently, many consumers turn to rankings to decide which music to purchase or listen to, and radio stations rely on the charts to determine which music to play on the air. When information is costly to obtain, rankings can be very valuable to consumers, especially for goods that have social externalities (e.g., when you play music at a party, you would like your guests to enjoy the music).

*Pollstar* produces three sets of rankings of bands: one based on gross concert revenue; one based on the number of tickets sold; and one based on the number of hits seeking information about each band’s schedule on its web page. Although useful, these methods have their limitations. An important limitation can be seen from [Figure 7](#). Hypothetical demand schedules for two bands, denoted A and B, are reported. As drawn, the demand curves all have the same slope but different intercepts. Band B is the most popular: at any given price, it has the greatest ticket demand.

We can write the demand curves in [Figure 7](#) as:

$$\text{Log } Q_i = a_i - \varepsilon \log P_i, \quad (4)$$

where  $Q$  is quantity of tickets sold,  $P$  is price,  $a_i$  is an indicator of band popularity, and  $\varepsilon$  is the elasticity of demand. The subscript  $i$  indicates the band. This constant elasticity demand curve is, of course, a simplification, but it illustrates a serious problem with current rankings, and provides an easily implemented solution. A more realistic model would also allow for the elasticity ( $\varepsilon_i$ ) to vary across bands, but greatly increase the parameters needed for implementation.<sup>14</sup>

<sup>14</sup> Another addition to the model would be to allow for rival bands’ prices to affect the demand for band  $i$ ’s concerts, and then take into account the effect of all other band’s price on the choice of band  $i$ ’s price. We will leave this extension for IO economists.

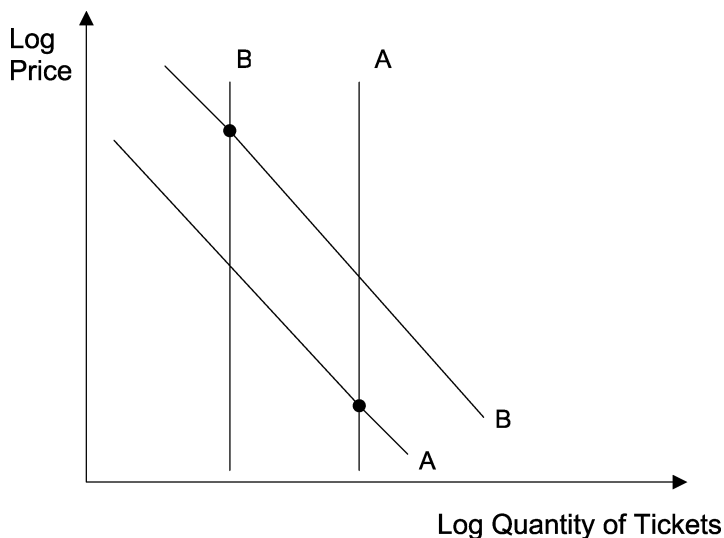


Figure 7. Hypothetical demand and supply curves for 2 bands.

Now if the market clears, the band's concert supply curve also affects revenues and ticket sales. (We will discuss disequilibrium shortly.) Bands have different concert supply functions. Suppose band B hardly tours (e.g., Barbra Streisand) and band A tours a great deal (e.g., Dave Matthews). Band A could sell more tickets than band B and collect more revenue if it wanted. This framework highlights problems with *Pollstar's* methods for ranking bands. First, it is clear from Equation (4) that as long as the band is on the demand curve, the quantity of tickets sold is not a good indicator of popularity because price varies. Second, total revenue would only be an appropriate measure of popularity in the unlikely event that the elasticity of demand equaled 1. Third, bands that tour frequently (or set prices lower) are likely to receive more hits from potential consumers on the web, but that reflects concert supply as well as popularity.

A simple solution is to rank the bands according to

$$a_i = \text{Log } Q_i + \varepsilon \log P_i.$$

To implement this solution, one needs an estimate of  $\varepsilon$ . The first two columns of Table 3 provide rankings of the top 50 bands using  $\varepsilon = 1$  and  $\varepsilon = 2$ .<sup>15</sup> The former corresponds to the *Pollstar* Top 100 Tour ranking, which is based on gross revenue.

If the market is in disequilibrium, the price and quantity may not be determined on the demand curve. In particular, if the band sets the ticket price below the market clearing level, the quantity of tickets demanded at the list price will exceed the quantity of

<sup>15</sup> For simplicity, we have ignored price discrimination, and just used the average concert price as a measure of  $p_i$ .

Table 3  
Alternative rankings of artists who toured in 2003

Artist	Rank1	Rank2	Rank3	Rank4	Rank5	Rank6
Bruce Springsteen & The E Street Band	1	2	1	3	5	1
Celine Dion	2	1	2	1	41	14
Fleetwood Mac	3	6	4	6	15	8
Eagles	4	3	6	4	9	13
Simon & Garfunkel	5	4	3	2	7	12
Cher	6	9	5	9	32	4
Aerosmith/KISS	7	8	7	8	12	7
Dixie Chicks	8	10	8	10	14	6
Billy Joel/Elton John	9	7	9	7	6	27
Dave Matthews Band	10	12	10	12	19	3
Summer Sanitarium Tour/Metallica	11	11	12	11	4	17
Toby Keith	12	16	13	17	54	5
The Rolling Stones	13	5	11	5	2	51
Kenny Chesney	14	21	14	21	56	2
Tim McGraw	15	14	15	15	36	15
Shania Twain	16	13	16	13	16	21
Justin Timberlake/Christina Aguilera	17	15	18	14	30	23
Jimmy Buffett	18	17	17	16	13	20
Phish	19	23	19	23	29	9
Pearl Jam	20	25	20	26	51	10
Ozzy Osbourne	21	20	22	22	21	24
James Taylor	22	24	21	25	53	16
Yanni	23	18	24	20	61	30
50 Cent	24	34	25	30	79	11
Bon Jovi	25	19	23	18	20	34
John Mayer/Counting Crows	26	32	26	31	43	18
Matchbox Twenty	27	35	28	37	111	22
Alabama	28	30	29	32	50	26
Red Hot Chili Peppers	29	36	27	36	59	25
The Dead	30	29	30	29	33	31
Michael Flatley's Lord of the Dance	31	31	32	34	392	32
American Idols Live	32	38	31	38	63	29
Alan Jackson	33	39	34	39	82	33
Brooks & Dunn	34	43	35	45	84	28
George Strait	35	27	33	27	26	45
Lollapalooza 2003	36	40	37	40	42	37
Steely Dan	37	33	40	35	75	55
Radiohead	38	44	36	41	31	39
Def Leppard	39	45	39	46	143	40
Bill Gaither & Friends Homecoming	40	67	41	70	125	19
ZZ Top	41	47	42	48	121	41
Santana	42	49	43	49	70	42
Widespread Panic	43	55	38	52	216	36
Journey/Styx/REO Speedwagon	44	41	46	42	60	49
Luis Miguel	45	28	45	28	47	81

(continued on next page)

Table 3  
(continued)

Artist	Rank1	Rank2	Rank3	Rank4	Rank5	Rank6
Elton John	46	37	44	33	34	77
Mana	47	46	48	44	44	60
Mamma Mia	48	42	52	43	226	70
Ben Harper/Jack Johnson	49	61	47	59	80	44
Trans Siberian Orchestra – East	50	58	50	56	136	54

Notes: *Rank1* assumes elasticity of demand is 1 (gross revenue); *Rank2* assumes elasticity of demand is 2; *Rank3* assumes elasticity of demand is 1 and that latent demand is 25% greater than ticket sales for sellouts; *Rank4* assumes elasticity of demand is 2 and that latent demand is 25% greater than ticket sales for sellouts; *Rank5* is based on revenue per performance; *Rank6* is based on number of tickets sold. Rankings are computed for all artists, but only the first 50 according to *Rank1* are shown.

tickets sold, and our method would not provide an accurate measure of  $a_i$ . We can still conceptualize notional demand curves in this situation, however. The challenge is to determine how much excess demand exists. As we saw in the previous section, at the Bruce Springsteen concert about 25 percent of tickets were purchased above the list price, suggesting that excess demand was at least 25 percent as large as the number of tickets sold. (An alternative method for estimating excess demand would be to use information on the number of willing – or at least interested – buyers who sought tickets from on-line sales venues after tickets were sold out.) A simple solution is to apply the 25 percent figure to all concerts that sell out. Accordingly, in columns 3 and 4 of Table 3, we inflated the quantity of tickets sold by 25 percent in all sold out concerts, and recomputed the rankings for  $\varepsilon = 1$  and  $\varepsilon = 2$ .

For comparison, in column 5 we present the ranking based on revenue per show, which can be thought of as a crude indicator of the performers' wage rate. If demand for artists' performances were infinitely elastic, as in a competitive market, this would provide a ranking of artists' potential income. And lastly in column 6 we report the ranking based on total tickets sold, which is one of *Pollstar's* criteria.

The results are sensitive to the type of ranking. Bruce Springsteen and the E Street Band, for instance, move from the top ranked artists by revenue in 2003 to second place when an elasticity of demand of 2 is used (or 3rd place if rationing is taken into account), to fifth place when revenue per show is used, and back to first place when the number of tickets is the basis of the ordering. Celine Dion is ranked second based on revenue and 14th based on tickets sold. The Rolling Stones move from 13th to 5th place when the elasticity is increased from 1 to 2. Overall, however, the rankings are fairly similar if popularity ( $a_i$ ) is the criteria (the correlation between the ranks in column 3 and 4 is 0.91), and quite different if revenue per show or total tickets sold is the criterion (the correlation between columns 4 and 5 is 0.56).

We should emphasize that our framework misses many important features of the concert industry. Most importantly, we have made an *ad hoc* assumption about the elasticity

of demand, and imposed the same elasticity for all bands. In addition, we have ignored advertising and promotion efforts, which are endogenous and undoubtedly influence ticket sales. A more complete approach would adjust for promotion efforts. Nevertheless, considering rankings in the framework of a simple supply and demand model highlights an often overlooked feature of existing rankings: popularity depends on both price and quantity. This simple insight applies to record sales as well as to concerts.

## 7. Superstar effects

As we saw in Section 4.3, the distribution of concert revenues is highly skewed, suggesting that the music industry is a superstar industry, where a small fraction of the performers earn a substantial share of the revenues. [Sherwin Rosen \(1981\)](#) was the first to provide a formal theoretical model to explain why “relatively small numbers of people earn enormous amounts of money and seem to dominate the fields in which they engage”. Building on the intuition of [Marshall \(1947\)](#), Rosen models a market where demand is characterized by imperfect substitution among the sellers (here, the performers), and where “the costs of production (writing, performing, etc.) do not rise in proportion to the size of a seller’s market”. At the heart of the imperfect substitution of performers is the notion of quality, or talent, of a performer. As [Rosen \(1981\)](#) puts it, “Lesser talent often is a poor substitute for greater talent. The worse it is the larger the sustainable rent accruing to higher quality sellers because demand for the better sellers increases more than proportionately: hearing a succession of mediocre singers does not add up to a single outstanding performance.” When combining the demand and supply as depicted above, Rosen ends up with a market equilibrium in which small differences in talent at the top of the distribution can account for large differences in revenue.

[Borghans and Groot \(1998\)](#) also address the issue of superstardom, arguing that a certain degree of monopolistic power of the artist generates higher revenues for superstars, on top of a difference in talent. They note that a stylized fact concerning superstars is that those whose talent is “suitable for media replication” earn much more than others. They argue that the availability of a mass media market gives the most talented artists an endogenous property right, derived from the fact that the public prefers to watch the best performances – consistent with the imperfect substitution assumption of Rosen’s model. Borghans and Groot conclude that, “Due to media production, only one person is needed to serve the whole market, where without this technology many producers are needed. Efficient allocation requires the most talented producer to be assigned to this task, but in practice the situation provides this person with an opportunity to exploit the number-one position.”

[Adler \(2006\)](#), in his [Chapter 25](#) in this volume, takes exception with Rosen’s view of talent, and maintains that superstars do not exist because of differences in talent, but because of “the need of consumers to have a common culture.” Because Adler’s chapter thoroughly addresses this and related issues, we tread lightly on theoretical aspects of superstar models.

Empirical evidence and tests of the superstar model are not straightforward, because of the lack of reliable income data and, more importantly, because of the inherent difficulty of objectively measuring talent or quality in a meaningful metric apart from economic success. What is a small difference in talent? On what objective, cardinal metric is Celine Dion only slightly more talented than Rod Stewart? As Krueger (2005) notes, “An objective measure of star quality for popular musicians is hard to define and even harder to quantify.” Measuring talent on a meaningful scale independently of economic success is an obstacle to testing the superstar model.

Hamlen (1991, 1994) looks at singers of popular music, and uses a measure of voice quality to assess the artist’s quality. The measure, which is external and objective, is the harmonic content of a singer’s voice sample. Hamlen then regresses the value of total record sales on harmonic content and a few observables for 107 singers, and finds an elasticity of 0.14. Since the Marshall–Rosen model would predict an elasticity above unity, Hamlen concludes that his empirical findings do not support the superstar model. Krueger (2005) points out, however, that “it is unclear whether the scaling of units of quality is appropriate (a different scaling could produce an elasticity above 1) and consideration of other dimensions of star quality could possibly rescue the theory”.

Krueger (2005) considers escalating superstar effects – perhaps due to the revolution in consumer electronics equipment which reduce the cost of copying and listening to music – as a possible explanation for the rising cost of concert tickets and increased concentration in concert revenue, which were documented in Section 4. Specifically, he tests whether the increase in prices (or revenue per artist) could be linked to a stronger superstar effect over in the 1990s. He uses a novel measure of star quality: namely, the number of millimeters of print that are devoted to each artist in *The Rolling Stone Encyclopedia of Rock & Roll*.<sup>16</sup> This information is then merged with the *Pollstar* data on concert revenue and prices, and the following regression is estimated:

$$\ln Y_{it} = \alpha + \beta_t S_i + \mathbf{x}'_{it} \boldsymbol{\gamma} + \delta_t + \varepsilon_{it}, \quad (5)$$

where  $\ln Y_{it}$  is the log average price (or log annual revenue or log revenue per show),  $S_i$  is the measure of Star Quality,  $\mathbf{x}'_{it}$  is a vector of covariates (number of supporting acts, years of experience of the band, and dummies for genre, gender and foreign status),  $\delta_t$  is a set of 22 unrestricted year fixed effects, and  $\varepsilon_{it}$  is an error term.

Notice that the coefficient on star quality,  $\beta$ , has a  $t$  subscript, indicating time period (1981–1986, 1987–1991, or 1997–2003). This allows the effect of star quality to vary across time periods. In the regressions, this is accomplished by interacting the amount of print with dummies indicating the four periods. The test of the rising-return-to-superstardom hypothesis amounts to a test of whether there is a discrete jump in  $\beta_t$  after 1996.

Krueger finds that the return to superstardom has indeed increased over time, but that the timing does not coincide with the increase in ticket prices. He concludes that we must look at other factors to explain the rising prices.

<sup>16</sup> See Krueger (2005) for a more detailed explanation of the data and procedures.

Empirical testing of superstar models lags behind the development of new theoretical versions of the model. At least when it comes to music, and probably for many other branches of the arts, a major limitation of tests of superstar models is the absence of natural units with which to measure talent. Rosen postulates that “small differences in talent become magnified in larger earnings differences”. But what is a “small” difference in quality of popular music performers? Surely there is an intrinsically subjective component to quality; some music appeals to a subset of listeners but not to others. Also, one might argue that talent should be measured within genres: otherwise, how is it possible to compare a jazz band to a heavy metal band? More empirical work could certainly be done on the superstar model, and perhaps the popular music industry could help shed light on some distributional and marginal return to labor issues in broader fields, but we are skeptical that current methods of measuring talent will shed much light on the superstar model.

## 8. The world of radio broadcasting

Ever since radio broadcasts started in the United States, England, and other countries in the early 1920s, the business of radio has been intertwined with that of music. As we will see, even if at first record companies and music publishers’ profits were threatened by the supply of “free” music on the radio, they quickly learned to promote their records and collect royalties from performing rights sold to radio broadcasters. Now, bands and composers also benefit from radio exposure. From a publicity standpoint, radio is an important part of record promotion. And from a royalties standpoint, composers can garner substantial returns if they have a hit song on the radio. [Table 1](#) documents that artists receive substantial revenue from performing rights royalties.

### 8.1. *Royalties from performing rights*

Under Section 106 of the US Copyright Act, a copyright on a musical work grants an exclusive right to reproduce, distribute copies, publicly perform, and create a derivative of the work in question. Thus, anyone who wants to legally play a copyrighted song on the radio, or press it on a compilation CD, must acquire a license to do so from the copyright owner. Artists generally transact with music publishing firms, which are often but not always affiliated with their record company, to collect their publishing income. Music publishers acquire administrative rights from the copyright owner, which entitle them to find users, issue licenses, collect money and pay the songwriter. The traditional split of publishing income is 50/50 between the publisher and the songwriter [see [Passman \(2000\)](#) and [Krasilovsky et al. \(2003\)](#)].

Various uses of a musical work are covered by different rights that must be purchased separately. [Table 4](#) summarizes these rights.

The *reproduction right* is the exclusive right of a music copyright owner to authorize the mechanical reproduction of the work in a record, cassette or CD. The license



Table 4  
Rights attached to musical compositions

Right	What it covers	Standard rate
Public performance right	The right to publicly perform a composition, for example, on the radio, in a club, in a concert, or on a jukebox	Blanket license via a performing rights organization (ASCAP, BMI, SESAC), rate based on factors such as advertising revenues and size of audience reached
Compulsory mechanical right (called compulsory because the composer cannot refuse to grant it once he gets paid)	The right to record and distribute recordings of a composition, only once it has been made public	8.5 cents per composition, or 1.65 cents per minute, whichever is greater
Synchronization right	The right to use a sound recording in a movie, commercial, or TV program (must be coupled with a performance right)	It depends on the length used and the use itself (background, integral part)

Source: Krasilovsky, Shemel and Gross (2003) and Passman (2000).

granting such a right is called a *mechanical license*, and the fees charged for it are the mechanical royalties, calculated at a certain rate per song and per unit manufactured and sold.<sup>17</sup> In the case of audiovisual productions, the license of reproduction rights is often referred to as a *synchronization license* because the music is to be synchronized with the images.

The public performance right gives the copyright owner the exclusive right to authorize the use of the musical work in public. Radio and television broadcasts, as well as jukeboxes and music played in bars, restaurants and any type of public establishment, all fall under the public performance right. As such, broadcasters and establishment owners must acquire public performance licenses anytime they want to use copyrighted music. Because searching and bargaining with every single copyright owner and publisher would prove costly and infeasible, and because an individual owner of copyrighted music could not possibly survey all the radio stations to enforce his public performance right, all licenses are handled by performance rights organizations (PROs). PROs issue public performing licenses to broadcasters and establishment owners, monitor and survey radio and television broadcasts to determine the amount of airplay for each composition, and then remunerate the copyright owners.

Performing rights also cover cellular phone *ringtones*, which cell phone owners increasingly download from the Internet. In 2003, ringtones were a \$2.5 billion industry

<sup>17</sup> Once a work has been made available to the public (after its first-use), a copyright owner is obligated to grant a mechanical license to anyone paying the statutory rate. For this reason, it is called a compulsory mechanical license.

worldwide [Flynn (2004)]! PROs have struck deals with the ringtone providers, and now compensate composers for each ringtone downloaded. On their website, BMI, one of the American performing rights organizations, boasts of having deals with 175 ringtone providers reaching more than 90 percent of US cell phone subscribers. BMI's payments are of 5¢ per ringtone, 2.5¢ each for the publisher and the composer.<sup>18</sup>

In the United States there are three performing rights organizations, typically known by the acronyms: ASCAP, BMI and SESAC. All offer blanket licenses, which grant the right to use all the songs in their respective catalogs.<sup>19</sup> Artists and composers can sign on with only one PRO. Radio stations can contract with multiple PROs. A radio station that wanted to play both Springsteen and Madonna, for example, would need to contract with both ASCAP and BMI.

Founded in 1914, the American Society of Composers, Authors, and Publishers (ASCAP) is the oldest of the performing rights organizations. It was the first organized effort to collect fees for public performances of music, which had been protected since the inception of the Copyright Act of 1897. In its first few years, ASCAP struggled to establish itself and persuade publishers to become members. Only in 1921 did it write its first royalty checks to publishers and composers. That is also the time when radio broadcasting began, suddenly creating a whole new market for ASCAP compositions. Broadcasters, however, were reticent to pay for the performing rights, arguing that once they had a copy of the record they were allowed to do whatever they pleased with it, including playing it on the radio. By 1932, the radio lobby had convinced seven states to outlaw ASCAP, on the basis of illegal racketeering practices and attempted extortion. Eliot (1993) relates: "In 1940, as many of the contracts ASCAP held were about to expire, the organization threatened to withdraw all member recordings if radio stations didn't agree to a broad-based, cohesive form of royalty payment . . . In retaliation, even as the FCC was threatening to outlaw the paying of records on the air, which ASCAP felt was largely the result of the broadcast lobby, station owners decided to start their own organization, to break what they claimed was ASCAP's monopolistic tactics."

Kleit (2000) notes that ASCAP's blanket license rate rose from 2 percent to 7.5 percent in the 1930s, parallel with the popularization of radio. Needless to say, the radio broadcasters were unhappy with the rising cost of the copyrighted music. To increase competition and to provide an alternative to writers and publishers not represented by ASCAP, the National Association of Broadcasters, together with NBC and CBS, created Broadcast Music Incorporated (BMI) in 1939.

ASCAP is now a not-for-profit entity owned by its members. The membership totals more than 180,000, including composers, songwriters, lyricists, and music publishers of every kind of music. Approximately 100,000 new songs are added to the catalog every year. The fees charged by ASCAP for its blanket license are not based on the amount of

<sup>18</sup> See BMI's website at [www.bmi.com/news/200406/20040616b.asp](http://www.bmi.com/news/200406/20040616b.asp).

<sup>19</sup> See Passman (2000), Krasilovsky, Shemel and Gross (2003), and Besen, Kirby and Salop (1992), as well as [www.ascap.com](http://www.ascap.com), [www.bmi.com](http://www.bmi.com), and [www.sesac.com](http://www.sesac.com) for more on the different American PROs.

airplay its music gets, but on the station's or venue's gross revenues less certain adjustments. The current basic rate is just under 2 percent of the adjusted gross advertising revenue for radio stations. BMI is a non-profit company owned by broadcasters. It now represents approximately 300,000 songwriters, composers, and music publishers in all musical genres, and its website mentions a repertoire of about 4.5 million compositions. BMI's blanket license rate for radio stations is about 1.6 percent of adjusted gross advertising receipts.

The Society of European Stage Authors and Composers, or SESAC, is the smallest of the three American performing rights organizations, with a market share estimated at 3 percent. Market shares are not easily computable, but [Krasilovsky, Shemel and Gross \(2003\)](#) report that a 1990 court proceeding concerning pay cable determined the current market shares at 54 percent for ASCAP, 43 percent for BMI and the remaining 3 percent for SESAC. SESAC is a for-profit private licensing company founded in 1930. It currently represents over 8000 publishers and writers and has a repertoire of more than 200,000 compositions. SESAC specializes in country and Latin music, and operates differently from ASCAP and BMI. It has a somewhat more selective procedure to accept new writers in their catalog, but also it charges fees for blanket licenses based on fixed determinants, such as market population served by the radio station and the station's standard advertisement rates.

While law papers on the topic abound, few articles have been written about performing rights in the economic literature. Notable exceptions include [Besen, Kirby and Salop's \(1992\)](#) article on copyright collectives and [Kleit's \(2000\)](#) study of competition among PROs. Besen, Kirby and Salop present an economic model that attempts to explain why copyright collectives are formed, how they operate and how they may compete. They start with a model of an unregulated monopoly copyright collective, where the cost for each individual copyright owner to collect fees from the broadcasters is prohibitively high, justifying the formation of the copyright collective as a means of saving costs. A collective also gives the copyright owners the possibility of cooperative price setting, and thus of more market power vis-à-vis big broadcasters. Besen et al. look at both models where the collective has the ability to limit the membership, and where it lacks such ability. They analyze the competition between the collectives as well as the effect of different types of government regulation. No formal statistical tests of their models are presented, although their predictions accord with certain stylized facts. As [Johnson \(1992\)](#) notes in a commentary, Besen et al. address the puzzle of the coexistence of ASCAP and BMI, but cannot test their suggested answer.

Interestingly, in most countries there is only one copyright collective. Besen et al. suggest three explanations for this fact: "First, government regulation may authorize only a single collective to administer a particular right. This is especially likely in countries like Austria, Germany, and Switzerland, where collectives must be licensed by the government. Second, government policies that mandate open entry and equal treatment of members may lead to a single collective . . . Third, efficient negotiation between the monopoly collective and user groups may eliminate any incentive for competitive entry." Why exactly did three organizations come to coexist in the US is still something

of a mystery. Besen et al. suggest that perhaps ASCAP miscalculated its hold on the market, and by requesting too high fees, it led excluded songwriters and broadcasters to form a collective of their own.

Kleit (2000) takes the blanket licenses offered by the PROs as a form of bundling, or block booking. He proposes a model of competition between PROs using blanket licenses, and shows that such licenses lead to higher profits for the PROs and higher costs for the users of the copyrighted music when there are a small number of competing licensing organizations.

### 8.2. *Music publishing in the US*

The United States is by far the largest market for music publishing. Table 5 gives a breakdown of revenues by source of income. In 2001, the performance-based revenues alone almost reached \$1 billion, for a total publishing income of nearly \$2 billion. That represents 29.3 percent of the world publishing income. By comparison, Germany, the second biggest market, shows a total income of just over \$800 million, for 12.2 percent of world income (see Table 7 below).

### 8.3. *Foreign markets*

In the US, both ASCAP and BMI earn just above 20 percent of their revenue from foreign sources. Performing rights organizations have agreements with their affiliates

Table 5  
Revenues from music publishing in the US in 2001 (millions of US dollars)

Performance-based income	
Radio	317.17
TV/Cable/Satellite	381.09
Live performance & recorded	216.40
	914.66
Reproduction-based income	
Phono-mechanical	552.70
Synchronization	102.31
	655.01
Distribution-based income	331.85
Interest investment income	37.10
Misc.	1.80
TOTAL	1940.42

Source: *NMPA International Survey of Music Publishing Revenues*, twelfth ed., Table 6, Master Survey Data [National Music Publishers' Association, Inc. and The Harry Fox Agency, Inc.].

Table 6  
List of performing rights organizations

Country	Organization	Acronym
Unites States of America	American Society of Composers, Authors, and Publishers; Broadcast Music Incorporated; Society of European Stage Authors and Composers	ASCAP; BMI; SESAC
Germany	Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte	GEMA
Japan	Japanese Society for Rights of Authors, Composers and Publishers	JASRAC
United Kingdom	Performing Rights Society	PRS
France	Société des Auteurs, Compositeurs et Éditeurs de Musique	SACEM
Italy	Società Italiana degli Autori ed Editori	SIAE
Spain	Sociedad General de Autores y Editores	SGAE
The Netherlands	BUMA-STEMRA	BUMA-STEMRA- CEDAR
Canada	The Society of Composers, Authors, and Music Publishers of Canada	SOCAN
Switzerland	Société Suisse des Auteurs, Suisse Auteurs	SSA, SUISA

in other countries to share revenues. They both collect revenues from abroad for their national members, as well as collect from users in the United States on behalf of foreign PROs. Table 6 lists the PROs in the US and the other top ten markets in the world.

While figures on the flows of revenues from copyright licenses from country to country are hard to obtain, we can observe total revenues for each country. Table 7 shows the top ten countries and the breakdown of the publishing revenues.

Table 8 shows ASCAP's flows of revenues from foreign publishing companies. Overall, we can see that the balance is positive, meaning that the United States (or at least ASCAP) is a net exporter of musical talent to the rest of the world.

#### 8.4. Payola

Payola is the practice of record companies giving cash or gifts to radio stations in exchange for airplay. Payola is interesting both for its history and from an economics standpoint because payola is an illegal economic transaction. Payola has been a federal criminal offense since 1960.<sup>20</sup> One may ask, What is wrong with payola? It could be argued that payola only creates a market for radio hits, a market in which the amount paid by the record company to the radio station becomes the price of a hit. It is akin to advertisement and promotion of a song by a record company. Indeed, if record companies

<sup>20</sup> It is legal for a record company to directly pay a radio station to broadcast a song as long as an announcement is made to the public. Undercover and undisclosed payments are illegal.

Table 7  
Publishing income in the top ten countries in 2001 (millions of US dollars)

Country	Performance-based income	Reproduction-based income	Distribution-based income	Interest investment income	Misc.	2001 Grand total	Total world income (%)	Cumulative (%)
USA	914.66	655.01	331.85	37.10	1.80	1940.42	29.3	29.3
Germany	305.28	318.81	153.72	30.55	0.00	808.36	12.2	41.5
Japan	291.17	350.74	49.64	0.50	67.60	759.64	11.5	52.9
United Kingdom	260.11	321.75	72.65	8.05	7.17	669.73	10.1	63.0
France	320.80	166.58	61.17	0.00	0.00	548.55	8.3	71.3
Italy	257.01	73.93	22.90	0.00	0.00	353.83	5.3	76.7
Spain	70.51	114.43	2.15	9.68	0.00	196.77	3.0	79.6
The Netherlands	78.03	53.21	29.22	16.12	0.00	176.57	2.7	82.3
Canada	71.40	44.39	18.84	4.53	0.00	139.17	2.1	84.4
Switzerland	50.08	24.71	25.83	5.01	0.00	105.63	1.6	86.0
<i>Top ten total</i>	<i>2619.05</i>	<i>2123.56</i>	<i>767.97</i>	<i>111.54</i>	<i>76.57</i>	<i>5698.67</i>	<i>86.0</i>	

Source: *NMPA International Survey of Music Publishing Revenues*, twelfth ed., Table 6, Master Survey Data [National Music Publishers' Association, Inc. and The Harry Fox Agency, Inc.].

Table 8  
ASCAP's foreign relations (millions of US dollars)

	2002	2003
Amount received from foreign publishing	148,027	180,309
Amount distributed to foreign companies	133,253	149,526
Balance	+14,774	+30,783

Source: Jim Steinblatt, ASCAP Media Relations, personal communication.

are willing to pay to promote their songs on the radio, it must be that radio promotion translates into higher record sales.<sup>21</sup>

So why has payola become illegal? Perhaps an analogy is instructive. Payola is analogous to a professor paying bribes to the editor of the *American Economic Review* to publish his paper. The professor would be willing to pay since a publication is good for career advancement, and eventually translates into higher future earnings. But AER readers expect the published articles to be the best and most relevant to the field, not the ones written by those with the deepest pockets or the most eager to get tenure. An essential function of a scientific journal is to screen papers. One could argue that an essential feature of a radio station is to screen records, especially since the right to broadcast on the radio waves is licensed by the government.

Payola has a colorful history. Payola is a contraction of the words “pay” and “Victrola”, an early type of LP record player.<sup>22</sup> The first laws and court cases involving payola were in 1960, but payola had been around for much longer, and still persists today, albeit under a different name. Coase (1979) traces the history of payola, going as far back as 1867 in England. Of course, back then the payments were not made to radio stations, but to public performers, with a request to play a song from the publisher's catalog. The agents that involved in this business were referred to as song-pluggers, and it became commonplace for vaudeville singers to be compensated for adding certain songs to their repertoire. When radio came about, song-pluggers turned to big bands performing live on radio stations to plug their songs. And then, when records made their appearance on the air, radio stations and their employees were approached by record companies to play their songs.

Coase (1979) explains, “Payola took the form of cash payments (which might be on a regular weekly or monthly basis), royalties on the sales of records, a share in a record company, advertisements in the disk jockeys' hit sheets, the reimbursement of recording stars' fees for appearances on the disk jockeys' programs or at record hops which

<sup>21</sup> Liebowitz (2004) points out that even though radio spins seem to increase sales of the particular record being spun, it does not mean that the recording industry as a whole benefits from radio broadcasting. Indeed, record sales fell in the first half of the 1920s after the popularization of the radio.

<sup>22</sup> See <http://www.history-of-rock.com/payola.htm>.

they organized, expensive gifts, and mortgage loans on disk jockeys' homes." Early on, payola was viewed as an impediment to competition. Many attempts were made to outlaw the practice, but these attempts only succeeded in pushing payola underground. The situation changed in 1959, when the president of the American Guild of Authors and Composers wrote a letter to the FCC (Federal Communications Commission) and the FTC (Federal Trade Commission) about payola and other deceptive practices, urging a congressional inquiry.<sup>23</sup> After a year of widely publicized hearings, the FCC amended the Communications Act of 1934 to make unannounced payments to deejays a criminal offense. Over twenty-five deejays and program directors were exposed in the scandal, but the top two deejays in the country, Dick Clark and Alan Freed, were the hardest hit.

Most of the pressure to outlaw payola came from ASCAP, which lost ground to BMI-licensed rock and roll records from small independent record labels during the 1950s. Coase (1979) points out that "during the period 1948 through 1955, 68 percent of the tunes which were number 1 on *Billboard's* top hits were controlled by ASCAP, and ASCAP's share was never less than 50 percent (in 1951). In 1956, its share was 23 percent, in 1957 and 1958, 25 percent and in 1959, 31 percent. In the circumstances, it is hardly surprising to find that the suppliers of 'good music' [ASCAP-licensed music] came to the conclusion that something was wrong with the economic organization of the popular music industry."

After 1960, program directors took over the playlist and left the disk jockeys out of the loop, shielding them away from payola charges. However, the pay-for-play business did not stop there. Soon, what became known in the industry as "independent record promoters" started acting as middlemen between the record companies and the radio stations, blurring the transactions and making this sort of payment not quite payola, but with a similar result and intent. Coase (1979) sees this as inevitable: "When a pricing system is not used and something of value is provided for nothing, people are willing to incur costs up to its worth in order to secure the benefits of that service." He goes on to argue that a payment system "is both natural and desirable", and that a ban on payola leads to a lower real income of the community. True enough, independent record promoters, or *indies*, could also be compared to food and beverage distributors who pay for placement in grocery stores, facilitating the connection between wholesaler and retailer. But Boehlert (2001) warns that "radio isn't really retail – that's what the record stores are. Radio is an entity unique to the music industry. It's an independent force that, much to the industry's chagrin, represents the one tried-and-true way record companies know to sell their product."

How big is the independent promoter business? Boehlert (2001) explains: "There are 10,000 commercial radio stations in the United States; record companies rely on approximately 1000 of the largest to create hits and sell records. Each of those 1000 stations adds roughly three new songs to its playlist each week. The *indies* get paid for

<sup>23</sup> This came at the end of a Congressional hearing on a television quiz show scandal, in which shows were exposed to be rigged and fixed in advance.



every one: \$1000 on average for an “add” at a Top 40 or rock station but as high as \$6000 or \$8000 under certain circumstances. That’s a minimum \$3 million worth of indie invoices sent out *each week*.” While it is easy to think that big record companies have a financial advantage in playing this game, Surowiecki (2004) argues that the big players already have the biggest names in show business, the biggest sales staff and the connections that go with it. Independent record promoters could thus enable small labels to get their artists on the radio, much the same way payola helped propel rock and roll in the 1950s. “Paying to play, then, creates a rough marketplace democracy: if you can come up with the cash, you get a shot. But that’s all. Labels can buy themselves exposure; they can’t buy themselves a hit. If people don’t want to hear a record, radio stations won’t keep playing it of their own accord”.<sup>24</sup>

So payola, even if disguised a bit, is still present.<sup>25</sup> Whether or not the current laws are optimal for the society is a good question for economists. Surowiecki (2004) says that pay for play is simply a signaling mechanism enabling record companies to signal which songs they think will be hits, thus reducing the radio station’s scouting efforts. Interesting developments are sure to come, with the growing consolidation of the radio business and power houses like Clear Channel Communications, which owns well over 1000 radio stations in the US, and with the advent of Internet radio stations and file sharing. Some stations, such as KROQ in Los Angeles and Clear Channel, also refuse to accept payment from independent record promoters. Perhaps the record companies will find new ways to promote their records, or perhaps big radio conglomerates will need to exert caution and stay away from the independent promoters business, to avoid payola charges or to enhance their reputation for independent judgment. It is also possible that large media conglomerates will use their position in multiple markets to extract even larger payments from record companies.

### 8.5. *Digital recordings in the Internet era*

With the advent of new technologies, such as streaming and downloading on the Internet, the Copyright Act no longer provided adequate protection for copyrighted works. In 1995, the Digital Performance Right in Sound Recordings Act was passed in an effort to strengthen copyright protection. The Act recognizes that digital transmissions of sound recordings are required to have an appropriate license. Interestingly, the license is administered by SoundExchange, a non-profit entity created by the RIAA (Recording Industry Association of America), and not by the performing rights organizations. Krasilovsky, Shemel and Gross (2003) note that between 1996 and March 2000, 80 million performances were licensed by SoundExchange. The revenues are split: 50 percent goes to the record company (NOT the publisher), 45 percent to the featured musicians

<sup>24</sup> Surowiecki (2004).

<sup>25</sup> Interestingly, these promotion payments are among the many costs that are deducted from record sales before bands receive royalties.

and vocalists, and 5 percent to an escrow fund for distribution to the non-featured musicians and vocalists. In addition, the Act establishes a statutory digital mechanical license rate, separate from the one for physical records.

Another Act was passed in 1998, the Digital Millennium Copyright Act. This Act was designed to implement two 1996 World Intellectual Property Organization Treaties dealing with copyrights in a digital environment. It provides restrictions on the use of technologies to copy and transmit copyrighted works, by making it illegal to circumvent measures put in place to guarantee the copyrights. The next section explicitly addresses new technologies and copyright issues.

## 9. File sharing and other new technologies

Throughout the 20th century, the rise and fall of various technologies have affected and shaped the way the world listens to music. Broadcasting, first via radio, jukeboxes, and movies, and then through television, cable television, and satellite television – including music channels like MTV and VH1 – and, very recently, via Internet webcasts, has allowed music to reach more and more listeners. Sound recordings also have evolved, with new formats – and along with them new playback machines – being introduced, and most often completely replacing the earlier generations. Recordings began with Edison's cylinders and Berliner's gramophone, then vinyl 33 1/3 rpm records, then 45 rpm singles, eventually followed by 8-track tapes, and then cassette tapes and Sony's Walkman. Records, as we know them today, in the form of laser compact disks (CDs), were introduced in the mid 1980s. By 1992, CD sales eclipsed cassette sales in the US.<sup>26</sup> Since 2000, CDs account for more than 90 percent of the market, whether one looks at total value of records sold or number of units shipped. In 2003, the CD market share was 95 percent.<sup>27</sup>

The supremacy of the compact disk is now threatened by a new format: the MP3, which stands for MPEG-1 Layer 3, a standardized digital file format that compresses audio to enable many songs to fit in a small amount of disk space.<sup>28</sup> Along with the spread of broadband Internet connections, file sharing and peer-to-peer (P2P) software, MP3 players have dramatically grown in popularity in the early 2000s. The actual number of song downloads seems to be impossible to pin down, but estimates suggest that more than one billion songs are downloaded *each week!* [See Oberholzer and Strumpf (2004) and Zentner (2006).] While many music lovers rejoice and engage in massive downloading and illegal file sharing, record companies and many music copyright holders deplore the practice, alleging that file sharing is responsible for declining album sales and lower profits. Industry executives were quick to put the blame on MP3 sharing. The

<sup>26</sup> In units per capita. See Liebowitz (2003a, Table 5).

<sup>27</sup> Source: RIAA, [www.riaa.com](http://www.riaa.com). Note that the figures do not include digital download sales.

<sup>28</sup> MPEG is the acronym for Moving Picture Experts Group.

Recording Industry Association of America (RIAA) successfully sued to shut down Napster in 2001, and as P2P networks provided an alternative platform for users, the RIAA is now suing thousands of individual users.<sup>29</sup>

### 9.1. Intellectual property issues

Economists have begun to look into the question of file sharing and CD sales. The situation can be considered from a normative perspective, questioning the legitimacy of the existence of copyright protection, especially since it can be seen as hindering the development of new technologies. Indeed, as Liebowitz (2004) notes: “It is common in the literature, particularly in the more popular press, to encounter the claim that copyright owners always cry wolf when a new technology appears to threaten the old, only later to discover that the new technology was nothing short of a bonanza. This claim implies that foolish copyright owners misunderstood the new technology and were fortunate enough to have been thwarted in their attempts to restrict the new technology.”<sup>30</sup>

How far does intellectual protection go? Are rights strong enough to encourage the optimal amount of innovation? The problem stems from the fact that musical compositions are non-rival goods, whose property rights, as laid out by Nordhaus (1969), generate a trade-off between under-provision of the non-rival good (with weak rights) on the one hand and monopoly distortions (when the property rights are strong) on the other.<sup>31</sup> The RIAA is clearly pushing for stronger rights, and is lobbying for greater governmental control over technology. Romer (2002) points out that “The relevant economic question is whether the net harm (if any) created by a shift along the Nordhaus trade-off justifies this kind of intervention.” He also warns that “giving an industry veto power over new technologies that threaten its current business model would set a very dangerous public-policy precedent”.

Boldrin and Levine (2002) argue against intellectual property protection. They present a model of competition where downstream licensing, in this case copyright protection, leads to the Pareto *worst* outcome, whereas a case without copyrights results in first-best. “‘Intellectual property’ has come to mean not only the right to own and sell ideas, but also the right to regulate their use. This creates a socially inefficient monopoly, and what is commonly called intellectual property might be better called ‘intellectual monopoly’.” Klein, Lerner and Murphy (2002) reject Boldrin and Levine’s model and reach the opposite conclusion: file sharing technologies reduce the value

<sup>29</sup> Napster was up and running in its original incarnation between June 1, 1999, and July 11, 2001.

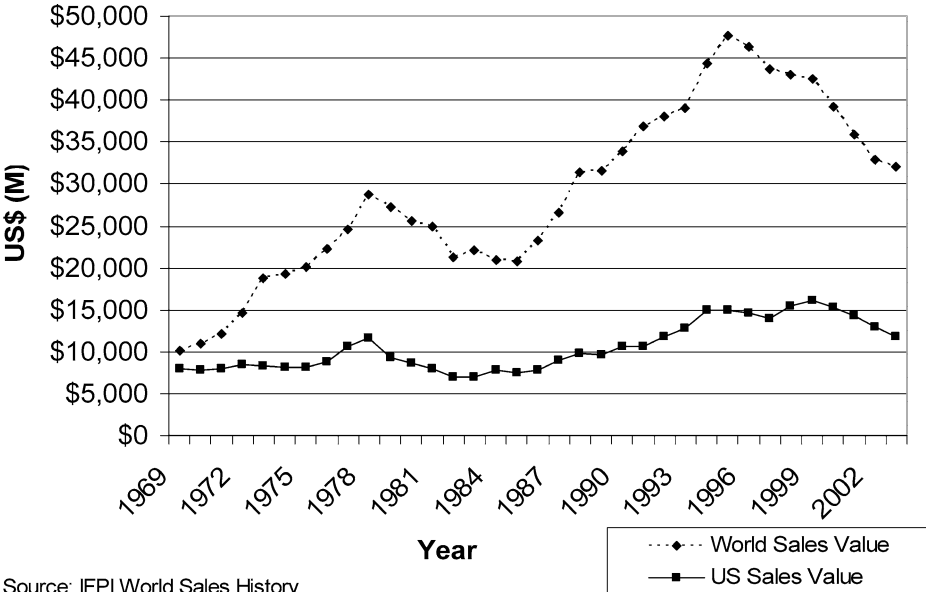
<sup>30</sup> Liebowitz (2004) goes on to argue that, for example, when the VCRs were introduced, television broadcasters had a legitimate concern because of the possibility that users would be able to time-shift television viewing and thus kill the possibility for broadcasters to sell advertisement, their principal source of revenue.

<sup>31</sup> Musical compositions are non-rival goods, since once their reproduction cost is paid, they can be simultaneously enjoyed by many. Efficiency would dictate a price (above reproduction costs) of zero, but then composers would be underpaid, and the production of music recordings would be too low. See Liebowitz (2003b).

of the copyright to its holder.<sup>32</sup> These models remain theoretical, without any support from empirical evidence.<sup>33</sup> The first question to ask here is whether or not there exists a causal relationship between file sharing and the slump in CD sales.

9.2. Does file sharing lower CD sales? Preliminary considerations

There is no unambiguous theoretical prediction regarding the effect of MP3 file sharing, or other illegal forms of music piracy, on CD sales. Various effects have been suggested that point in opposing directions. Furthermore, available evidence is, at best, mixed. Liebowitz (2005) and Peitz and Waelbroeck (2004b) present the most thorough reviews of the existing theoretical and empirical literatures. Before we review the results, it is informative to take a look at trends in record sales. Figure 8 shows the evolution of



Source: IFPI World Sales History  
 Notes: All values in millions of 2003 constant US dollars.  
 Turkey and China are excluded as they do not comply with IFPI standards and definitions.  
 Other audio formats (MiniDisc, DVD-A, SACD) included in totals from 1997 onwards.  
 Music video figures included in totals from 2001 onwards. Digital download sales excluded.

Figure 8. Total value of record sales, 1969–2003.

<sup>32</sup> Klein, Lerner and Murphy (2002) reject Boldrin and Levine’s model on the basis that it is based on “the innocuous assumption that the copyright holder’s demand is elastic”.

<sup>33</sup> Hui and Png (2002) provide empirical evidence on the movie industry by estimating the impact of economic incentives on the supply of motion pictures.

recording sales (of all formats) worldwide and in the United States from 1969 to 2003, in constant dollars. It is apparent that the value of sales has declined in recent years, after peaking in 1999 in the US and in 1995 worldwide.<sup>34</sup> Sales have dropped an average of 7 percent per year since 2000 in the US. The picture is similar for the global music industry. At least for the US, the downturn coincides with the launch of Napster and new portable MP3 players, such as Diamond's Rio in 1999. One should also note that CD copying became widely feasible on home computers in the late 1990s. Could these technologies be responsible for the drop in record sales?

The timing for the US is certainly suggestive, but it should be noted that there have been periods of sharp declines in sales before. In the late 1970s and early 1980s, sales plummeted, though not as sharply as in the last few years. Furthermore, the fact that the decline in sales began outside the US before it did in the US is suspicious, because Internet technology was more widespread in the US than elsewhere in the 1990s.

Before jumping to conclusions from the coincidence of these trends, one needs to also consider other factors that affect record sales. Liebowitz (2003a, 2005) lists the price of records, income, population, changes in taste, and prices of substitutes and complements as relevant factors. Liebowitz (2005) looks at these standard demand determinants and concludes, "They do not appear capable of explaining the decline in sound recordings that had occurred."

On a theoretical level, file sharing, or more broadly piracy, can have many potential effects on CD sales.<sup>35</sup> The main argument set forth by Napster and other "pro-file-sharers" is commonly called the *sampling* argument.<sup>36</sup> Sampling is thought to have a positive effect on CD sales by allowing potential customers to hear songs before they purchase them. Because of file sharing, customers would be better informed, making CDs a less risky purchase. Like advertising, sample could have the effect of increasing sales. However, Liebowitz (2005) questions the force of the sampling argument, pointing out that once someone has in his or her possession a song obtained for free, he or she might not go the extra step of actually paying to legally purchase the CD. He concludes that "the effect of sampling (more music-listening services at a constant CD price) is to lower the price of music-listening services. The net effect should be to lower the revenues generated by music-listening services. With a price per CD that is independent of the sampling effect, this implies that the quantity of CDs will fall due to sampling." Thus, sampling could be viewed as a supply shift as well as an information source, with opposing effects on sales.

Another effect, known as the *substitution* or *replacement* effect, clearly is expected to have a negative impact on sales: here, music downloaded simply replaces purchased CDs. Even if MP3s and CDs are not perfect substitutes since CDs come in a package

<sup>34</sup> One limitation of these data is that the dollar value is based on the suggested retail list price, not on the actual sales generated by the albums.

<sup>35</sup> This section focuses on file sharing. For more on piracy and the effect on CD sales, see Hui and Png (2003). Using international data from 1994 to 1998, they find that piracy reduces CD sales by 6.6 percent.

<sup>36</sup> See Gopal, Bhattacharjee and Sanders (2006) for a formal model of sampling.

with the CD jacket and perhaps lyrics and liner notes, and since the sound quality of MP3s can be inferior to CDs, we would nevertheless expect that if people can download a song for free, it will to a certain extent replace their purchases of music.

A third effect that [Liebowitz \(2005\)](#) cites is the *network* effect, but again, there is no clear prediction as to whether this would have a positive or a negative impact on CD sales. A network effect occurs when the value of a commodity varies with the number of people that are using it.

Another point to consider is that perhaps what has occurred is not just substitution of CDs for MP3 files, but a shift in leisure activities brought about by the new technologies. Internet and computers could have created a change in how people spend their time, possibly reducing the demand for pre-recorded music. [Peitz and Waelbroeck \(2004b\)](#) look at different surveys of time use and daily Internet activities, and conclude that “there is evidence that the increasing availability of broadband is changing the spare time activities of consumers in favor of online activities”.

### 9.3. *Empirical studies*

Perhaps the first empirical study of the effect of illegal file sharing on CD sales was produced by the RIAA during the Napster trial. SoundScan’s CEO, [Michael Fine \(2000\)](#) had been engaged by the plaintiffs to produce evidence on the question. His report is not very compelling. His main claim is that because sales declined more at stores near colleges and universities, and because college students are heavier downloaders than the rest of the population, then it must be that file sharing reduces CD sales. However, this analysis does not take into account the fact that the students might use the Internet to legally buy CDs online, thus also reducing the sales at local music stores.

[Liebowitz \(2005\)](#) surveys the existing empirical literature, classifying papers in terms of their unit of analysis. He distinguishes between countries, records, cities, households, and genres as possible units of analysis. [Liebowitz \(2005\)](#) notes that a methodology relying on genres would be interesting, but dismisses such studies because of lack of consistent and reliable data. He reviews the household methodology quickly, citing [Michel \(2004\)](#) as an example. Michel builds a model in which the consumer has the choice between buying a CD and copying music illegally. He then derives the market demand for CDs, and finds that the introduction of new file sharing technologies actually brings to the market people who were not previously buying any music. This would imply that CD sales should not decrease because of P2P networks. Michel uses household level data from the Consumer Expenditure Survey, taking computer ownership as a proxy for file sharing. This has obvious problems, since it measures neither Internet access nor file sharing behavior. Michel uses a difference-in-differences estimator to assess the impact of MP3 downloads on CD sales between 1998 (pre-treatment) and 2001 (post-treatment). He finds an insignificant effect, and is therefore unable to reject the hypothesis that “some file sharing (prior to 2002) was undertaken by consumers formerly not in the market for music”. One caveat that we need to mention here is that this result would hold if nothing else had changed between the two years studied. We can

however suspect that computer ownership and Internet use has greatly expanded in that time period, so we would not necessarily be capturing the behavior of the same type of households.

Liebowitz (2005) also criticizes Boorstin's (2004) study, which used cities as the unit of analysis, with Census data from 1998, 2000, and 2001. Boorstin regressed CD sales by city on the number of people with Internet access. Liebowitz argues that the regression is flawed because it includes dummy variables for the years 2000 and 2001, which are likely to pick up the effect of file sharing. He redoes Boorstin's analysis without the year dummies, and regressing on *per capita* CD sales (not total sales), and ends up finding that file sharing could explain a decline in CD sales of 12 percent in the US, when the actual total decline was of 15 percent. Liebowitz concludes that "These two values are so close that we can say that this evidence is consistent with a view that all of the decline is due to file sharing. This is a charming story, but it isn't clear how reliable these results are".

A study by Zentner (2006) uses data from a European consumer mail survey from October 2001. In an OLS regression of a dummy for buying CDs on a dummy indicating regular downloading, he finds an insignificant effect. However, as he points out, "simultaneity between tastes for music and peer-to-peer usage makes it difficult to isolate the causal effect of music downloads on music purchases". Consequently, Zentner instruments for regular downloading, using variables such as the speed of the individual's Internet connection and measures of Internet sophistication.<sup>37</sup> He then finds a negative and significant effect, indicating that music downloads reduce the probability of CD purchases by around 30 percent, which would explain a drop of 7.8 percent in the sales of music in the countries covered by the survey.

Peitz and Waelbroeck (2004a) use countries as a basis of analysis, with data from 16 countries representing 90 percent of world sales and from an IPSOS-REID survey, for 2000 and 2001. Taking first-differences, they run a regression of CD sales (expressed in units of CDs sold) on GDP, downloads (defined as the percentage of adults who downloaded MP3s at least once during the period) the percentage of households with broadband connection, and two variables indicating the sales of musical cassettes and the number of CD players per household. They find a significant and negative effect of downloads on music sales, reducing the sales by about 11 percent between 2000 and 2001. They then use survey data from the US for the period 2000–2002 to try to assess the partial effect of Internet piracy on CD sales. After making necessary assumptions, they conclude that "Internet piracy alone can only explain 22.5 percent of the CD decline in 2002 and is most likely not to be a significant factor in 2003 as the percentage of Internet users who download music is reported to have declined further after the series of legal actions undertaken by the RIAA in the summer of 2003". They note

<sup>37</sup> Zentner's (2006) measures of Internet sophistication include whether or not the individual publishes his own webpage, participates in online auctions, asks for technical support online, reads computer magazines, and how long he or she has used email and the Internet.



that this implies that a coefficient more than 4 times bigger would be needed to fully explain the drop of 8.9 percent in CD sales in the US in 2002. Peitz and Waelbroeck note, however, that their “results should be taken with caution since we consider a small number of countries in the econometric analysis. Also, the use of aggregated data and the particular choice of explanatory variable can be questioned.”

One study that stands out from the others in terms of sophistication and data is Oberholzer and Strumpf (2004), which reaches the controversial conclusion that file sharing does not have a significant impact on CD sales. They have access to unique data on actual downloads and sales, whereas the rest of the literature mostly relies on small-scale survey data or national and international aggregates.<sup>38</sup> Their data set contains 1.75 million file downloads, which represent about 0.01 percent of the world downloads for the seventeen-week period spanning from September 8, 2002, to December 31, 2002. They link this data set to album sales from Nielsen SoundScan from the second half of 2002, and also merge on information on the artists and track time taken from the website [www.allmusic.com](http://www.allmusic.com). They use the data set to regress observed record sales on album characteristics and the number of downloads for that album, using a fixed effects model to control for album-specific, time-invariant characteristics. To avoid endogeneity problems they instrument for the number of downloads, using shifters related to download costs, which they argue influence downloads but should have no effect on sales. The shifters they use as instruments are: album average and minimum track length, time length of albums in the same music category, and also the percentage of German students on vacation due to German school holidays. They maintain that the last variable is a valid shifter of the supply of files available for downloads because in their sample, one out of six US download is from Germany. When German children are on vacation, they would spend more time at their computer at home sharing files, thus shifting up the supply of MP3s. This choice of instruments has been criticized on various grounds. One criticism is that if, as the authors argue, file sharing leads to more CD sales through an advertising or (sampling effect), then cost shifters should enter the demand equation for CDs directly, rendering the identification strategy invalid. In any case, throughout their various specifications, they find that downloads have an insignificant effect on album sales. When comparing their estimates with the sharp drop in record sales, Oberholzer and Strumpf conclude that “At most, file sharing can explain a tiny fraction of this decline.”

Liebowitz (2005) criticizes this conclusion and the underlying methodology. He warns about a potential fallacy of composition that would arise because records are the unit of analysis. Just as the elasticity of demand at the industry level is expected to be lower than at the firm level, downloading could increase the sales of one particular album and reduce overall CD sales. This effect, as Liebowitz points out, can

<sup>38</sup> Their download data are logs from two OpenNap servers (centralized P2P network), which tells them which files users searched for and which files they downloaded, for seventeen weeks from September 8 to December 31, 2002.



seriously change the interpretation of Oberholzer and Strumpf's results: "A regression using downloads to explain sales would return a positive coefficient, assuming that all other simultaneity problems were overcome. After all, increases in downloads, by assumption, lead to an increase in the sales of the downloaded recordings in this example. But that does not mean that downloads increase overall record sales. A positive coefficient could be entirely consistent with record sales being severely harmed by downloads and thus couldn't answer the question about the overall impact of downloading."

Needless to say, the effect of file sharing on record sales remains a hotly contested issue. This is one area where we expect a good deal of research in the near future.

#### 9.4. Searching for a new business model

The jury might still be out on the effect of file sharing on CD sales, but one thing is certain: the record industry is suffering. And it is likely that the business model for distributing music will change dramatically in the near future. Zhang (2002) claims that the current music distribution system is inefficient and that peer-to-peer file sharing networks might be a solution: "P2P networks help to provide a better information environment for music listeners to experience the music works." He further predicts, "While smaller labels and unknown artists welcome the new technology, the big labels and stars suffer from the transition. The overall effect on social welfare is positive, but it is harmful to the music industry if only a small proportion of P2P users buy albums." Gayer and Shy (2006) present a model of an artist and her publisher, and show that the artist's revenues are greater under file sharing since the more revenue comes from live concerts, which get better publicity from the distribution of songs on P2P networks. However, in that model, music publishers lose from file sharing. In an interesting twist on Rosen's superstar model, Gopal, Bhattacharjee and Sanders (2006) predict that sharing technologies erode the superstar phenomenon widely prevalent in the music business. This implies that top artists actually lose from file sharing, but that less popular artists may gain from the extra exposure and lower distribution costs that the Internet has to offer. Michel (2003, Chapter 3) similarly predicts, "It appears that the artists and the consumers will reap most of the benefits of the new technologies". It is not surprising then to see how strongly the record companies react to the technological changes.

Legal issues are also prominent. Millions of people are infringing copyright laws, and the RIAA, as well as the Motion Picture Association of America (MPAA), its equivalent in the movie industry, are actively suing users and P2P software companies alike. In the midst of all this, some are proposing new copyright systems. A compulsory license system, much like the performing rights system right now, where radio stations acquire a blanket license to have the right to broadcast songs, and artists and publishers get compensated via a performing rights organizations (ASCAP, BMI, SESAC in the US), has been proposed, for example, in a recent book by Fisher (2004).<sup>39</sup>

<sup>39</sup> See also Liebowitz (2003b, Footnote 2 on p. 2), for a list of references. Liebowitz (2003b) discusses the pros and cons of the compulsory license, insisting on the cons and concluding that "only as a very last resort should we replace the current system with a compulsory license".

Peitz and Waelbroeck (2004b) describe technology developments known as Digital Rights Management, or DRM. DRM refers to technologies aimed at monitoring and blocking the use of copyrighted files. Some companies have already included such features in their software. It is unclear whether blocking technology will succeed in the race against file sharing.

### 9.5. *New business practices*

Napster is now back – as a legitimate service selling songs over the Internet. Other competitors include Apple's iTunes and RealNetworks' Rhapsody. The terms they offer vary. Some offer a sort of rental service where, for a monthly subscription fee, the user can download an unlimited quantity of songs onto his PC or portable device, but cannot burn CDs; once the subscription expires, the files can no longer be read. Others sell songs for a fee – 49¢ or 99¢ per song, or \$9.99 per album – and the tracks belong to the buyers forever. These services originally received a lukewarm reception: why pay for songs that are available for free on P2P networks? But interest at universities and colleges, fertile grounds for illegal file sharing (and lawsuits) with their broadband connections and student population, is growing. Some schools are starting new partnerships with music providers, in order to save bandwidth and curtail piracy. Napster has signed deals with eight colleges, including Penn State University, the University of Rochester, Cornell and George Washington University, through which the students receive free subscription to the regular Napster service (which is usually \$9.99 per month).<sup>40</sup> Berkeley and the University of Minnesota have signed agreements with RealNetworks.<sup>41</sup> Others, including Yale, Duke, Wake Forest and the University of Colorado at Boulder, have a similar deal with Cdigix (formerly Cflix), to receive not only music but also movies.<sup>42</sup>

## 10. Conclusion

Rather than summarize our lengthy survey, we conclude by suggesting some worthwhile questions for further study, which might stimulate research on the popular music industry. Below is our list of 11 areas that seem particularly worthy of further research:

- Why do contracts in the popular music industry take the form that they take? Are they efficient?

<sup>40</sup> See Young (2004a).

<sup>41</sup> See Young (2004b).

<sup>42</sup> The movie industry is also confronted with illegal file sharing, albeit perhaps on a smaller scale. Like the RIAA, the MPAA has taken legal action. On August 19, 2004, the 9th US Circuit Court of Appeals in Los Angeles ruled that P2P software developers were not infringing the copyright law by making products that people use to illegal download copyrighted material. The case was against the Grokster and Morpheus softwares. This is probably only an early opening round in the battles to come.

- Why have prices for popular music concerts grown so much faster than prices of other entertainment events since the late 1990s? Can more appropriate price indices for concerts – that take into account price discrimination, rationing, shifts in demand, and other factors – be constructed?
- What determines the amount of price differentiation within concerts, and why has price discrimination grown since the 1980s? Is there less regional variation in prices for the same concerts than one would expect in an efficient market? If so, why?
- How has increased concentration in promoters and media affected the popular music industry? Will continued technological change cause the industry to become more or less concentrated?
- There is a paucity of evidence on demand elasticities for concerts. As always, identifying demand and supply parameters requires some assumptions or exclusion restrictions. One potential approach is to use supply shocks, caused by factors like bad health (especially for older performers), to identify the elasticity of demand for concerts. Once a set of parameters is available, more elaborate rankings could be computed.
- The Internet lowers the cost of band promotion. How will the continued development of the Internet change the music industry? If bands rarely receive much income from record sales, will they seek other means for distributing their music? Will start-up bands have greater bargaining power with record companies because they can directly promote their music themselves on the Internet?
- How will future technological developments, which are hard to predict at present, affect the concert industry and the distribution of recorded music? Will the variety in popular music increase because of new distribution technologies?
- We lack systematic data on concert production costs over time. What are the trends in concert production costs? Can these costs account for the trend in prices?
- Tickets are beginning to be distributed in auctions. How do ticket auctions affect the average price and the size of the secondary market? What strategies do fans use when they have the option of purchasing tickets in an auction.
- Why is there a secondary market for tickets? Why do tickets appear to be underpriced for many concerts?
- Does the practice of legal payola (i.e., payments to radio stations via independent record promoters) affect the popularity of bands? Will payola become a common practice in new domains, such as webcasts, as technological change continues to shape the popular music industry?

## References

- Adler, M. (2006). "Stardom and talent". In: Ginsburg, V.A., Throsby, D. (Eds.), *Handbook of the Economics of Art and Culture*, vol. 1. Elsevier, Amsterdam. Chapter 25 in this volume.
- Ashenfelter, O. (1989). "How auctions work for wine and art". *Journal of Economic Perspectives* 3 (3), 23–36.

- Baumol, W.J., Bowen, W.G. (1966). *Performing Arts: The Economic Dilemma*. Twentieth Century Fund, New York.
- Becker, G. (1991). "A note on restaurant pricing and other examples of social influences on price". *Journal of Political Economy* 99 (5), 1109–1116.
- Besen, S.M., Kirby, S.N., Salop, S.C. (1992). "An economic analysis of copyright collectives". *Virginia Law Review* 78 (1), 383–411. Symposium on the Law and Economics of Intellectual Property.
- Boehlert, E. (2001). "Pay for play", salon.com (March 14, 2001). Available at <http://dir.salon.com/ent/feature/2001/03/14/payola/index.html>.
- Boldrin, M., Levine, D. (2002). "The case against intellectual property". *American Economic Review* 92 (2), 209–212. AEA Papers and Proceedings.
- Boorstin, E.S. (2004). "Music sales in the age of file sharing". Senior Thesis. Department of Economics, Princeton University. <http://www.cs.princeton.edu/~felten/boorstin-thesis.pdf>.
- Borghans, L., Groot, L. (1998). "Superstardom and monopolistic power: Why media stars earn more than their marginal contribution to welfare". *Journal of Institutional and Theoretical Economics* 154 (3), 546–571.
- Caves, R.E. (2000). *Creative Industries: Contracts between Art and Commerce*. Harvard University Press, Cambridge, MA.
- Coase, R.H. (1979). "Payola in radio and television broadcasting". *Journal of Law and Economics* 22 (2), 269–328.
- Courty, P. (2000). "An economic guide to ticket pricing in the entertainment industry". *Louvain Economic Review* 66 (1), 167–191.
- Courty, P. (2003). "Some economics of ticket resale". *Journal of Economic Perspectives* 17 (2), 85–97.
- De Vany, A.S. (2006). "The movies". In: Ginsburg, V.A., Throsby, D. (Eds.), *Handbook of the Economics of Art and Culture*, vol. 1. Elsevier, Amsterdam. Chapter 19 in this volume.
- De Vany, A.S., Walls, W.D. (2004). "Motion picture profit, the stable Paretian hypothesis, and the curse of the superstar". *Journal of Economic Dynamics and Control* 28, 1035–1057.
- Eliot, M. (1993). "Rockonomics: The Money behind the Music", second ed. Carol Publishing Group, New York.
- Fine, M., (2000). "SoundScan study on Napster use and loss of sales". Report of the CEO of SoundScan, engaged by the plaintiffs in the action, A&M Records, Inc. et al. v. Napster, Inc., <http://www.riaa.com/news/filings/pdf/napstert/fine.pdf>.
- Fisher, W.W.I. (2004). *Promises to Keep: Technology, Law, and the Future of Entertainment*. Stanford University Press, Stanford.
- Flynn, L.J. (2004). "The cellphone's next makeover: Affordable jukebox on the move". *The New York Times* (August 2, 2004).
- Gayer, A., Shy, O. (2006). "Publishers, artists, and copyright enforcement". *Information Economics and Policy*. In press.
- George-Warren, H., Romanowski, P., Pareles, J. (2001). *The Rolling Stone Encyclopedia of Rock & Roll*. Revised and Updated for the 21st Century. Fireside, New York.
- Gopal, R.D., Bhattacharjee, S., Sanders, G.L. (2006). "Do artists benefit from online music sharing". *Journal of Business* 79 (3), 1503–1533.
- Hamlen, W.A. Jr. (1991). "Superstardom in popular music: Empirical evidence". *Review of Economics and Statistics* 73 (4), 729–733.
- Hamlen, W.A. Jr. (1994). "Variety and superstardom in popular music". *Economic Inquiry* 32 (3), 395–406.
- Hui, K.-L., Png, I.P.L. (2002). "On the supply of creative work: Evidence from the movies". *American Economic Review* 92 (2), 217–220. AEA Papers and Proceedings.
- Hui, K.-L., Png, I.P.L. (2003). "Piracy and the legitimate demand for recorded music". *Contributions to Economic Analysis & Policy*, Berkeley Electronic Press 2 (1), Article 11.
- Johnson, W.R. (1992). "Commentary on 'An economic analysis of copyright collectives'". *Virginia Law Review* 78 (1), 417–419. Symposium on the Law and Economics of Intellectual Property.
- Kafka, P., Powers, K. (2003). "Celebrity 100, The Road to Riches". *Forbes.com* (July 7, 2003).
- Kahneman, D., Knetsch, J., Thaler, R. (1986). "Fairness as a constraint on profit seeking: Entitlements in the market". *American Economic Review* 76 (4), 728–741.

- Klein, B., Lerner, A.V., Murphy, K.M. (2002). "The economics of copyright 'fair use' in a networked world". *American Economic Review* 92 (2), 205–208. AEA Papers and Proceedings.
- Kleit, A.N. (2000). "ASCAP versus BMI (versus CBS): Modeling competition between and bundling by performance rights organizations". *Economic Inquiry* 38 (4), 579–590.
- Krasilovsky, M.W., Shemel, S., Gross, J.M. (2003). *This Business of Music; The Definitive Guide to the Music Industry*, ninth ed. Billboard Books, New York. J.M. Gross as contributor.
- Krueger, A.B. (2005). "The economics of real superstars: The market for concerts in the material world". *Journal of Labor Economics* 23 (1), 1–30.
- LaFranco, R. (2003). "The 2nd annual Rolling Stone money report. Rock's 50 richest". *Rolling Stone* 919, 57–61.
- Liebowitz, S.J. (2003a). "Will MP3 downloads annihilate the record industry? The evidence so far". In: Libecap, G. (Ed.), *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth*, vol. 15. JAI Press, Greenwich, CT.
- Liebowitz, S.J. (2003b). "Alternative copyright systems: The problems with a compulsory license". Unpublished paper. School of Management, University of Texas at Dallas. [www.utdallas.edu/~liebowit/intprop/complpff.pdf](http://www.utdallas.edu/~liebowit/intprop/complpff.pdf).
- Liebowitz, S.J. (2004). "The elusive symbiosis: The impact of radio on the record industry". *Review of Economic Research or Copyright Issues* 1 (1), 93–118.
- Liebowitz, S.J. (2005). "Pitfalls in measuring the impact of file-sharing on the sound recording market". *CESifo Economic Studies* 51 (2–3), 439–477.
- Marshall, A. (1947). *Principles of Economics*, eighth ed. MacMillan, New York.
- Michel, N.J. (2003). "A theoretical and empirical analysis of the impact of the digital age on the music industry". Doctoral Thesis. Department of Financial Economics, University of New Orleans.
- Michel, N.J. (2004). "The impact of the digital age on the music industry: A theoretical and empirical analysis". Unpublished paper. The Heritage Foundation, Washington, DC.
- Moulton, B. (1996). "Bias in the consumer price index: What is the evidence?" *Journal of Economic Perspectives* 10 (4), 159–177.
- National Music Publishers' Association, Inc. and The Harry Fox Agency, Inc., NMPA International Survey of Music Publishing Revenues, twelfth ed. National Music Publishers' Association, Inc. and The Harry Fox Agency, Inc. Available online at [www.nmpa.org](http://www.nmpa.org)
- Nordhaus, W.D. (1969). *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change*. MIT Press, Cambridge, MA.
- Oberholzer, F., Strumpf, K. (2004). "The effect of file sharing on record sales, an empirical analysis". Unpublished paper. Harvard Business School and Department of Economics, University of North Carolina, Chapel Hill. [www.unc.edu/~cigar/papers/FileSharing\\_March2004.pdf](http://www.unc.edu/~cigar/papers/FileSharing_March2004.pdf).
- Osbourne, S. (2002). "Interview: Women in rock". *Rolling Stone* 908, 54–56.
- Pareles, J. (2002). "David Bowie, 21st-century entrepreneur". *The New York Times* (June 9, 2002).
- Passman, D.S. (2000). *All You Need to Know about the Music Business*, fourth ed. Simon & Schuster, New York.
- Peitz, M., Waelbroeck, P. (2004a). "The effect of Internet piracy on CD sales: Cross-section evidence". CESifo Working Paper No. 1122.
- Peitz, M., Waelbroeck, P. (2004b). "An economist's guide to digital music". CESifo Working Paper No. 1333.
- Piketty, T., Saez, E. (2003). "Income inequality in the United States, 1913–1998". *Quarterly Journal of Economics* 118 (1), 1–39.
- Romer, P. (2002). "When should we use intellectual property rights?" *American Economic Review* 92 (2), 213–216. AEA Papers and Proceedings.
- Rosen, S. (1981). "The economics of superstars". *American Economic Review* 71 (5), 845–858.
- Rosen, S., Rosenfield, A.M. (1997). "Ticket pricing". *Journal of Law and Economics* 40 (2), 351–376.
- Slichter, J. (2004). *So You Wanna Be a Rock & Roll Star, how I Machine-Gunned a Roomful of Record Executives and Other True Tales from a Drummer's Life*. Broadway Books, New York.
- Surowiecki, J. (2004). "Paying to play". *The New Yorker* (2004-07-12).

- Tirole, J. (1988). *The Theory of Industrial Organization*. MIT Press, Cambridge, MA.
- Weinraub, B. (2002). "For the industry, less to celebrate at the Grammys". *The New York Times* (February 25, 2002): C1.
- Young, J.R. (2004a). "Napster and 6 colleges sign deals to provide online music to students". *The Chronicle of Higher Education* (July 30, 2004).
- Young, J.R. (2004b). "RealNetworks announces deals with Berkeley and U. of Minnesota to offer online music". *The Chronicle of Higher Education* (August 24, 2004).
- Zentner, A. (2006). "Measuring the effect of music downloads on music purchases". *Journal of Law and Economics* 49 (1), 63–90.
- Zhang, M.X. (2002). "Stardom, peer-to-peer and the socially optimal distribution of music". Unpublished paper. School of Management, MIT.

## THE ECONOMICS OF BOOKS

MARCEL CANOY\*

*European Commission*

JAN C. VAN OURS

*Tilburg University, CentER and CEPR*

FREDERICK VAN DER PLOEG

*European University Institute, Florence, CESifo and CEPR*

### Contents

Abstract	722
Keywords	722
1. Introduction	723
2. Cross-country statistics and trends	724
2.1. General overview	724
2.2. A closer look at book reading	726
2.3. Producing books and selling them	729
3. The characteristics of the book market	734
3.1. Perspective from industrial organisation	734
3.2. Differentiated products and uncertain prospects	735
3.3. Other characteristics of books	736
3.4. Authors, publishers, retailers and readers	737
3.4.1. Author versus publisher	737
3.4.2. Publisher versus retailer	739
3.4.3. Readers	740
3.5. Books and culture	741
3.6. Grounds for government intervention	742
4. A critical appraisal of fixed book price agreements	743
4.1. Welfare analysis	744
4.2. Retail price maintenance may increase non-price competition	747
4.3. Imperfect competition: Is the cross-subsidy argument valid?	747

\* The chapter was written while M. Canoy was still at the Central Planning Bureau, Netherlands Bureau for Economic Policy Analysis.

4.4. Fixed book price policies	749
5. Other public policies	751
5.1. Stimulating demand: Lower value-added tax	751
5.2. Stimulating supply: Prizes and grants for writers and subsidies for bookshops	751
5.3. Make reading cheap: Public libraries	752
5.4. Bringing the pieces of information together	753
6. Concluding remarks	756
Acknowledgements	759
References	759

## Abstract

This chapter analyses the tensions between books and book markets as expressions of culture and books as products in profit-making businesses and includes insights from the theory of industrial organisation. Governments intervene in the market for books through laws concerning prices of books, grants for authors and publishers, a lower value-added tax, public libraries and education in order to stimulate the diversity of books on offer, increase the density of retail outlets and promote reading. An overview of the different ways by which countries differ in terms of market structures and government policies is given. Particular attention is paid to retail price maintenance. Due to differences between European countries it is not a good idea to harmonise European book policies. Our analysis suggests that the book market seems quite able to invent solutions to specific problems of the book trade and that, apart from promoting reading, there is little need for government intervention.

## Keywords

books, publishers, authors, diversity, monopolistic competition, retail price maintenance, subsidies, libraries, Internet

*JEL classification:* Z11, D4, D6, L1, L4



## 1. Introduction

The arts have a special character. This combined with market failures may provide grounds for government intervention. The nature of intervention depends on the characteristics of the art and culture involved and on the potential failures of the market to provide an adequate and diverse enough supply of goods and services. Books are sufficiently different to warrant special attention. According to a 1964 UNESCO definition, a book is a “non-periodical, printed publication consisting of at least 49 pages, excluding cover pages”. Usually, three categories of books are distinguished: educational, scientific and general books. Educational books are intended as a means of learning for the institutionalised educational system up to higher vocational training, often based on government rules and regulations. Scientific books aim at users starting from higher vocational training. General books are all those that are not educational or scientific. Each category has its own characteristics and warrants different treatment. In this chapter we focus primarily on a sub-category of general books: cultural books. There are many other general books, such as cookbooks, travel guides or commercial fiction, which do not deserve the interest of ministers of culture. With cultural books there is always a tension between books as instruments of culture and books as products in a profit-making business as recognised by the Director-General of UNESCO at the World Book and Copyright Day, 23 April 2002:

Being one of the oldest means of communication and distribution, books not only have spiritual, educational and cultural implications, but also involve the legitimate industrial and economic aspects of the publishing trade. The association of these two factors – cultural impact and economic interests – results in a complex system of parameters that may seem incompatible. Developing the publishing and reading universe requires understanding of the existing internal relations between the different elements comprising the book chain: literary intervention, the respective functions of the publisher, printer, distributor and bookseller, and the reader.

Governments influence book markets through subsidies for libraries, authors and publishers, tax concessions on the sale of books, and laws concerning the pricing of books. This chapter first provides an overview of differences between countries in terms of market structures and reading behaviour. Section 3 discusses the special features of the book market and stresses the importance of principles such as *nobody knows, time flies, A-list/B-list* and *infinite variety* put forward by Caves (2000, 2006). It also discusses the grounds for government intervention. The market is quite capable of inventing solutions to specific problems (think of contracts for authors, literary agents, gate keeping by publishers, joint distribution by wholesalers cooperating on distribution, agreements concerning stocks between retailers and publishers, joint publicity, best-seller lists, reviews, etc.). Apart from stimulating reading, it is not clear what role there is for government intervention. The *pièce de résistance* as seen by most people in the cultural sector and among ministers of culture primarily in Europe is rightly or wrongly the fixed book price agreement. It may be viewed as a solution of the book

industry to specific problems, as laws are typically inspired by the book trade. Section 4 therefore provides a critical appraisal of retail price maintenance and fixed book price agreements. Section 5 discusses a spectrum of different government policy instruments to stimulate reading and to promote the diversity of publishing and distribution. It also gives some empirical details. Section 6 concludes.

## 2. Cross-country statistics and trends

### 2.1. General overview

To illustrate general trends in the book market and highlight differences between countries, this section presents stylised facts for 20 OECD countries. International comparative information is scattered and it is difficult to get a full overview for all countries. Although the focus is on cultural books, statistical information is only available for all books or for a different classification (e.g., the Universal Decimal Classification – UDC). Furthermore, due to differences in definitions and measurement, it is difficult to compare countries at a particular moment in time or differences within a country across time periods. Nevertheless, on the basis of the general overview presented in Table 1, some clear cross-country differences are visible in book reading, book production, and book services provided by public libraries.

There are large differences in book reading. About half of Portuguese adults never read a book. This is in sharp contrast with the 20 percent of readers in Belgium, Denmark, Italy and Norway. Reading is popular in Finland, Sweden and Switzerland where about 90 percent of adults read. Nevertheless, even in Sweden almost 30 percent failed to read a book during the past year. Although in most countries a majority of the adults read, there are also large numbers of people who never read a book. Among those who read, a substantial part read a book only every now and then.

UNESCO provides statistical information on the number of titles. These are non-periodic publications (first editions and re-editions) published in a particular country and made available to the public.<sup>1</sup> Table 1 presents cross-country differences in title production for ‘arts and culture’ and for all titles. At the low end of the distribution one finds the US, with 24 titles per 100,000 inhabitants of which 6 concern arts and culture. At the high end, Denmark produces 275 titles per 100,000 inhabitants, of which 80 concern arts and culture. Most titles per inhabitant are produced in the Scandinavian countries, in Switzerland and in the UK. Relatively few titles are produced in Italy, Japan, Greece and Australia.

For most of the countries for which information is available, the average annual number of books sold per inhabitant is about 5 to 6. The exceptions at the lower end are

<sup>1</sup> The term ‘title’ is used to describe a printed publication which forms a separate whole, in one or several volumes. Different language versions of the same title published in a particular country are considered as individual titles.

Table 1  
Economics of books – international comparison

	Reading		Titles		Copies sold	Revenues	Value added	Public libraries	
	Ever	Last year	Arts & cult.	Total				Books	Loans
Australia	78	–	–	37	6.3	35	–	–	–
Austria	–	43	27	100	–	80	0.06	1.2	2.0
Belgium	70	23	49	96	–	–	–	3.0	6.7
Canada	81	–	18	74	–	–	–	2.4	–
Denmark	76	55	80	275	5.3	90	0.14	5.5	13.7
Finland	89	–	56	225	5.4	115	0.08	7.2	19.2
France	–	40	34	66	6.9	45	0.08	1.5	1.5
Germany	83	40	23	98	–	75	0.10	1.4	4.0
Greece	–	36	18	39	–	20	–	0.9	0.2
Ireland	80	40	–	221	–	60	–	2.9	3.3
Italy	70	56	19	56	4.8	40	0.12	0.7	4.5
Japan	–	–	17	52	6.1	60	–	–	3.9
Netherlands	81	53	37	110	–	40	0.17	2.6	10.0
Norway	76	–	55	112	–	–	–	4.6	5.0
Portugal	49	15	37	82	2.6	60	0.10	0.9	0.3
Spain	–	39	58	148	4.7	40	0.14	1.0	0.6
Sweden	92	72	43	141	3.6	50	0.09	5.2	8.0
Switzerland	88	–	59	253	–	–	–	3.9	0.8
UK	82	63	55	188	4.7	50	0.18	2.1	7.8
USA	84	–	6	24	–	60	–	–	–

*Notes. Reading:* Percentage 'ever' reading books at home in population 16–65 years, 1994–1998; Belgium = Flanders, Norway = Bokmal. Source: International Adult Literacy and Life Skills Survey. Last year = for reasons other than work or study, percentage of population of 15 years and over, 2002. Source: Eurostat.

*Titles:* Annual book title production per 100,000 inhabitants, 1996–1999 (most recent year available), Belgium (1995) and Ireland (1999). Source: International Publisher's Association; all other countries = UNESCO Statistical Yearbooks.

*Copies sold:* Number of book copies sold per inhabitant, 2000–2002 (most recent year available). Sources: UNESCO, Statistical Yearbooks (Italy, 1996, Portugal, 1994, Spain, 1994), Stichting Spoorwerk betreffende het boek (Netherlands, 2000), Book Market Ltd. (UK, 2001), International Publishers' Association (all other countries).

*Revenues:* Publishers' revenues from book sales, Euro per inhabitant (US dollars for Australia, Denmark, Japan, Sweden, UK and USA), rounded numbers, 2000–2002 (most recent year available). Sources: International Publishers' Association (Australia, Italy, Japan), Association of American Publishers (US), European Commission (2004; all other countries).

*Value added:* Percentage of GDP contributed by the book publishing industry. Source: European Commission (2004). Spending on books in 1999 (percentage of GDP) was 0.30% in France, 0.42% in Germany, 0.36% in the UK, and 0.35% in the US. Source: Publishers Association.

*Public libraries:* Number of book volumes and number of loans to users, both per inhabitant, 1997–1999 (most recent year available). Source: UNESCO, Statistical Yearbooks.

Portugal and Sweden with an average of 2.6 and 3.6 books per inhabitant, while at the high end there is the France with 6.9. Publishers' revenues from sales also vary a lot from a low 20 Euro per inhabitant in Greece to a high 115 Euro in Finland. In most countries the revenue from book selling is about 40 to 60 Euro per inhabitant. In terms of GDP the book publishing industry is not very important. In absolute terms the largest industries are located in the US, Germany, the UK, France and Italy. In 2001, value added of the book publishing industry represented about 0.11 percent of GDP and employed some 140,000 in the EU-15. The industry is stable in terms of turnover and per capita sales. Books thus remain popular with readers and have not lost out to other media.<sup>2</sup>

Table 1 also provides information about cross-country differences in services provided by public libraries. The number of books available through public libraries is low in Greece, Italy, Portugal and Spain. It is much larger in Denmark, Finland and Sweden. The number of loans per inhabitant correlates highly with the number of books available. It ranges from less than one in Greece, Portugal, Spain and Switzerland to at least ten in Denmark, Finland and the Netherlands.

## 2.2. A closer look at book reading

People read books in a variety of ways. They read books at home, at work, while travelling, during holidays, etc. To find out how much time individuals read one has to clearly define time spent reading. This is especially important when different activities are performed simultaneously, say reading and travelling or reading and listening to music. Furthermore, reading may not be distributed evenly across the year. Some people may read on a regular basis while others read especially during holidays. Reading time measured as a regular activity in a typical week leads to a much lower estimate than reading time measured as the product of the number of books read and the average time spent reading a book. In time-use surveys it is therefore not easy to establish how much time individuals spend reading. While comparing results within the same time-use survey is easy, cross-comparisons between different surveys is difficult.

Table 2 presents information about the frequency of reading as measured in the International Adult Literacy and Life Skills Survey, which is based on similar surveys in a series of countries. Cross-country comparison is informative. As shown, differences in book-reading frequency are clear and large. Reading a book daily varies from about a quarter of all adult males in Australia, Canada, Ireland, Sweden, Switzerland, UK and US to a mere 5 percent for Portuguese male adults. In most countries 10 to 20 percent of adult males read daily. A common element is that females read substantially more often than males. The differences are smallest in Belgium (Flanders) and Portugal, and very large in Australia, Canada, Denmark and the Netherlands.

Cross-country differences in reading behaviour are present already at a young age. Table 2 also presents the reading behaviour of 15-year old children. Greece and Finland

<sup>2</sup> See European Commission (2004).

Table 2  
Frequency of reading books – adults and 15-year olds

	Adults aged 16–65 (daily reading)		15-year olds (read at least 1 hour per day)
	Males	Females	
Australia	24	41	16
Austria	–	–	12
Belgium	12	15	12
Canada	25	44	13
Denmark	20	36	14
Finland	16	30	22
France	–	–	14
Germany	20	29	13
Greece	–	–	29
Ireland	26	39	15
Italy	16	25	17
Japan	–	–	12
Netherlands	18	34	9
Norway	15	30	10
Portugal	5	9	16
Spain	–	–	11
Sweden	24	39	12
Switzerland	24	39	11
UK	25	37	12
US	25	39	12

*Notes.* Percentages of population. Sources: Adults – International Adult Literacy and Life Skills Survey, 1994–1998, 15-year olds – Programme for International Student Assessment, 2000.

are countries where many children like to read. Note that there are also differences in reading behaviour between adults and children.

The level of education is also an important determinant of reading habits but here no systematic cross-country information is available. We illustrate the situation for France and Italy.<sup>3</sup> In France 71 percent read at least one book during the past twelve months.<sup>4</sup> These percentages are 62, 78 and 92 percent for lower-, medium-, and higher-educated individuals, respectively. The same differences in reading behaviour are present for Italy, where on average 44 percent of the population read a book in the past year. The percentages are 32, 64 and 82 percent for lower-, medium- and higher-educated people, respectively.

There is not much cross-country information concerning trends in reading. We use information about the Netherlands as an illustrative example. As shown in Table 3, there is

<sup>3</sup> See [www.readingeurope.org](http://www.readingeurope.org).

<sup>4</sup> This number is slightly different from the number in Table 1, which refers to reading at home.

Table 3  
Reading books: males and females, Netherlands, 1975–2000

	1975	1980	1985	1990	1995	2000
<i>Males</i>						
Hours per week per capita	1.4	1.5	1.1	1.1	0.9	0.7
Reading books (%)	43	43	36	33	30	22
Hours per week per reader	3.3	3.5	3.1	3.3	3.0	3.2
<i>Females</i>						
Hours per week per capita	1.9	1.7	1.6	1.9	1.5	1.2
Reading books (%)	56	53	52	54	46	40
Hours per week per reader	3.4	3.2	3.1	3.5	3.3	3.0

Table 4  
Reading books: Finland, Germany, Norway, Sweden and UK, 1998–2002

	Finland	Germany	Norway	Sweden	UK
<i>Males</i>					
Hours per week per capita	1.1	0.7	0.8	1.1	0.6
Reading books (%)	13	9	12	13	8
Hours per week per reader	8.1	7.8	6.8	8.1	7.3
<i>Females</i>					
Hours per week per capita	1.6	0.9	1.3	1.5	0.8
Reading books (%)	20	13	20	23	13
Hours per week per reader	8.2	7.2	6.4	6.6	6.3

*Notes.* Reading connected with work and for an exam and reading as a joint activity, for instance with travelling, is not included. To some extent other reading (not reported here) may include reading books too, which may lead to underestimation of the time used for reading books. Source: Sociaal en Cultureel Planbureau for the Netherlands and also Time Use Surveys of Eurostat (2004).

a clear downward trend in book reading both for males and females. Furthermore, fewer people indicate that they read books, though the average time spent reading has hardly changed. Table 4 gives similar information for other countries. Females are more likely to read than males, but conditional on reading there is not much difference between males and females. The main differences in reading behaviour are related to the participation in reading and not to the time spent reading by readers. All readers irrespective of gender or country spend about 6.5–8 hours per week reading books.<sup>5</sup> In Europe, people spend most their time watching television. In the US trends suggest that Internet use is

<sup>5</sup> In the Netherlands reading participation is much higher but the number of hours read per week per reader is lower. This is likely to be a measurement issue, since the number of hours read per capita is similar to other countries.

increasing, mainly at the expense of watching television rather than reading [European Commission (2004)].

### 2.3. Producing books and selling them

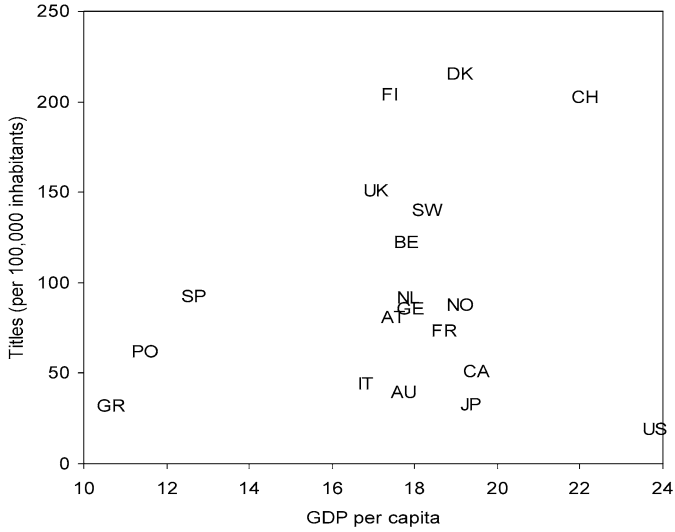
Table 5 shows that during the past decades production of book titles increased in all countries, but there are differences between countries. From the mid 1970s to the late 1990s there was virtually no increase in Norway and relatively mild increases in Austria, France, Greece and Sweden. In contrast, up to 1990 production more than doubled in Belgium, Finland, Spain and the UK. Although the number of titles produced increases each year in most countries, the number of publishers is stable. The average size of a publishing enterprise in the EU is small. Most publish only between 20 and 40 titles per year [European Commission (2004)].

Differences in the number of titles published may be related to economic prosperity, to the educational level of the population, or to population density. With rising incomes people may buy more books, thus increasing the supply of book titles. Figure 1(a) illus-

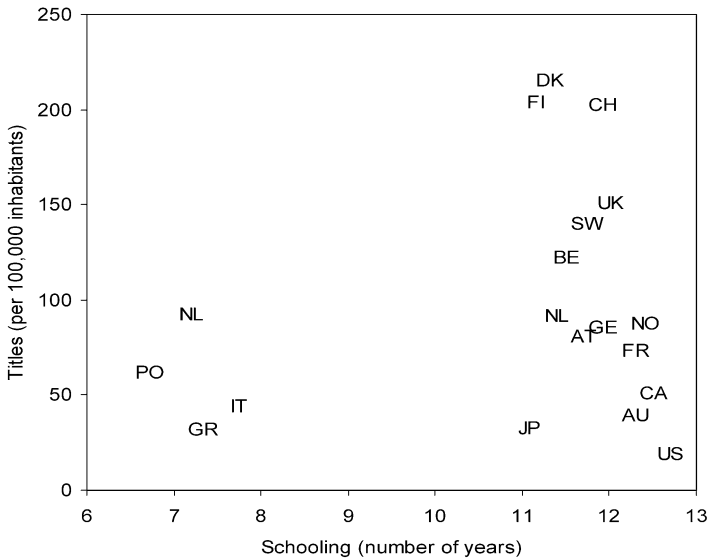
Table 5  
Annual book title production, 1975–1999 (per 100,000 inhabitants)

	1975	1980	1985	1990	1995	1999
Australia	40	66	67	39	38	37
Austria	74	94	112	81	102	100
Belgium	60	91	84	122	97	96
Canada	29	78	–	51	61	74
Denmark	140	181	187	216	238	275
Finland	97	136	182	204	264	225
France	54	60	69	74	60	66
Germany	–	–	–	86	91	98
Greece	29	42	47	32	40	39
Ireland	15	–	23	–	182	221
Italy	17	21	27	44	60	56
Japan	31	36	38	33	42	52
Netherlands	88	103	87	92	117	110
Norway	122	137	86	88	167	112
Portugal	63	62	104	62	78	82
Spain	66	76	90	92	122	148
Sweden	110	91	114	140	143	141
Switzerland	155	162	180	202	220	253
UK	63	85	93	151	174	188
US	39	33	21	19	23	24

Notes. The calendar year for which information is given is sometimes slightly different from the year indicated. The definition of ‘book title’ differs between countries and sometimes changes over time. Data for Belgium and Ireland (1995, 1999) are from the International Publishers’ Association, so they may not be fully comparable to the information for previous years. Source: UNESCO, Statistical Yearbooks.



(a) Book titles and GDP per capita, 1990



(b) Book titles and level of schooling, 1990

Figure 1. Book titles, GDP per capita, schooling and population density.

trates the relationship between titles per capita and GDP per capita. There seems to be a positive relationship, but it is not very strong. Countries with a low per-capita GDP, such as Greece, Portugal and Spain, have the lowest numbers of per-capita book titles, but



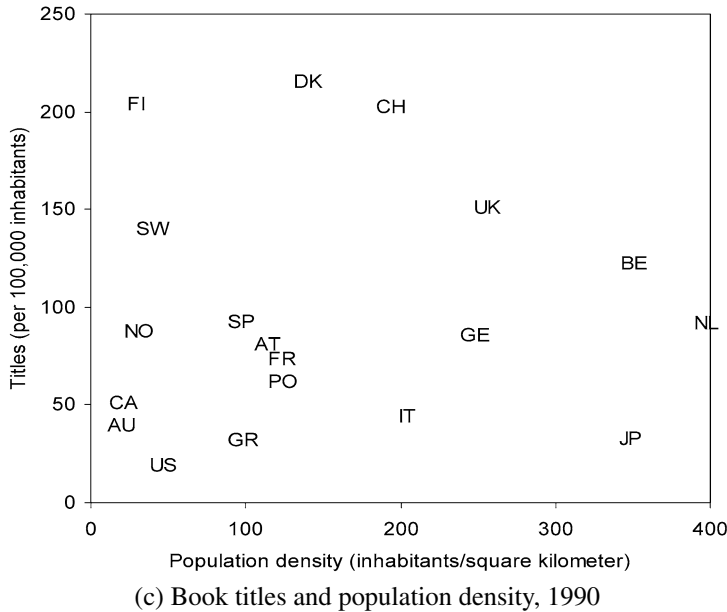


Figure 1. (continued)

among the countries with a higher per-capita income there is a lot of variation. The US, for example, has a lower per-capita book title production than some southern European countries with half the per-capita income of the US.

Figure 1(b) shows that there is also a positive relationship between the average educational level of a population and the number of titles produced. Again, the relationship is not very strong, and fuzzy for countries with a higher level of education. Figure 1(c) shows that there is no association between book titles and population density.

In terms of book title production distinguished by the Universal Decimal Classification the most important categories in each of the countries are social sciences, applied sciences and literature. Nevertheless, in terms of distribution of titles across UDC classes, there are big cross-country differences. In Austria, Canada, Portugal and Switzerland more than a quarter of all titles concern the social sciences. In France, Greece, Norway, Portugal and Spain more than 30 percent concerns literature. In other countries the distribution across the three main categories is more even. Table 6 shows that in the late 1990s in Canada, Finland and Germany only 20–25 percent of all titles concern arts and literature. For countries such as Greece, Norway and Portugal this percentage is 45–50. Table 6 also shows the evolution of the number of titles in arts and literature since 1975. The share of titles in arts and literature changes but no clear patterns emerge. For some countries, e.g., Belgium, France, Norway and Portugal, the share of literature increases, but in Sweden and the UK there is a decline.

Table 6  
Book titles: arts and literature as a percentage of total production

	1975	1999
Austria	22	26
Belgium	35	36
Canada	–	24
Denmark	30	29
Finland	38	22
France	32	42
Germany	–	26
Greece	51	46
Italy	36	35
Japan	–	39
Netherlands	33	43
Norway	32	49
Portugal	29	45
Spain	32	39
Sweden	36	30
Switzerland	23	23
UK	34	29
US	–	26

Notes. See Table 5.

European Commission (2004) presents a recent overview of the European book market. Total sales amount to 27 billion Euro in 2000. The biggest market is Germany with some 9.5 billion Euro (2000/2001). Both Germany and UK are strong exporters of books to countries that share their languages. Other large book markets are found in France, Spain and Italy. During the first two years of this decade, the UK book publishing industry has grown to be the largest in Europe. In contrast, there has been a decline in Germany.

Table 7 provides some information about publishers' revenues distributed by type of books and by distribution channel. About half the revenues in most countries come from general books. Table 8 shows that most of the sales are made through retail channels (trade), except in the US. In some countries there are strong retailers, but in others there are many independent bookshops. In France, the multimedia retailer Fnac accounts for around 15 percent of sales. In Italy Feltrinelli commands 25 percent of the retail market [European Commission (2004)]. However, in Germany, the largest bookseller, Thalia, has only 3 percent of the market and there are many small independent bookshops. The largest retailers in the UK in 1998 were Waterstones and W.H. Smith with 20 percent and 18 percent of the market [Latcovich and Smith (2001)]. The US book industry has limited opportunities for growth in a mature market and competition is focused on growth through market shares.<sup>6</sup> The US has seen consolidation among retail chains.

<sup>6</sup> See also Szenberg and Lee (1994), Greco (1999, 2000) and Clerides (2002).

Table 7  
Publishers' revenues – an international comparison by type of books (%)

	Year	Text books	Scientific books	General books	Children books	Total
Australia	2000	21	19	50	10	100
Denmark	2002	11	9	66	14	100
Finland	2000	22	8	56	14	100
France	2002	15	27	49	9	100*
Germany	2001	2	39	55	4	100
Ireland	2000	62	4	31	3	100
Italy	2000	23	6	68	3	100*
Spain	2001	21	24	46	9	100
Sweden	2001	40	2	48	10	100
UK	1999	15	28	48	9	100
US	2001	26	22	41	11	100

*Notes.* Text books = primary and secondary text books.

Scientific books = college, higher education, university, reference, dictionaries, encyclopaedias, professionals, STM excluding journals.

General books = general trade, consumer, fiction, non-fiction, and religious.

Source: International Publishers' Association.

\*Indicates that the columns do not add up to the total because information about the distribution is only available for part of the revenues.

Table 8  
Publishers revenues – an international comparison by distribution channel (%)

	Year	Trade	Book clubs	Direct	Total
Australia	2000	74	26	0	100
Denmark	2002	83	15	2	100
Finland	2000	59	16	25	100*
France	2002	69	18	13	100*
Italy	2000	71	4	25	100*
Sweden	2001	76	17	7	100*
UK	2001	78	10	12	100*
US	2001	41	9	50	100*

*Notes.* Trade = retail and wholesale bookstores, distributors and supermarkets.

Information about the distribution is only available for part of the revenues of France, Italy, Sweden, UK and US.

Source: International Publishers' Association.

\*Indicates that the columns do not add up to the total because information about the distribution is only available for part of the revenues.

Barnes and Noble commands 30 percent of the market and independent booksellers struggle to compete in the market [European Commission (2004)].

Table 8 shows that the share of book clubs is high in Australia (26 percent), about 15–20 percent in Denmark, Finland, France and Sweden, and low in Italy, UK and US. Although Internet sales have grown in importance, they are still rather small. In the UK around 17 percent of books go through Internet retailers, a percentage that is no longer thought to be growing very fast. For Germany estimates suggest between 4 and 5 percent of sales are made through Internet retailers, although recent growth has been much faster than in the UK. Some reports have estimated Internet sales in France and Italy at 1–1.5 percent. Spain has even lower Internet sales than France.

Internet is mainly used as a channel for books and so far not for digital products. For example, E-books are not sold much in the European market. In the US E-books are more important; over 7000 titles were published in 2003 while over 1.3 million E-books were sold. Concentration in the world wide online book market is high with 60 percent for Amazon.com [Latcovich and Smith (2001)].

### 3. The characteristics of the book market

Section 2 yields a colourful but opaque picture. The book market seems to flourish in one dimension (production) but not in the other (reading). The functioning of the market and its impact on both reading and production is not so clear. However, missing data frustrate an empirically sound judgement on the functioning of the market. There are also substantial differences between countries, both in descriptive data (number of books, beginning authors, Internet selling) as well as in government policies. This section follows up on Section 2 by zooming in on the nature of books and the book market.

#### 3.1. Perspective from industrial organisation

Many cultural goods share a number of properties; for example, in the words of Caves (2000) these properties can be listed as: *nobody knows* (uncertain demand), *time flies* (short period of profitability), *infinite variety* (horizontal differentiation) and *A-list/B-list* (vertical differentiation). To this Beck (2003) adds spontaneous purchases of about half books [Fishwick (1989)], non-convexities in production with large fixed costs and small marginal costs, and free entry for the book trade. Relevant questions are: in what mix do these properties come in the case of books; to which problems does that give rise; in what sense does the market solve the problems; and finally is there a role left for the government?

Economics can shed light on these questions. A book is a private good, since its consumption is rival and excludable. This implies that at first blush there is no fundamental market failure, so that government provision is not a serious option. Books can be borrowed by other people. However, in so far as that yields utility to the owners, there is no market failure. The market for books has a traditional supply chain: production,

wholesale, distribution and retail. In each part of the chain there is competition between private entrepreneurs. Government provision only occurs with libraries, but that does not exclude competition between private firms in the rest of the chain. There is substantial product differentiation in each part of the chain, which generates niche markets. Branding is important. Making a new product successful often requires substantial investment and innovation. This includes accepting that some products will never make it.

As a result of these features, most parts of the supply chain are characterised by a fairly large number of players. Consumers of books can easily switch from one product to the other. The book market knows relatively few consumer lock-ins, which helps the market to function properly. Transparency adds to that. Even though books are experience goods, author reputation, book reviews, book clubs and word-of-mouth create a fair amount of transparency. The book market is also dynamic: there is innovation, market shares fluctuate and there is entry and exit. An exception might be the European retail market, where government policy may have stifled innovation. From the perspective of industrial organisation, the characteristics of book markets are not that different from many other markets [Allen and Curwen (1991)]. This suggests that the book market should not be exempted from competition law. Still, in many countries retail price maintenance is tolerated. This is discussed further in Section 5.

### 3.2. *Differentiated products and uncertain prospects*

What are the consequences of the above features? The book market seems to be characterised by monopolistic competition. It has the following features:

- (a) the products sold are differentiated;
- (b) firms set the price of the goods;
- (c) the number of sellers is large and each firm disregards the effects of its price decisions on the actions of its competitors;
- (d) entry is unrestricted.

There then exists a trade-off between efficiency (exploiting scale economies by producing more of the same product type) and diversity.

This notion can be illustrated by looking, for example, at the market for cereals. Consumers have a love for variety, but variety can come at a cost. Each individual cereal variety becomes more expensive. In addition, the market becomes less transparent. Since firms do not take the potential downside of the variety decisions of other firms into account (the business stealing effect), there could be a market failure and optimal product diversity is not guaranteed. The book market is different in the sense that consumers do not engage in repeated purchases in the same way as they do for cereals. Book consumers rarely buy the same book twice. This changes the traditional trade-off in the sense that it greatly reduces possibilities for exploiting economies of scale. This is particularly true in the light of *nobody knows*. This does not mean that the book market can never have too much variety, but the argument then rests on lack of transparency and not on the more common economies of scale argument. Unlike in standard monopolistic

competition models, the book market is not characterised by repeated entry by publishers with each publisher filling a niche. It is books that occupy niches, not publishers. Publishers have a portfolio of authors and books that serve as a way of risk smoothing. Some books will make it while others will not, but publishers either have difficulties of forecasting the success or are happy to accept differences in success out of cultural motives. Additional complexities arise for two other reasons. First, the book market is characterised by the fact that a single product (a book) has a very short life cycle. This is not unlike products that depend on fashion. Bertarelli and Censolo (2000) formalise the idea that firms exercise monopoly power for a short period of time, but then have to cut prices under influence of entry and 'running-out-of-fashion'. Second, publishers may face a trade-off between risk smoothing and specialisation; that is, a publisher specialised in science fiction may have a competitive edge to non-specialised publishers, but may face the extra risk that science fiction lovers switch to video games.

A publisher thus has a quickly changing portfolio of books. Its strategy consists of deciding on the portfolio (trading off risks and specialisation) and on the prices of the portfolio. An 'industrial organisation' translation of portfolio may be economies of scope. In Ottaviano and Thisse (1999) multi-product firms in a monopolistic competitive market face the decision whether to engage in new product lines (exploiting economies of scope) or not (reducing cannibalisation). Depending on the parameters, in such a market there can be too much or too little variety [Dixit and Stiglitz (1997)]. This is akin to the decision by a publisher whether to employ a new author in the same field as his current portfolio. Here too there is a trade-off between economies of scope and cannibalisation. Similarly, the publisher has to face up to the decision whether to engage in a new field or not. It is not clear whether the accumulation of the complex decisions by publishers generates too much or too little variety. This trade-off combined with publishers' differences in 'love for culture', leads to a mix of publisher types. There are specialised publishers, small publishers and large publishers. This has been the case for many years in many countries.

### 3.3. Other characteristics of books

There exist several features that distinguish books from other products.<sup>7</sup> First, books are experience goods. One only learns the value after consumption. Second, books are characterised by high fixed and low marginal costs. Many other information goods markets share these two features. Third, some books are extremely successful, while most are unsuccessful. Success is hard to forecast (the *nobody knows* principle). In some circumstances this leads to 'winner takes all' economics as developed by Rosen (1981), but clearly not as extreme as in other information goods. Cowen (1998) reports that the number of copies sold in 1990 of the top-fifteen books accounts for less than one

<sup>7</sup> See also Appelman and van den Broek (2002), Throsby (2001), Cowen (1998), and Chapter 34 by van der Ploeg in this volume.

percent of total sales. It also means that booksellers and publishers – should they wish to do so – can cross-subsidise higher-risk books with profits on other books. Indeed Appelman and van den Broek (2002) present evidence that such cross-subsidies exist. These potentially welfare-enhancing cross-subsidies can be thwarted by non-branch shops (typically supermarkets) which might use books as a sales product (see Section 5). Fourth, the opportunity costs of consuming a book (i.e., time) typically outweigh the price of a book. This is one of the rare truly unique features of a book and contributes to a low price elasticity compared to other goods. In fact, the little evidence that is available suggests that the market for books other than best sellers is price-inelastic, probably because most purchasers have high incomes or buy them for study purposes.<sup>8</sup> Fifth, reading a book can be interpreted as a private investment in culture rather than consumption. Sixth, there is an (almost) free substitute for buying books, namely libraries. However, one can argue that the quality of the product is lower, which makes substitutability imperfect. Seventh, there is a public good nature associated with (the cultural value of) a book. Throsby (2001) mentions that a book (or, more generally, art) can be seen as possessing an option value ('I always have the option of buying a book'), an existence value ('I like the fact that there exist books') and a bequest value. Added to this are values associated with national identity, social cohesion, national prestige and the development of criticism and experiments [O'Hagan (1998)]. None of these values is (fully) reflected in the price, so that indeed the total value of books is higher than what has been paid. It turns out that these characteristics influence the way publishers, wholesalers, retailers and readers interact.

### 3.4. *Authors, publishers, retailers and readers*

The characteristics mentioned above create various problems. For most of these the market has found a solution. This subsection is organised around the solutions in various parts of the supply chain.

#### 3.4.1. *Author versus publisher*

The properties listed above make the book market relatively simple [compared to other cultural markets, see Caves (2000)]. A theatre production or movie is much more complex for a number of reasons. First there is the *motley crew* property. A play or movie involves a complex set of different professionals to interact. The success of the play or

<sup>8</sup> Van Ours (1990) and Appelman and van den Broek (2002) find a price elasticity of  $-0.8$  for the Netherlands, de Grauwe and Gielens (1993) find  $-0.6$  for Belgium, and Ecalle (1988) and Fishwick and Fitzsimons (1998) find  $-0.9$ . However, Hjorth-Andersen (2000) find  $-1.4$ , which suggests that demand for Danish books is highly elastic. Bittlingmayer (1992) finds an even higher price elasticity between  $-2$  and  $-3$  for the demand for individual books, which suggests more substitution possibilities between varieties of books than between books and other (cultural) goods. For example, the falling price of TV may have diverted demand away from books.

movie crucially depends on how these different professionals get along. Many parts of the chain have the possibility to break it and kill the project. This leads to a complex set of contracts and other institutions, largely unnecessary and therefore absent in the book industry. Second, the *nobody knows* and *time flies* principles are even more applicable to a play or movie than to a book. Third, the production costs of a play or movie are much higher than those of a book. As a consequence, movies and plays involve problematic financing, subsidies and complex contracts.<sup>9</sup>

Authors and publishers share the risk associated with the *nobody knows* and *time flies* principles. This implies that authors get a percentage of the sales (typically 10 percent) and a split of the gross profits (typically 58–42 percent) between author and publisher. Only with celebrity authors or authors with a strong reputation, there are advances (which can be substantial). While celebrity authors do reduce the risk of publishers somewhat, there are also serious large scale flops. Former US President Clinton's *Between Hope and History* had 70 percent of copies shipped to bookstores returned unsold [Caves (2000)].

Changing the terms of the contract either in favour of the author or the publisher can lead to misallocations. A higher fee for the publisher leads to a higher number of published books, since it becomes more lucrative to publish books and there still exists a reservoir of authors wanting to accept lower fees [Caves (2000, p. 57)]. However, there will be less commercial success per book on average and lower quality as good authors may spend their time on more profitable activities. This could be justified if the perception is that there is a lack of supply of books. There is no evidence of that however (the contrary is more likely). A higher percentage for the authors implies higher risk for the publisher, less books and less possibilities for beginning authors. Section 2.3 showed that there was substantial variation in production of books even corrected for underlying factors such as GDP per capita or education. It is thus not clear whether any given country is in the 'right' equilibrium.

In a simple world a contract, such as the one described above, would do and there is not much more to say on the matter. The world is not so simple though. One reason is that incentives differ between publishers and authors. Publishers want to make money. After a publisher has decided to accept a certain manuscript (even some potentially low-selling poetry) he still wants to make as much money as possible. Even the culture-loving publisher wants to use opportunities of money making to compensate for the failings; otherwise he simply does not survive. This can contrast with the interest of authors, who want to maximise sales and impact. The reason has to do with the payment schemes of the authors, who receive a certain percentage on books sold, but can supplement this with other sources of income. The potential of related side incomes (lectures, TV, film) has grown. With globalisation and the Internet the likelihood of superstar incomes for authors has become a real possibility. The increased importance of media

<sup>9</sup> See Chapter 19 by De Vany in this volume for an explanation of the functioning of the particularly complex market for movies.



makes it easier for them to leverage their reputation into sources of income. Indeed, Coser, Kadushin and Powell (1982) suggest that there is asymmetry of power with publishers having the right to refuse to publish even after accepting a manuscript, but there seem to be sufficient countervailing powers. Authors can change publishers, should they wish to do so. A large number of beginning authors find their way to the book market. In addition, sales of a novel increase the probability of future sales, a factor that influences an author more than his publisher. As a result of these differing incentives, authors may want to use agents. There is no marketplace for literary reputations of beginning authors. The chance that a publisher accepts a manuscript is extremely low; Caves (2000) mentions one in 15,000 for novels. Agents reduce the cost of publishers by filtering out good and bad manuscripts. The publisher can then use the reputation of a good agent as a proxy for quality. Agents can also perform useful commercial activities for authors.

The differences in incentives have created a love–hate relationship between authors and publishers. Coser, Kadushin and Powell (1982, pp. 224–225) report a number of funny incidents:

Thus, the nineteenth-century British prime minister and author Benjamin Disraeli, had received the unprecedented sum of ten thousand pounds from the house of Longman for his last work of fiction, *Endymion* (1880); when it did not sell as well as expected, he told an associate ‘My conscience will force me to disgorge’ and offered Longman a new contract that virtually amounted to returning three thousand pounds to the firm. Longman at once replied that it ‘could not think of availing [itself] of Beaconsfields [Disraeli] liberal and considerate suggestion’. But such instances of ‘*Après vous Gaston*’ are rare indeed.

And on the other side of the spectrum:

... a letter from the nineteenth-century writer of books on Japan, Lafcadio Hearn, to his New York publisher, Harper’s (which had resented something Hearn had done): ‘Please understand that your resentment has for me less than the value of a bottled fart, and your bank account less consequence than a wooden shithouse struck by lightning.’

The relationship between authors, publishers and readers can be better understood by considering the temporary monopolies created by intellectual property rights of authors [Plant (1934); Landes and Posner (1989)]. Copyright is a strong feature of book markets.

### 3.4.2. *Publisher versus retailer*

The properties of books also create problems in the relationship between retailers and publishers. Most notably, *nobody knows* and *time flies* create problems with stocks in retail outlets. If a book does not perform, the retailer wants to get rid of it as quickly as possible. Shelf space is scarce and new potentially successful books are looming. Market solutions to this problem include second hand sales shops, sales of remainders, pricing strategies and policies that aim at sharing risks between publishers and retailers.

In many countries book retailers have a right of returning books for full credit (variants of this principle exist). Again, and similar to the contracts between authors and publishers, history has determined a sort of equilibrium here. Too much leeway to retailers makes them 'lazy' and puts too much risks on publishers. Increasing the risk for retailers also creates problems, perhaps not all covered by market solutions. Retailers can further reduce risks by smart wholesaling agreements. There are distinct differences in market shares of wholesale firms in Europe. The largest wholesaler in the Netherlands has 70 percent of the market. In France, Finland and Denmark the wholesale market is also concentrated. In contrast, in Belgium the wholesale market is more scattered and efficiency of distribution is lacking [EIM (2001)]. In Anglo-Saxon countries, on the other hand, wholesale is less concentrated, but there are many integrated firms (publisher and wholesale). The fact that publishers are larger makes it worthwhile for them to vertically integrate into distribution. All in all, there is no reason to believe that the market will necessarily fail to solve the coordination problems needed to sort out the economies of scale.

It remains unclear in what sense the retailing sector is in the 'right' equilibrium in any individual country. The fact that there are such large differences between countries in retailing – and these differences are very unlikely to reflect just differences in preferences – suggests that the market solutions to the problems created by the properties of books are most vulnerable in the retail sector. Indeed, the number of independent book retailers reduces and they are replaced by chains [Epstein (2001)]. We come back to this later on.

There also exists a trade-off between exploiting economies of scale in retail and other policy goals. Examples are the reduction of transportation costs for consumers or equity 'universal service' type of arguments. The hermit on the Shetlands wants his bookshop. This trade-off is, at least in Europe, also apparent in banking, where efficiency requires shutting down local outlets, potentially frustrating elderly people who value the service, and supermarkets where there is a trade-off between environmental and planning issues and efficiency of large stores. Various trends tilt towards scale. First and most importantly is the possibility of the Internet. Even more than banking and supermarket products, books are easy to sell on the Internet. They are well-defined, easy to transport and personal contact with the seller is not (always) needed. In fact, interactive service and personal advice from Internet bookstores is often excellent. The storage, review and search possibilities are unlimited. The success of Amazon.com is no surprise.<sup>10</sup>

### 3.4.3. Readers

Since books are experience goods, consumers – when left to themselves – have a hard time deciding which book to buy. To facilitate the choices there are a number of institutions, mainly independent experts in various outlets. There are book reviews in

<sup>10</sup> See also Chapter 11 by Baumol in this volume on the role of the 'New Economy' in Arts and Culture, and Soon-Yong, Stahl and Whinston (1998), Creemers (1999), Yetkiner and Horvth (2000), Klein (2000) and Goolsbee and Chevalier (2002) for a discussion of book retailing on the Internet.

newspapers and the Internet, best-seller lists, and a fairly strong word-of-mouth culture. There is also some information transmitted through prizes and awards, and book clubs that pre-select titles. One of the best known clubs is the one associated with Oprah Winfrey's Book Club selections. The books chosen by Oprah often immediately hit the bestseller lists and are able to generate extra demand. For example, the first pick *Deep End of the Ocean* by Jacquelyn Mitchard was already on the market for some time and immediately became a top bestseller. All these institutions serve as intermediary between readers and booksellers. There is no reason to believe that the market for information is failing, with the possible exception for payola [Caves (2000)], a system in which the author (or his agent) 'bribes' a gatekeeper to influence his choices. The phenomenon is best known in Radio channels for Pop music.<sup>11</sup> Payola makes sense if airplay leverages future incomes. Since the government does not want the media to lose its independence, it has forbidden payola (although it still prevails). In the book market payola is less frequent and the argument runs through the best sellers list. An American consultant once bought so many copies of his own management book that he topped the best-sellers list for many weeks, leveraging his income in his regular consultant job. Another channel through which payola comes into play in the book market is through sticky prices. Chain bookstores can offer deals to book publishers to selectively display books in eye-catching positions. Caves (2000, p. 295) mentions the following example:

Barnes and Noble's 'Discover Great New Writers' program assures that a book appears face-out in every store for every two or three months and gets a review in a special brochure for \$1700 per title . . . These practices have entered into the controversy between publishers and the traditional independent booksellers over promotional allowances and other terms that disproportionately benefit chains and superstores.

Since payola runs against the vital role of objectivity (for culture) that gatekeepers perform, one is inclined to treat payola with scepticism.

### 3.5. *Books and culture*

The characteristics of Sections 3.1 and 3.2 and the market solutions of Section 3.3 lead to an assessment on how the book market performs in reaching cultural goals, which are:

- (i) a diversified portfolio of supply of books;
- (ii) books must be available for all, both in term of price and in terms of distance.

A diversified portfolio can imply several things: number of titles, number of genres, number of cultural titles or cultural genres. Similarly, availability (distance) can imply number of retailers, (cultural) stock of the retailers or variety in retailing. Finally, availability in price may refer to the prices of books or the possibility of reading [Appelman

<sup>11</sup> See Chapter 20 by Connolly and Krueger in this volume.

and van den Broek (2002)]. Obviously, difficulties in defining what cultural performance is, makes general assessments on performance somewhat heroic.

There are a number of observations that help assessing the cultural performance of the book market. First, books are rival and excludable. They share this with other cultural goods (CDs), but not with all of them (radio, TV, monuments are non-rival, buildings are non-excludable). The consequence of this (already mentioned earlier), is that the book market is more 'normal' than some other cultural markets, and hence requires less government interference. It is also important in the light of the discussion on technological trends, some of which transform non-excludable or non-rival goods into excludable or rival goods (think of pay-TV again). For books these discussions are unnecessary, but with the Internet one may expect a demand-driven growth in the sale of selected parts of handbooks and guidebooks. Second, books are reproductive cultural goods (unlike, say, the 'Nightwatch' by Rembrandt), implying that spreading books is easier than non-reproductive forms of art. Third, most fiction books are *not* luxury goods as are visits to the opera. One reason is that there is little social aspect to reading. Another reason lies in the presence of libraries, even though higher educated and richer people read more.

As a result, the market produces a large variety of books, with prices that are low enough (with libraries as a fallback as well) to make books available to everybody interested. O'Hagan (1998) and Cummings and Katz (1987) report that despite of this there are often additional policies towards arts directly aimed at equal access. We conclude that the market solutions, complemented with the presence of libraries, seem to be reasonably effective in reaching cultural goals, with the retailing sector as the most vulnerable part. This is a general assessment of the book market performance compared to other cultural markets. It does not say anything on the book market's cultural performance in any given country. From the empirical analysis in Sections 2.2 and 2.3 it follows that there are substantial differences in reading, production and retailing. These differences persist when correcting for factors such as GDP per capita and education, and are unlikely to be attributed only to differences in preferences. The assessment in this section points to existent market solutions to problems created by the characteristics of books. There are two possible explanations for these persistent differences. One is that some countries are in the wrong equilibrium. The fact that the market provides solutions does not imply that these are always used in a way that maximises social welfare. So it may well be that in some countries retailers are unsuccessful in dealing with the stock risks. This may lead to too few books, too few cultural books, too little reading or too many authors. Another explanation has to do with differences in public policy towards the book market.

### 3.6. *Grounds for government intervention*

The most important reason to intervene in the book market is to protect a dense network of well-stocked, high-quality bookshops and stimulate the publication of a large variety of books. Indeed, the number of high-quality bookshops is decreasing. This might be due to commercial motives if it does not pay to invest too much in variety in low-selling

books. Monopoly profits and cross-subsidies from profitable to less profitable books may allow bookshops to store a greater variety of books and publishers to take more risks. The current practice in many European countries of a fixed book price (FBP) in combination with a variety of subsidies handed out by literary funds is often motivated by these considerations. Critics argue that a FBP or subsidies for high-brow books may harm reading by the general public, since monopoly prices and cross-subsidies for less popular books are paid for by ordinary people reading popular books. Furthermore, subsidies for authors, translators, bookshops and publishers are paid for by ordinary people who may not be interested in more culturally valuable books or high-quality bookshops.

When considering policy instruments for reaching cultural objectives, there are at least two trade-offs. The first is between efficiency and density/distance. Increasing the scale of booksellers can enhance efficiency, but leads to longer travelling time for consumers. The second trade-off is between efficiency and cultural goals. Diversity of books in a bookstore may conflict with productive efficiency. The optimal choice of policy instruments is not the same for all countries.<sup>12</sup> It depends on culture-political preferences and on country-specific characteristics that determine the market outcome. Certain characteristics (e.g., a large 'language size') generate market outcomes where cultural objectives are more easily. In such countries the use of cultural policy instruments could be counterproductive. It is mainly for that reason that the United States, Australia and Canada do not have policies aimed at the book market, while they do have policies aimed at other cultural markets. This also implies that harmonising book policies in Europe is not a good idea.

Governments may wish to stimulate reading of worthwhile books, production of a diverse menu of titles and/or an extensive network of high-quality bookshops. None of the countries investigated by EIM (2001) specified the explicit targets for, say, the number of titles or bookshops. It thus is hard to evaluate the effectiveness and efficiency of the instruments used to attain cultural objectives.

#### **4. A critical appraisal of fixed book price agreements**

The fixed book price agreement (FBP) involves retail price maintenance, by which the publisher reserves the right to set the retail prices of books. Since the publisher also influences wholesale prices, he effectively sets gross margins for retail outlets. The cultural merits ascribed to such agreements have almost reached mythical proportions. No public debate in Europe on the cultural value of books is complete without a discussion of the FBP.<sup>13</sup>

<sup>12</sup> See Appelman and Canoy (2002).

<sup>13</sup> Tietzel (1995) and Rürup, Klopffleisch and Stumpp (1997) provide a thorough analysis with applications to Germany, Fishwick and Fitzsimons (1998) deal with the UK case. Hjorth-Andersen (2000) analyses the Danish situation, and Appelman and van den Broek (2002) discuss the Netherlands. See also Ornstein (1985), Uitermark (1986), Whyte (1994) and Ringstad (2004).

#### 4.1. Welfare analysis

We first compare a competitive equilibrium (CE) with a monopoly outcome for the book market. We assess the effects of a FBP which lasts forever, ignoring that in some countries the fixed price on a book expires after a few years and after which discounts are permitted and competition can play its role [van der Ploeg (2004)]. The full cost of reading a book  $Q$  includes the price of the book and the sales tax, but also the opportunity cost of the time needed to read a book ( $\phi W$  where  $W$  is the wage and  $\phi$  the hours spent reading books). Households use their time to work ( $1 - \phi B$ ) or read books ( $\phi B$ ), where  $B$  is the number of books bought. There is no utility of leisure, but utility of taking time to read books. Utility is quasi-linear in consumption of other goods  $C$ , say  $U(B) + C$  with  $U' > 0$ ,  $U'' < 0$ , and is maximised subject to the household budget constraint. Book demand follows from setting the marginal rate of substitution between books and other consumption goods equal to the ratio of the full cost of books to the price of other consumption goods. With quasi-linear preferences the marginal utility of private consumption equals one, so book demand follows from  $U'(B) = Q$  or:

$$b \cong -\varepsilon[(1 - \beta)(p + t) + \beta w] \\ \text{with } \varepsilon \equiv -Q/BU'' > 1/(1 - \beta) \text{ and } 0 < \beta \equiv \phi W/Q < 1, \quad (1)$$

where lower-case romans denote logarithmic deviations (e.g.,  $b \equiv dB/B$  except  $t \equiv dT/(1 + T)$ ),  $\varepsilon$  is the demand elasticity with respect to the full cost of reading a book, and  $\beta$  is the share of the opportunity cost of the time needed to read a book in the full cost. The demand elasticity,  $\varepsilon(1 - \beta)$ , is less than  $\varepsilon$ , since the price is only part of the full cost of reading a book. To have positive marginal revenue, assume  $\varepsilon(1 - \beta) > 1$ .

Publishers/booksellers maximise profits,  $PB - K(B) - F$  with  $K' > 0$  and  $K'' > 0$ , where  $K(B)$  denotes variable costs and  $F$  fixed costs. They set prices above marginal cost,  $P = K'\varepsilon(1 - \beta)/[\varepsilon(1 - \beta) - 1] \equiv P^{**} > K'$ . Prices are higher if the price elasticity is low, that is if few substitutes are available and the book price is only a small part of the full cost. Booksellers may find price discrimination, whereby high-income earners get charged a higher price for books, profitable. They may also use hardcovers and paperbacks as a tool of price discrimination, especially if the former are less price elastic than the latter. Equilibrium sales are given by

$$b = -\varepsilon[(1 - \beta)(t + m) + \beta w]/[1 + \varepsilon(1 - \beta)K''B/K'], \quad (2)$$

where  $m$  is the change in the mark-up. Book sales rise if the sales tax is cut or the opportunity cost of reading (the wage) falls. Figure 2 shows the effects of the FBP. In the competitive equilibrium,  $P^*$  and  $B^*$ , willingness to pay is the area under the demand curve  $x + y + z + v + w + a + b$ , which is more than consumers actually pay  $v + w + a + b$ . Subtracting total production costs  $v + w + F$ , we obtain total surplus, that is profits  $a + b - F$  plus consumer surplus  $x + y + z$ . Publisher/booksellers would have been prepared to deliver books below the equilibrium price. In the FBP equilibrium,  $P^{**}$  and  $B^{**}$ , willingness to pay equals  $x + y + a + v$ . Again subtracting production

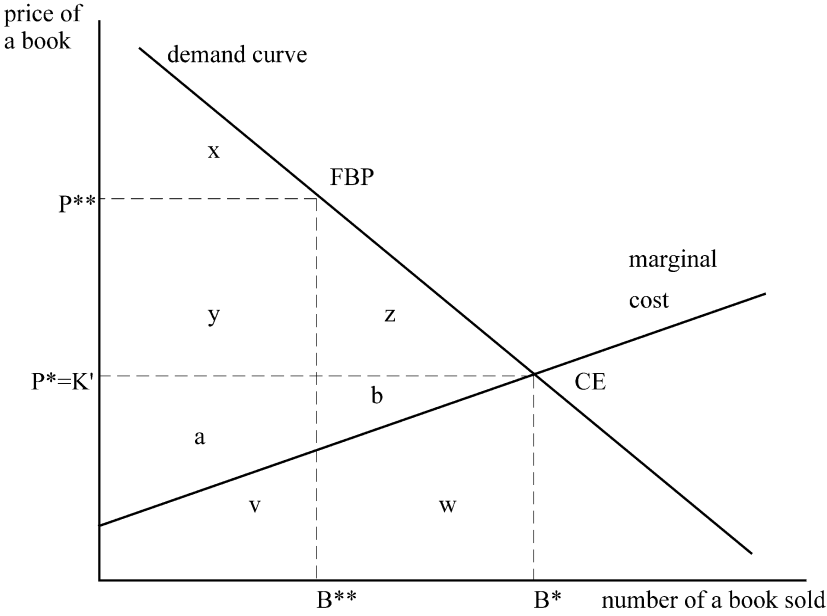


Figure 2. Welfare costs of the fixed book price agreement.

costs  $v + F$ , we obtain total surplus, that is profits  $y + a - F$  plus consumer surplus  $x$ . The loss in welfare resulting from the fixed book price is thus equal to the area of the familiar “triangles”  $z$  (the consumer loss if profits  $y$  are distributed to households) and  $b$  (the producer loss).

Under the FBP firms only publish/sell a particular book title if sales revenues  $y + a + v$  exceed costs  $v + F$ , that is if profits are positive or  $y + a > F$ . In a competitive equilibrium an individual title is published/sold if profits are positive, that is if  $a + b > F$ . If fixed costs  $F$  are very high, fewer titles are published and sold because it is less likely that sales revenue minus variable costs will be high enough to cover fixed costs.

Since monopoly profits are higher than profits in competitive equilibrium ( $y + a - F > b + a - F$ ), more titles are profitable and are published/sold under the FBP than in competitive equilibrium. It is possible to print and sell extra books at low and almost non-increasing marginal cost, so the producer loss  $b$  is likely to be small. Also, the price elasticity of the demand for books  $\epsilon$  is likely to be small as a large part of the full cost of reading is the opportunity cost of time. Hence, the monopoly mark-up is likely to be large and monopoly profits  $y$  are large. It thus seems likely that more titles are published under the FBP than under perfect competition. Some monopoly profits are necessary for marginal titles to recoup fixed costs, which is not feasible under perfect competition. However, if a particular title is published, fewer copies will be

sold at a higher price than in the competitive equilibrium. Even though the FBP leads to more variety in book titles published, prices will be higher and sales of each title lower.

The FBP also has dynamic costs. Of course, price competition between retail outlets becomes impossible but it also is more difficult to vary prices in response to local conditions. A store on a remote island may want to charge more for the same book than a store in the capital, but under the FBP it cannot do so. Also, it is more difficult to vary prices for different types of customers or for different seasons. Some customers need no service and low prices, while others prefer service at a higher price. Most important is that the FBP discourages the development of innovative distribution channels, since realised cost savings cannot be passed on to customers. Competition triggers independent small bookshops to be more attractive for the customer than big chains, supermarkets and the Internet. This seems indeed to be the case for the UK and the US. With the FBP, unconventional distribution channels (bookclubs, supermarkets, petrol stations, the Internet, etc.) have less of a chance. Against these costs there is the benefit that independent small bookshops may be able to recommend interesting books and order books from the publisher or distributor.

Tullock (1980) argues that publishers and booksellers lobby and spend time, energy and other resources to get and maintain the privileges of the FBP. Monopoly profits  $y + a - F$  are then not handed back to consumers. Publishers and booksellers go on with lobbying and rent seeking until a large part of these profits is dissipated, so monopoly profits  $y + a - F$  should be added to the sum of the consumer and producer surplus welfare loss “triangles” of the monopoly agreement. The total welfare loss is then  $y + a - F + z + b$ . If these profits are dissipated, the gain in the diversity of book titles will evaporate as well.

The FBP may thus lead to a bigger diversity of book titles (if rent seeking does not dissipate all profits) but to worse incentives to develop new distribution channels. It also leads to higher prices and less sales revenue for each title published. However, the experience of the UK after liberalisation has been that, while the number of bookstores has declined and new channels of distribution have opened up, book prices have gone up by more than the retail price index. Liberalisation of taxi fares often raises some taxi fares in order to reap the benefit from the uninformed tourist trade while others charge lower prices and enjoy a higher business volume targeted at the residents’ market.<sup>14</sup> Competition in the book market may fail for different reasons than in the taxicab business, since consumers can with the aid of the Internet, quickly be informed about prices. A different problem is that the public has difficulty in assessing the quality of particular books before they are read. This may lead to a “lemon” problem in which bad-quality books drive out good-quality books [Akerlof (1970)]. Book reviews and book clubs may prevent this.

<sup>14</sup> See Salop and Stiglitz (1977) and Sutton (2000).



#### 4.2. *Retail price maintenance may increase non-price competition*

Advocates of the FBP also argue that, even though price competition is eliminated, non-price competition may intensify. For example, a bigger sale margin stimulates booksellers to give better service to customers [Holahan (1979); Mathewson and Winter (1998); Deneckere, Marvel and Peck (1997)]. With a bigger profit margin, it pays to spend more effort on service in order to get extra customers. If the extra service (more attractive presentation in bookshops, better information to customers, more promotion, etc.) generates more sales than the fallback in sales due to higher monopoly prices, the FBP may be desirable. Otherwise, the market fails to deliver sufficient service, because bookshops have an incentive to operate as free-riders by offering discounts and expecting their customers to get their information and service elsewhere. Bookshops hardly refuse service or charge for information provided to people who in the end may not buy a book. Still, most customers rarely engage in such a strategy, as the costs of roaming around various bookshops seem high in relation to the possible discount one might obtain. Much of this service is already made available through publishers' advertisements or book reviews in newspapers and other media or on the Internet. In any case, it is questionable whether the demand for books really depends on service. Better service does not seem a good argument for supporting a FBP.

The book trade also argues that a bigger margin provides incentives for better-stocked bookshops. Booksellers may take over some of the inventory risks from publishers, so that more titles will be published. At the margin it is more profitable for retail outlets with relatively high costs to open up. This argument only works if customers want to purchase their books at particular high-cost bookshops. The gain in sales from these outlets may then offset the drop in sales resulting from higher monopoly prices. Although a dense network of bookshops may be desirable from a cultural point of view, this argument for the FBP is difficult to justify on grounds of market failure. Another popular argument is that higher margins encourage more retail outlets to put new book titles with uncertain sales prospects on their shelves. Given that there seems to be no problem for beginning authors to get their first book published, this is not a strong argument either. Marvel and McCafferty (1984) suggest that resale price maintenance may sustain a luxury image, but that seems more relevant for the markets for perfumes and jewellery than for books.

In sum, the above discussion suggests that there is not a clinching economic argument for the FBP. Even if there is a greater variety of book titles being published under the FBP, there may be more efficient instruments to achieve this. In any case, lowering of production costs due to technological progress will benefit the diversity of books being published.

#### 4.3. *Imperfect competition: Is the cross-subsidy argument valid?*

The novel *Endurance* by Ian McEwan is not a perfect substitute for *Il Nome della Rose* by Umberto Eco. They are different books, because the authors have different styles,

the themes of the two novels are different, and last but not least the original languages in which the books are written are different. Still, Umberto Eco's books are closer substitutes for the novels of Ian McEwan than, say, a cookbook or a travel book. On the other hand, Martin Amis may be a closer substitute than Umberto Eco for Ian McEwan. One must therefore leave the realms of homogeneous goods and adopt a framework of Chamberlinian monopolistic competition in which books are imperfect substitutes. Publishers/booksellers carve out a niche and make monopoly profits, which enable them to recoup fixed costs. It is thus profitable to publish books. In fact, an important argument of the lobby of booksellers/publishers rests on imperfect competition. They argue that the FBP allows for cross-subsidies from best-sellers to less popular books and leads to a more diverse supply of book titles and bookshops. In addition, the book lobby suggests that publishing and stocking a large selection of books enhances reputation, yields economies of scope and satisfies the idiosyncratic taste of individual publishers and booksellers even though these arguments do not seem very strong (also see Section 4.2).

The cross-subsidy argument seems at first blush irrelevant. In competitive markets with imperfect information about the success of a product, it is common to invest in many products and reap a success on only a few. Even without a fixed horse price agreement, horse owners purchase lots of yearlings, many of which subsequently are sold to the riding school or the butcher if they do not win races. Similarly, in a market without a FBP publishers invest in beginning authors, just like horse owners invest in yearlings. Indeed, the industry's rule of thumb formulated by Denis Diderot in 1767 suggests that one out of ten new editions is a profitable success, four cover costs, and five make losses [Beck (2003)]. There are few barriers to beginning authors in the book market even though publishing is a risky business with only a third of published books being profitable. The FBP then has all the welfare and political economy costs of a monopoly. This situation may arise if best-sellers are easily digestible, require little time to read and have high price elasticities of demand, while, say, poetry readings demand a lot of time and effort and have low price elasticities of demand. Indeed, anything worthwhile from a cultural point of view takes time and effort to appreciate and contributes to a low price elasticity of demand.

Non-fiction books (dictionaries, cookbooks, travel guides, textbooks, etc.) are likely to be close substitutes within each genre and will thus have high price elasticities. Fiction books (children books, mysteries, etc.) often have close substitutes (perhaps with the exception of *Harry Potter*), especially for the pocketbook versions of old titles, and thus high price elasticities. We do not expect large monopoly profits on such titles, and there is little room for cross subsidies to books with a special or unique character. Such books have low price elasticities and generate high monopoly profits. If this is the situation, the cross-subsidy argument is likely to be wrong. The problem with a FBP is that there is no guarantee that publishers/booksellers will use the monopoly profits to make sure that more esoteric titles will be published and stocked in the stores. Monopoly profits may well be directed towards unproductive managerial slack.

4.4. *Fixed book price policies*

In countries like Australia, Canada and the US a fixed book price is not an issue. However, in Europe books are subject to European competition law. Nevertheless, some countries allow for fixed book price policies for cultural reasons. Table 9 presents an overview. There are twelve countries with and eight countries without a fixed book price. The details of the book price regimes differ across countries [see [European Commission \(2004\)](#)]. Belgium, for example, has a self-regulation agreement restricting price competition for six months, after which pricing is unrestricted. In some countries the fixed book price policy may change, because they are tested by the competition authorities. These countries include Austria where the fixed book price is in force until 2005, Denmark where some changes to fixed pricing are under consideration by the Competition Authority, Greece where a recent court case has allowed newspaper stands to sell books at less than the fixed retail price, Italy where the law concerning the fixed book price was extended until 31 December 2004, and the Netherlands where the current law expired in January 2005 and a new law is currently presented to the Parliament. In the

Table 9  
Public policies on books

	VAT rates (2001)		Fixed book price policy
	Books	Standard	
Australia	10	–	No, not since 1972
Austria	10	20	Yes
Belgium	6	21	No
Canada	7	15	No
Denmark	25	25	Yes, adjusted since 2001
Finland	12	22	No, not since 1971
France	5.5	20.6	Yes
Germany	7	15	Yes
Greece	4	16	Yes, since 1997
Ireland	0	21	No, not since 1995
Italy	4	19	Yes
Japan	5	5	Yes
Netherlands	6	17.5	Yes
Norway	0	23	Yes
Portugal	5	17	Yes, since recent
Spain	4	16	Yes, since 1974
Sweden <sup>a</sup>	6	25	No, not since 1970
Switzerland	2.4	7.6	Yes
UK	0	17.5	No, not since 1995
USA <sup>b</sup>	1–7	1–7	No

<sup>a</sup>Before 2002, the VAT rate on books was 25%.

<sup>b</sup>Sales tax.

UK the fixed book price was abolished in the mid 1990s. Because of this smaller independent bookshops claim to have found it harder to remain in business despite offering high-quality service. Prices also went up in the UK, but this may have been due to a shift from pocket to hardcover books.

The UK, Sweden, Finland, Belgium and Ireland do not have a FBP, but the latter two countries are thinking of introducing it. France, Italy, Spain, Portugal, Greece, Austria, Luxembourg, Denmark and most recently Germany introduced FBP agreements in their laws. Most of these are based on the French law (the “*loi Lang*”). The Netherlands exempts collusive agreements within the book trade from the competition bill, but it has not anchored the FBP in the law as such. Some countries (e.g., Italy, Denmark, Spain and soon the Netherlands) exclude educational books from the FBP. Hence, the FBP is with the exception of the UK, Sweden and Finland popular throughout Europe [EIM (2001)]. Some European countries with FBP practise *collective* retail price maintenance, so that all associated publishers impose fixed retail prices on all associated retail outlets. Other countries (e.g., Germany) have *individual* retail price maintenance with prices of some titles free, which may have the advantage of some residual competition and less distortions.

The European Commission is not in favour of the FBP, which it sees as infringement of the ideal of a common market. However, it recently gave up its competition proceedings against the German book price fixing system (the ‘*Sammelrevers*’) because German publishers and booksellers gave sufficient evidence that their FBP did not hinder trade appreciably between member states and thus did not violate the European Union’s competition rules. Effectively, this guaranteed the freedom of direct cross-border selling of German books to final consumers in Germany, particularly, via the Internet. German publishers and booksellers thus will not hinder or put an embargo on direct cross-border Internet sales or on advertising of cheaper German books by foreign retailers. They have agreed not to violate the ‘*Sammelrevers*’, so that they will not cooperate with foreign retailers in order to circumvent the FBP. This prompted former European Commissioner Mario Monti to say:

On the basis of EU competition law the Commission has no problem with national book price fixing systems which do not appreciably affect trade between member states. By clearing the German price fixing system the Commission, in a perspective of subsidiarity, also takes account of the national interest in maintaining these systems which are aimed at preserving cultural and linguistic diversity in Europe.

Thus, there appear to be no European obstacles to national book price fixing agreements, provided that they do not hamper cross-border trade.

In sum, a FBP may induce higher prices and less sales of any book title that is published. It may also hinder innovation and distribution, but more titles will be published and there will be more bookshops with a diverse assortment of titles.<sup>15</sup> In any case,

<sup>15</sup> However, German data suggest that retail price maintenance does not facilitate above average focal pricing where prices are bunched around focal points [Beck (2004)].

many FBPs are of limited duration and characterised by sensible exceptions. The welfare costs are probably not very large, but may be reduced a little by reducing the term and coverage of the agreement. It may also be helpful to abolish certification and exclusive trade arrangements, scrap the fixed discount for recognised booksellers, and move to individual rather than vertical price agreements.<sup>16</sup> Since educational and scientific books typically have relatively low price elasticities and are more susceptible to monopoly abuse, it helps to exclude them from the FBP. As a dogma, the FBP diverts attention and energy away from making the book trade more innovative and customer-oriented. It may be more worthwhile to stimulate reading of a wide variety of books by investing in public libraries and education, subsidising authors to write books of high cultural value, translating the best books into other languages and promoting them abroad.

## 5. Other public policies

Apart from influencing competition in the market for books, governments actively interfere in the book market through prizes, grants, subsidies to bookshops, public libraries and special VAT-regimes for books. This way the government wishes to recognise that books are not products with just economic value, but also have cultural value.

### 5.1. *Stimulating demand: Lower value-added tax*

The general consumption of books can be increased by lowering the specific VAT-rate on books. This is a general instrument, which is not well suited to direct at special books of literary value. The lower VAT on books applies to cookbooks as well to poetry. This instrument is therefore mainly used to stimulate the purchasing and hopefully reading of books. Administrative costs are low, since no apparatus of literary experts has to be called upon. [Table 9](#) presents an overview of VAT policies on books. All countries of Europe, except Denmark, use this instrument. The UK and Ireland even abolished the VAT on books altogether. The European Commission misguidedly attempts to harmonise VAT-rates on books making it difficult for other member states to abolish VAT on books. The Commission fails to take account of the subsidiarity principle. Since book trade especially between the non-English speaking countries hardly distorts intra-European book trade, there is no danger of tax competition and no harm in countries pursuing their VAT-policies on books independently of each other.

### 5.2. *Stimulating supply: Prizes and grants for writers and subsidies for bookshops*

Governments and commercial sponsors do many things to encourage writers. There are many prestigious and less prestigious prizes for the best novelist, the best detective

<sup>16</sup> See also Appelman and van den Broek (2002).

writer, the best poet, the best translator, etc. All of these are meant to encourage quality. More important, they might guide the uninitiated reader to better books. Book clubs, best-seller lists and book programmes on television also help in this respect. They also probably increase sales. Literary funds help struggling authors to make a living if their project is deemed to be of literary interest. Since only best-seller authors can make a living on royalties and related incomes, others may need some help especially if their output has cultural value but is perhaps of less general interest. These policies are designed to stimulate quality rather than quantity. Sometimes subsidies for publishers of high-quality books may help as well (witness Sweden).

Many politicians attach cultural importance to a dense network of retail outlets. Section 2.3 already suggested that density seems to be falling in some countries, perhaps more in countries without a FBP; and concentration is increasing as well. From a cultural point of view this is bad news. Consumers have to travel longer and there is less variety of bookshops. If the main objective of cultural policies is to increase the density of high-quality outlets, subsidies for high-quality bookshops may be more effective than the FBP. If they act as cultural centres in less-populated areas, they may deserve public support.

Subsidising in order to maintain well-stocked bookshops would probably prove an administrative nightmare, which may explain why there is not much experience. Subsidising publishers to publish books of literary and cultural value would also seem to hinder the market mechanism and lead to adverse effects. In Sweden the government subsidises in this manner roughly a third of all fiction and a fifth of books for children. However, Swedish retailers do not stock all titles as the government, rather surprisingly, does not require subsidised books to be offered for sale.

### 5.3. *Make reading cheap: Public libraries*

Table 10 overviews the use of public libraries. Cross-country differences are substantial. With 0.4 service points per 10,000 inhabitants the density of public libraries is very low in France, Italy and Portugal. However, it is quite high with more than 3 service points in Switzerland and the UK. Figure 3(a) shows that in cross-country terms there is a clear positive correlation between public library service points and book titles. Of course, this correlation does not necessarily imply causality. More likely, the correlation has to do with preferences for books. In some countries preferences are low and therefore lead to few public library service points and titles, whereas in other countries the reverse is true. The highest number of library employees is found in Austria, in the Nordic countries and Switzerland. Less than 20 library employees per 10,000 inhabitants are present in Greece, Japan, Portugal and Spain. Two indicators of library use are available: the number of registered users, and the number of library visits. There is a large variation in the first (11 percent in Austria and 59 percent in the UK) but, with the exception of Greece, little variation in the second. Figure 3(b) shows that there is also a positive correlation between public library loans and book titles, though Switzerland is an outlier.

Table 10  
Public libraries, 1997/1999

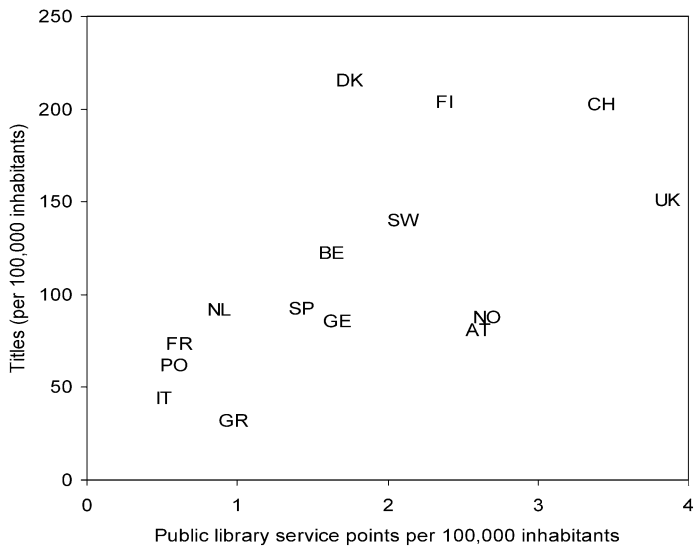
	Year	Service points	Library employees	Number of books	Registered users	Number of visits	User loans
Austria	1998	2.4	116	1.2	0.11	–	2.0
Belgium	1997	1.5	38	3.0	0.23	–	6.7
Canada	1999	1.2	42	2.4	–	–	6.6
Denmark	1999	1.6	93	5.5	–	–	13.7
Finland	1999	2.2	82	7.2	0.47	–	19.2
France	1997	0.4	22	1.5	–	4.5	1.5
Germany	1998	1.5	29	1.4	0.14	–	4.0
Greece	1997	0.8	17	0.9	–	0.2	0.2
Ireland	1998	0.9	35	2.9	0.23	–	3.3
Italy	1997	0.4	41	0.7	–	4.8	4.5
Japan	1999	0.3	15	1.5	0.28	–	3.9
Netherlands	1997	0.7	53	2.6	–	4.4	10.0
Norway	1997	2.5	42	4.6	–	4.4	5.0
Portugal	1999	0.4	7	0.9	0.41	–	0.3
Spain	1998	1.3	17	1.0	0.18	–	0.6
Sweden	1997	1.9	71	5.2	–	4.7	8.0
Switzerland	1997	3.2	74	3.9	–	4.3	0.8
UK	1999	3.7	45	2.1	0.59	–	7.8

*Notes.* Most recent year. Service points and library employees are per 10,000 inhabitants; book volumes, registered users, number of visits, loans to users are per inhabitant. Source: UNESCO, Statistical Yearbook 1999.

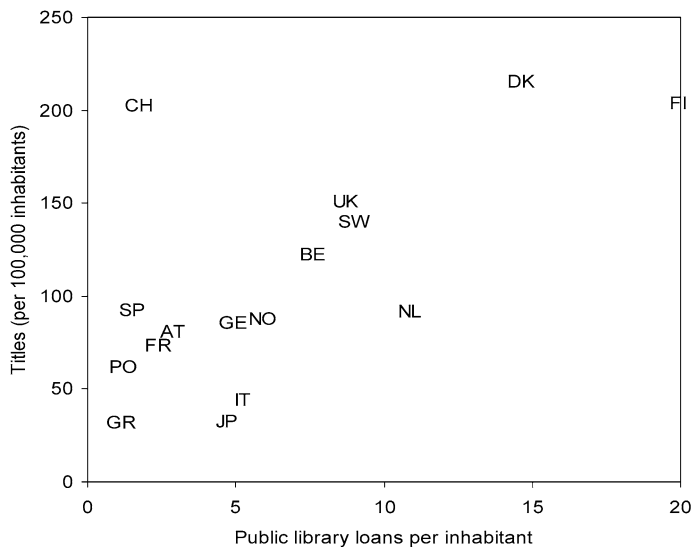
#### 5.4. Bringing the pieces of information together

From the stylised facts presented some conclusions may be drawn. People read fewer books and there are cross-country differences in reading. Some differences are already present at a young age. Females read more than males and higher educated people read substantially more than less educated individuals. Title production is increasing over time. Per-capita title production varies between countries, which suggests that some countries have more best-sellers than others. There are also cross-country differences in VAT and FBP policy. Furthermore, there are important differences in the use of public libraries. In Scandinavian countries, the Netherlands and the UK the use of public libraries is large. In Southern Europe the use is quite low.

To investigate the potential determinants of title production (see Table 5) we investigate to what extent per-capita GDP (taken from the Groningen Growth and Development Dataset), the average schooling level [taken from Barro and Lee (2000)] and fixed book price policies (Table 9) are relevant. We use information for 1975, 1980, 1985, 1990, 1995 and 1999, and allow for country-specific effects to account for possible cultural and taste differences and for time effects to account for changes in production



(a) Book titles and public library service paints, 1990



(b) Book titles and public library loans, 1990

Figure 3. Book titles and public libraries.



costs. We thus estimate the following equation:

$$\log(\text{TITLES}_{it}) = \alpha_i + \alpha_t + \beta_1 \log(\text{GDP}_{it}) + \beta_2 \log(\text{SCHOOLING}_{it}) + \beta_3 \text{FBP}_{it} + \varepsilon_{it}, \tag{3}$$

where TITLES is the number of titles per 100,000 inhabitants, GDP is real GDP per capita, SCHOOLING refers to average years of schooling of the population of age 25 and more, and FBP is a dummy equal to 1 if a country has a fixed book price regime and zero otherwise. Furthermore,  $i$  is a country index, and  $t$  a time index. The  $\alpha_i$  represent country-specific effects, the  $\alpha_t$  are time effects, and  $\varepsilon_{it}$  is an i.i.d. error term.

The first column of Table 11(a) shows the parameter estimates if the fixed effects are excluded. There is a significant positive effect of GDP per capita with an elasticity of 0.86, but neither the schooling level or the fixed book price regime have significant effects. If we introduce country-specific (random) effects, the parameter estimates hardly change. If we also introduce time effects, GDP per capita is no longer significant. Obviously, there is a high correlation between time and country-specific developments in GDP per capita (growth rates are correlated). Because of this, GDP per capita may have a positive effect on title production, but we cannot distinguish this effect from other correlated time effects representing production costs or taste changes. Table 11(a) also

Table 11  
Determinants of book titles and library loans

(a) Titles	(1)	(2)	(3)
GDP	0.86 (2.4)*	0.84 (2.9)*	-0.11 (0.3)
Schooling	-0.23 (0.6)	0.09 (0.2)	-0.02 (0.0)
FBP	0.02 (0.2)	-0.18 (0.9)	-0.20 (1.0)
Country effects	no	RE	RE
Time effects	no	no	FE
$R^2$	0.06	0.35	0.42
$\chi^2$	-	1.6	2.4
(b) Library loans	(1)	(2)	(3)
Schooling	3.55 (5.6)*	1.43 (5.8)*	0.92 (1.9)
Service points	0.41 (2.6)*	0.43 (4.1)*	0.41 (3.7)*
Country effects	no	RE	RE
Time effects	no	no	FE
$R^2$	0.45	0.39	0.38
$\chi^2$	-	4.0	1.5

Notes. Estimation period 1975–1999; titles: 20 countries, 109 observations; library loans: 18 countries, 90 observations;  $t$ -values in parentheses; all variables (except for FBP) are specified as natural logarithms; RE = random effects, FE = fixed effects;  $R^2$  = (within) correlation coefficient;  $\chi^2$  = test-statistic comparing random country effects and fixed country effects.

\*Indicates significance at the 5% level.

shows that the number of titles produced is unrelated to the average level of education and to fixed book price policy.

We also study the potential determinants of the use of public libraries, estimating the following equation for the same countries and years:

$$\log(\text{LOANS}_{it}) = \gamma_i + \gamma_t + \beta_4 \log(\text{SCHOOLING}_{it}) + \beta_5 \log(\text{SP}_{it}) + v_{it}, \quad (4)$$

where LOANS is the number of public library loans per inhabitant, SP is the number of service points per 100,000 inhabitants and  $v_{it}$  is an i.i.d. error term. Estimates are presented in Table 11(b). Average schooling in a country has a positive effect on the number of loans. The effect becomes smaller if country-specific (random) effects are introduced. This suggests that there could be correlation between schooling and library loans, caused by a joint preference for both. The effect becomes even smaller if time fixed effects are introduced, which suggest that over time there is a non-causal correlation between schooling and loans. Nevertheless, even after country-specific and time fixed effects are introduced, schooling has a positive effect on loans. This is not surprising, since (as discussed above) there is a positive correlation between education and reading. Table 11(b) also shows that there is a positive effect of library service points on loans.

Table 12 presents a similar analysis based on information for seven countries over the period 1990–1999 (production numbers are from the International Publishers Association). The estimated equation is similar to the previous one, where country-specific time trends  $\delta_i$  account for possible changes in schooling variable and tastes:

$$\log(\text{TITLES}_{it}) = \alpha_i + \alpha_t + \delta_i t + \beta_1 \log(\text{GDP}_{it}) + \beta_3 \text{FBP}_{it} + \varepsilon_{it}. \quad (5)$$

Now we find that GDP per capita has a positive effect, which is however reduced when time effects and country-specific time trends are introduced. The difference in results between Tables 11(a) and 12 may arise from GDP per capita being more correlated across countries over a long than over a short period. The first column of Table 12 shows that countries with a FBP price regime produce 25 percent more titles. However, this effect is estimated on cross-country differences. If we introduce country-specific fixed effects, the coefficient becomes significantly negative. This negative effect of the FBP on titles is driven completely by the abolishment of the FBP in 1995 and the subsequent increase in titles in the UK.

## 6. Concluding remarks

The book market ensures reasonable cultural performance with little government intervention, especially in large language areas. Yet there are differences between countries in reading, retail outlets, wholesale and production. Due to lack of data and research it is not easy to explain these differences. They may be due to differences in preferences, logistics, population density or public policies or due to being stuck in the wrong equilibrium. One important trend is that people seem to read fewer books over time. Perhaps

Table 12  
Determinant of book titles

Titles	(1)	(2)	(3)	(4)
GDP	0.87 (3.5)*	1.24 (7.6)*	0.50 (2.1)*	0.93 (1.7)
FBP	0.25 (2.9)*	-0.11 (2.2)*	-0.11 (2.6)*	-0.10 (2.3)*
Country effects	no	RE	RE	RE
Year effects	no	no	FE	no
Country specific trends	no	no	no	yes
$R^2$	0.16	0.59	0.75	0.73
$\chi^2$	-	2.4	0.8	23.4

*Notes.* Estimation period 1990–1999; 7 countries, 70 observations;  $t$ -values in parentheses; all variables (except for FBP) are specified as natural logarithms; RE = random effects, FE = fixed effects;  $R^2$  = (within) correlation coefficient;  $\chi^2$  = test-statistic comparing random country effects and fixed country effects.

\*Indicates significance at 5% level.

they are reading on the Internet or spending time on other cultural leisure activities. Here are some important areas for further research: investigate the relationship between production of titles, books sold and prices; use survey data to study the effects of personal characteristics of readers on market outcomes; analyse empirically differences between the book market and other cultural markets; and use industrial organisation to understand pricing and stocking behaviour of publishers and retailers.

The book industry is characterised by relatively few market failures and these can be relatively easily corrected with market instruments. The book industry can fend well for itself, in contrast to opera, movie or theatre, characterised by high production costs, high risk and complex interactions between a large number of different professionals. Even though there are obvious returns to scale, production costs are low. Thresholds for beginning authors, publishers and retailers are small, contracts are relatively simple and fairly uniform. The market is quite capable of inventing solutions to specific problems and public policies are not always called for, except perhaps to stimulate reading.

Nevertheless, there is a strong lobby for government intervention. Prizes and grants for authors, translators, publishers, bookshops, special VAT-regimes for books, stimulating reading through public libraries, and the fixed book price (FBP) are possible policy instruments. The standard case against the FBP is that book prices are higher and sales lower than under perfect competition. This hurts the interests of buyers, particularly those with lower incomes since prices will be higher. One possible argument in favour is that the FBP may induce more and better-stocked bookshops and lead to publication of more marginal book titles. The cross-subsidy argument of the lobby in favour of the FBP is not convincing, however. First, even without the FBP, the market cross-subsidises beginning authors and other risky projects in the hope of a possible best-seller. Second, even if this policy “works”, there is no accounting for what is done with the cross subsidies and no democratic checks. Third, there is no guarantee that profits on best-sellers will be used to cross-subsidise less popular books. In fact, pub-

lishers and booksellers have an incentive not to do this. Fourth, if less popular books are less price elastic than popular books (perhaps as they take more time to read), monopoly profits on less popular books are higher and the cross-subsidy argument does not work. Fifth, even if cross-subsidisation does occur, one should evaluate whether its cultural gains outweigh the distortionary costs of the FBP. Arguments put forward to defend the FBP, stressing improved service, better distribution and retail networks, and other forms of increased non-price competition, do not stand up to scrutiny either. The book industry produces many titles and beginning authors do not experience severe problems. The FBP may slow down or even stop the declining number of well-stocked bookshops outside big cities, but hinders sales through the Internet and supermarkets.

A comparison of policies towards the book industry in different European countries teaches us that harmonisation is a bad idea. There is not much inter-European book trade, so that book policies hardly distort the single European market. Also, characteristics of book industry, cultural and social features and political preferences of the different countries of Europe differ substantially. It is therefore best to allow member states of the European Union to design their own book policies. For example, a FBP makes more sense for Greece than for the UK as it has a smaller 'language size' and fewer people have access to the Internet. Although there may be a problem of a 'race to the bottom' if VAT-rates are not harmonised, tax competition seems pretty irrelevant for the book market. European countries should be free to lower or abolish VAT on books in order to promote reading.

Many granted privileges in the book industry will eventually be undermined by technical changes. Digital cameras, recording and editing equipment have made low budget radio and television as well as narrow casting possible, thus undermining the monopoly power of public and other broadcasters. Similarly, the Internet has stimulated virtual book suppliers, printing and publishing on demand and E-books. Virtual dictionaries, encyclopaedia and other handbooks have already overtaken, to a large extent, their physical counterparts. A dense network of well-stocked bookshops remains important. While more retailing will take place through the Internet, for some the physical bookshop where one can feel the book and bump into surprise titles and people, will remain indispensable.

There are, however, trends that endanger books, the most important being that people read less and less. Some worry that the next generation will stop reading books altogether, but this may be too pessimistic. First, the population is ageing so that more leisure time becomes available and the opportunity costs of reading decrease. Second, books are doing great. Back in 1947 85,000 books were in print in the United States, against 1.3 million in 1996. This is, in part, due to sharp reductions in production and printing costs. Third, there is no reason to believe that a cultural carrier as old as the book suddenly disappears. Modern technology more than anything else complements books rather than substitutes for it [Cowen (1998)]. However, as (among others) the French historian of written culture Roger Chartier argues:

The image that has become so familiar, that of surfing the web, clearly indicates the characteristics of a new way of reading: segmented, fragmented, discontinu-

ous. If such reading is suited to encyclopaedic texts, whose fragmented structure corresponds to that type of reading, it is disturbed or disoriented by genres the appreciation of which implies a less choppy reading, a familiarity to be maintained with the work, and perception of the text as an original and coherent creation . . . One of the great questions of the future is whether or not digital textuality will be able to overcome the tendency toward fragmentation that characterizes both the structure of texts and the modes of reading that it proposes.<sup>17</sup>

Each new development in the craft has led to outbursts of cultural pessimism allegedly indicating the end of the book. Most of the developments only improved the book business [Cowen (1998)]. Also, prices fell considerably and steadily. The future of the book market may look very different. Chartier's analysis indicates that E-books will replace parts of the market where E-reading already outperforms traditional reading. As for novels, nobody knows. We do not see ourselves reading Shakespeare from screen, but perhaps our children will.

### Acknowledgements

We are grateful for the helpful comments of Marja Appelman, Jonathan Beck, Françoise Benhamou, Richard Caves, Annelies Fontijne, Victor Ginsburgh, David Throsby, an anonymous referee and the participants of the conference on 'The Economics of Art and Culture', Princeton University, 10–12 September 2004.

### References

- Akerlof, G. (1970). "The market for 'lemons': Quality uncertainty and the market mechanism". *Quarterly Journal of Economics* 84, 488–500.
- Allen, W., Curwen, P. (1991). *Competition and Choice in the Publishing Industry*. Institute for Economic Affairs, London.
- Appelman, M.D., Canoy, M.F.M. (2002). "Horses for courses – Why Europe should not harmonise its book policies". *De Economist* 150 (5), 583–600.
- Appelman, M.D., van den Broek, A. (2002). *Boek en Markt. Effectiviteit en Efficiëntie van de Vaste Boekenprijzen*. SCP/CPB, The Hague.
- Barro, R.J., Lee, J.-W. (2000). "International data on educational attainment: Updates and implications". CID Working Paper No. 42, Harvard University.
- Beck, J. (2003). "Monopoly vs. oligopoly in the debate on retail price maintenance for books: A preliminary empirical result". *Diplomarbeit*, Humboldt University, Berlin.
- Beck, J. (2004). "Fixed, focal or fair? Book prices under optional retail price maintenance". SP II 2004-15, Wissenschaftszentrum, Berlin.
- Bertarelli, S., Censolo, R. (2000). "Preference for novelty and price behaviour". Working Paper, Università di Bologna, Bologna.

<sup>17</sup> Available online at [www.uchicago.edu-research/jnl-crit-inq/features/artsstatements/arts.chartier.htm](http://www.uchicago.edu-research/jnl-crit-inq/features/artsstatements/arts.chartier.htm).

- Bittlingmayer, G. (1992). "The elasticity of the demand for books, resale price maintenance and the Lerner index". *Journal of Institutional and Theoretical Economics* 148 (4), 588–606.
- Caves, R. (2000). *Creative Industrie, Contracts Between Art and Commerce*. Harvard University Press, Cambridge, MA.
- Caves, R. (2006). "Organisation of arts and entertainment industries". In: Ginsburgh, V., Throsby, D. (Eds.), *Handbook of the Economics of Art and Culture*, vol. 1. Elsevier, Amsterdam. Chapter 17 in this volume.
- Clerides, S.K. (2002). "Book value: Intertemporal pricing and quality discrimination in the US market for books". *International Journal of Industrial Organization* 20 (10), 1385–1408.
- Coser, L.A., Kadushin, C., Powell, W.W. (1982). *Books: The Culture and Commerce of Publishing*. University of Chicago Press, Chicago.
- Cowen, T. (1998). *In Praise of Commercial Culture*. Harvard University Press, Cambridge, MA.
- Creemers, M. (1999). "In de elektronische boekhandel is informatie belangrijker dan prijs". *Holland Management Review* 66, 58–63.
- Cummings, Jr., M.C., Katz, R.S. (1987). *The Patron State*. Oxford University Press, New York.
- de Grauwe, P., Gielens, G. (1993). "De prijs van het boek en de leescultuur". CES Working Paper No. 12, University of Leuven.
- Deneckere, R., Marvel, H.P., Peck, J. (1997). "Demand uncertainty and price maintenance: Markdowns as destructive competition". *American Economic Review* 87 (4), 619–641.
- Dixit, A., Stiglitz, J.E. (1997). "Monopolistic competition and optimal product diversity". *American Economic Review* 67, 297–308.
- Ecalte, F. (1988). "Une évolution de la loi relative du 10 août au prix de livre". *Economie et prévision*, 5/88, Paris.
- EIM (2001). *De Vaste Boekenprijs – Een Internationale 'Quick Scan*. EIM, Zoetermeer.
- Epstein, J. (2001). *Book Business: Publishing Past, Present and Future*. W.W. Norton, New York.
- European Commission (2004). "Book publishing, publishing market watch". Sectoral Report 2, Brussels.
- Eurostat (2004). *How Europeans Spend Their Time – Everyday Life of Women and Men*. Office for Official Publications of the European Communities, Luxembourg.
- Fishwick, F. (1989). "Les implications économiques du Net Book Agreement". *Cahiers de l'Économie du Livre* 2, 4–31.
- Fishwick, F., Fitzsimons, S. (1998). *Report into the Effects of the Abandonment of the Net Book Agreement*. Cranfield School of Management, Cranfield.
- Goolsbee, A., Chevalier, J. (2002). "Measuring prices and competition online: Amazon and Barnes and Noble". Working Paper No. 9085, NBER, Cambridge, MA.
- Greco, A.N. (1999). "The impact of horizontal mergers and acquisitions on corporate concentration in the U.S. book publishing industry: 1989–1994". *Journal of Media Economics* 12 (3), 165–180.
- Greco, A.N. (2000). "Market concentration levels in the U.S. consumer book industry: 1995–1996". *Journal of Cultural Economics* 24, 321–336.
- Hjorth-Andersen, Chr. (2000). "A model of the Danish book market". *Journal of Cultural Economics* 24, 27–43.
- Holahan, W.L. (1979). "A theoretical analysis of resale price maintenance". *Journal of Economic Theory* 21, 411–420.
- Klein, D.C. (2000). "Web strategies for professional publishers: Developing an information services portal". *Learned Publishing* 13, 83–94.
- Landes, W.M., Posner, R.A. (1989). "An economic analysis of copyright law". *Journal of Legal Studies* 18, 325–363.
- Latcovich, S., Smith, H. (2001). "Pricing, sunk costs, and market structure online: Evidence from book retailing". *Oxford Review of Economic Policy* 17 (2), 217–234.
- Marvel, H.P., McCafferty, S. (1984). "Resale price maintenance and quality certification". *Rand Journal of Economics* 15, 346–359.
- Mathewson, F., Winter, R. (1998). "The law and economics of resale price maintenance". *Review of Industrial Organisation* 13 (1–2), 57–84.

- O'Hagan, J.W. (1998). *The State and the Arts: An Analysis of Key Economic Policy Issues in Europe and the United States*. Edward Elgar, North Hampton, MA.
- Ornstein, S.I. (1985). "Retail price maintenance and cartels". *Antitrust Bulletin* 30 (2), 401–432.
- Ottaviano, G.I.P., Thisse, J.F. (1999). "Monopolistic competition, multiproduct firms and optimum product diversity". Discussion Paper No. 9919, CORE, Louvain.
- Plant, A. (1934). "The economic aspects of copyright in books". *Economica* 1, 167–195.
- Ringstad, V. (2004). "On the cultural blessings of fixed book prices: Fact or fiction". *International Journal of Cultural Policy*. In press.
- Rosen, S. (1981). "The economics of superstars". *American Economic Review* 71, 848–858.
- Rürup, B., Klopffleisch, R., Stumpp, H. (1997). *Ökonomische Analyse der Buchpreisbindung*. Hessischer Verleger- und Buchhändler-Verband, Büchhändler-Vereinigung, Frankfurt.
- Salop, S., Stiglitz, J.E. (1977). "Bargains and ripoffs: A model of monopolistic competitive price dispersion". *Review of Economic Studies* 44, 493–510.
- Soon-Yong, C., Stahl, D.O., Whinston, A.B. (1998). "Gutenberg and the digital revolution: Will printed books disappear?" *Journal of Internet Banking and Commerce* 3 (1).
- Sutton, J. (2000). *Marshall's Tendencies. What can Economists' Know?* Leuven University Press, Leuven & MIT Press, Cambridge, MA.
- Szenberg, M., Youngkoo Lee, E. (1994). "The structure of the American book publishing industry". *Journal of Cultural Economics* 18 (4), 313–322.
- Throsby, D. (2001). *Economics and Culture*. Cambridge University Press, Cambridge.
- Tietzel, M. (1995). *Literaturökonomik*. Mohr, Tübingen.
- Tullock, G. (1980). "Efficient rent seeking". In: Buchanan, J.M., Tollison, R.D., Tullock, G. (Eds.), *Towards a Theory of the Rent-Seeking Society*. Texas A&M Press, College Station, pp. 3–15.
- Uitermark, P.J. (1986). "Verticale prijsbinding van boeken; concurrentie en cultuur – een aanzet tot analyse". In: *De Vaste Boekenprijs: Pro's en Contra's*. Instituut voor Onderzoek van Overheidsuitgaven.
- van der Ploeg, F. (2004). "Beyond the dogma of the fixed book price agreement". *Journal of Cultural Economics* 28, 1–20.
- van Ours, J.C. (1990). "De Nederlandse boekenmarkt tussen stabiliteit en verandering". *Massacommunicatie* 18, 22–35.
- Whyte, J.L. (1994). "Breaking the bookshop cartel". *Long Range Planning* 27 (3), 75–87.
- Yetkiner, I.H., Horvth, C. (2000). "Macroeconomic implications of virtual shopping: A theoretical approach". University of Groningen.

## ARTISTIC LABOR MARKETS: CONTINGENT WORK, EXCESS SUPPLY AND OCCUPATIONAL RISK MANAGEMENT\*

PIERRE-MICHEL MENGER

*École des Hautes Études en Sciences Sociales, Paris, France*

### Contents

Abstract	766
Keywords	766
1. Introduction	767
1.1. A steady growth in numbers of artists	767
1.2. Definitional issues and methodological caveats	770
2. Employment status and careers	771
3. The rationales of occupational choice and risky careers	775
4. Talent, tournaments and the manufacturing of inequalities	779
4.1. Talent	779
4.2. Teams	781
4.3. Tournaments	781
5. The excess supply of artists	782
5.1. Causes of excess supply	782
5.2. Agglomeration and congestion	784
5.3. Monopolistic competition	786
6. Remedies to excess supply? Creative prospects and chances of self-actualization	789
7. Managing the risks of the trade	794
8. Collective action and public support in occupational risk management	797
9. Art as a model for creativity-enhanced work in advanced societies?	800
10. Summary and conclusion	805
References	806

\* I am grateful to Guido Friebel and David Throsby for helpful comments and suggestions on an earlier version of this chapter.



## Abstract

This chapter studies how and why artistic labor markets have expanded along a path of unbalanced growth. Long-term employment which nurtures the Baumolian cost disease persists only in large, heavily subsidized and sponsored organizations. The now dominant project-based system of production, with its functional needs for flexibility, relies on short-term assignments. Large parts of the business risk are transferred down onto the workforce in vertically disintegrated organizational settings. Artists and technical workers act mainly as contingent workers, freelancers and independent contractors; labor supply is patterned by repeated and discontinuous alternations between work and unemployment, and workers cycle between multiple jobs inside and outside the arts. Thus artistic labor markets display the main characteristics of a textbook model of imperfect monopolistic competition: excess supply of labor, unbounded differentiation of production, reputational rents, a population of small firms that has been growing as fast as the number of artists. On the supply side, the attractiveness of artistic occupations has to be balanced against the risk of failure that turns ideally non-routine jobs into ordinary or ephemeral undertakings. Learning by doing plays such a decisive role that in many artforms initial training is an imperfect filtering device. Individuals learn to manage the risks of their trade through multiple jobholding, occupational role versatility, portfolio diversification of employment ties, and income transfers from public support, social insurance and social security programs. Ironically, the study of the artists' risk management shows how rationally they behave, although artistic work may be highly idiosyncratic. Thus artists may be seen less like rational fools than like Bayesian actors.

How do vertically disintegrated systems of production shape individual careers and organizational behavior? Loose employment relationships do not preclude contractual stability. Employers use reputations as screening devices and signals of employability. Artists learn how to compose balanced sets of recurrent and non-recurrent hiring ties in order to secure a living as well as to increase their human capital. Considerable inequalities in amounts of work and earnings are observed, caused by the skewed distribution of talent and by joint consumption technologies that turn small differences in talent into huge earnings differentials. Inequalities may also trace back to the way a disintegrated labor market operates, since both the allocation of piecemeal work based on reputational rankings and team formation based on selective matchings magnify the power of differences in talent and work opportunity to increase inequality. These factors should not cause the kind of permanent excess supply of labor in the arts that has been noted for decades if the occupational commitment of artists were not combined with the management of business uncertainty through overproduction of infinitely differentiated goods and services.

## Keywords

labor markets, excess supply, monopolistic competition, uncertainty, occupational choice, occupational risk management, career, creativity

*JEL classification:* J41, J44, Z11

## 1. Introduction

### 1.1. A steady growth in numbers of artists

Evidence of sustained growth in artistic employment over recent decades is amply documented by several surveys and Census sources, and trends are quite similar in most advanced countries. For instance, in France, over the period 1982–1999 the number of artists grew at a rate of 98 percent; in the USA, from 1980 to 2000, the rate of increase was 78 percent. In both cases, the growth in numbers of artists was much higher than for the civilian labor force. All artistic occupations with the exception of musicians have seen a steady shift towards a higher proportion of women; by contrast, ethnic composition of the artistic workforce remains unbalanced, the non-classical music sphere being one of the relatively few exceptions.

One may speculate as to why artistic employment growth has been so rapid. On the demand side, increases in real disposable per capita income have shifted demand curves for the arts and resulted in an increasing fraction of national income and employment being devoted to the arts. In Europe much of the employment gain, especially in the 1970s and the 1980s, has been attributable to the steady growth of federal and local government subsidies leading to a large expansion of the non-profit sector providing services for artistic training and for conservation and display of cultural heritage. Public spending under non-arts headings (e.g., local economic development, urban regeneration) as well as support for the cultural industries has also stimulated opportunities for cultural employment. Furthermore several industrial sectors which draw heavily on the skills of artists and other creative occupations underwent rapid expansion during the 1980s. The most striking change within the cultural industries was the rapid growth of the audiovisual and broadcasting sector, along with growth in the advertising industry, the new media industries (video, corporate video) and the computer game industry. No less striking is that employment in these growing sectors was mainly on a short-term contract or freelance basis, which magnifies the shift towards numerical flexibility observed elsewhere in the economy [Smith (1997)]. One should also mention the expansion of the crafts and of the design sector, which increasingly contribute to the rise in the numbers of artists in Census data [Feist (1998)].

The economics of artistic labor markets has long paid attention on one hand to the unbalanced growth of stable organizations employing workers on long-term contract and subject to the well-known Baumol disease, and on the other hand to how individuals acting as self-employed workers have to be compensated for highly uncertain prospects in artistic occupations. Artistic labor markets have now evolved to approximate the spot-market model of textbook economics; employment relationships on an unfixed-term basis have largely vanished, and short-term hirings and self-employment strongly dominate. In that respect, arts have often been mentioned as forerunners in experiencing the trend toward increasingly flexible high-skilled labor markets where workers may be hired for only two or three hours, without any costly dismissal proce-

dures. In fact, some parts of the skilled labor force have long been experiencing high flexibility in employment relationships. As stated by Okun (1981, pp. 82–83),

... these weak employer–worker attachments also seem to apply to certain types of blue collar craftsmen who have relatively high skills and earn high pay – construction and dock workers, workers in the printing industry, and so on. These characteristics seem to arise most prominently when 1) an industry has many firms within a locality; 2) a firm has extremely variable demands for labour; 3) the worker’s skill is “general” in Gary Becker’s sense, that is, readily transferable across firms within an industry; and 4) the individual worker’s degree of skill categorized by conventions that develop among employers or unions or through government-sponsored occupational licensing. Carpenters thus may be classed as apprentices, journeymen, or masters; and references from one employer to the next carry weight. In such cases, workers develop an attachment to a local industry rather than to an individual employer.

Okun might have mentioned artists and technical cultural workers as well. Thus the labor market in the arts is a rather paradoxical competitive one. On one side employment is more and more contingent, as in secondary labor markets; on the other side individuals are highly skilled and non-substitutability is a core value, as in the so-called primary labor market. Therefore, though the segmentation of the workforce is fairly strong, the distinction between primary and secondary markets could hardly apply to the arts.

How do short-term assignments translate into worker flows and careers? From a *labor supply* standpoint, one artist equals one long-term occupational prospect, especially when employment relationships are long-term and careers are well patterned. But the gap is widening between the vocational commitment and the way it transforms into a career: self-employment, freelancing and contingent work bring in discontinuity, repeated alternation between work, compensated and non-compensated unemployment, searching and networking activities, and cycling between multiple jobs inside or outside the arts. From a *labor demand* standpoint, the spot-market profile of the arts makes things simpler: the focus is on contracts, on hirings and on works sold on the market. Thus the labor market here can be investigated at its most disaggregated level, that of the series of hirings, of work opportunities and of bargaining relationships. These represent an individual artist’s working life in a given period; when considered longitudinally, they display a career trajectory.

What is the impact of the fact that labor demand is expressed mainly in terms of contingent work? Numerous studies have shown that an increase in the number of artists may be far from corresponding to a similar increase in the level of activity. If there is more work but an ever more rapidly growing number of individuals, a fiercer competition takes place that implies higher inequalities in access to employment, more variability in the level and schedule of activity and on the whole work rationing for those who share the labor pie and cycle more often from work to unemployment or from arts work to arts-related or non-arts work.

The resulting overall picture of artistic labor markets and of their growth is however quite a paradoxical one: employment, underemployment and unemployment have all been increasing steadily and simultaneously. The pattern of change may vary across the different artistic occupations, but the trend is almost everywhere the same. Obviously, fluctuations in supply and demand of artistic labor do not provide a satisfying explanation of what appears to be a highly unbalanced growth. Several historical studies on artistic professions have repeatedly insisted on an “oversupply of artists” phenomenon, which they have associated with changes in the organizational apparatus of the art worlds or with technological innovations or, more radically, with the emergence and expansion of a free market organization for the arts. But in each case, ad hoc arguments may overshadow structural disequilibria; the present development of labor markets for the arts, by highlighting an apparently irresistible trend towards flexibility, helps to understand the underlying processes of such a course of development, namely the pervasive uncertainty of artistic careers, and the ways for individuals and organizations to handle uncertain prospects and to manage individual and business risks.

From a large sample of studies on artistic labor markets, the following picture emerges. Artists as an occupational group are on average younger than the general work force, are better educated, tend to be more concentrated in a few metropolitan areas, show higher rates of self-employment, higher rates of unemployment and of several forms of constrained underemployment (non-voluntary part-time work, intermittent work, fewer hours of work), and are more often multiple jobholders. Not surprisingly, artists earn less than workers in their reference occupational category (professional, technical and kindred workers), whose members have comparable human capital characteristics (education, training and age). And they experience larger income variability, and greater wage dispersion.<sup>1</sup> Taken together, these features portray oversupply disequilibrium [Stigler (1962)]. Moreover, they have been documented for so long that excess supply of artistic labor appears to be permanent and may act as a true structural condition of the arts’ unbalanced growth.

A closer examination of descriptive statistics would provide us with considerable detail about each of these traits and would allow for the kind of fine-tuned differentiation between the several categories of artists that we find in the comprehensive NEA report on *Artists in the Work Force* [Alper et al. (1996)], in the Australian report by Throsby and Hollister (2003), in the British one by O’Brien and Feist (1995) or in French official annual reports based on Census and Labor Survey data [Observatoire de l’emploi culturel (2004a, 2004b)]. However, our main aim here is rather to review explanatory models of work organization and labor supply in the arts and to focus on four main issues: the status of employment and career patterns, the rationales of occupational choice, the oversupply of artists, and occupational risk diversification. Our approach will deliberately be a multidisciplinary one, bringing together a number of studies in sociology, economics and history.

<sup>1</sup> See also Chapter 23 by Alper and Wassall in this volume.

## 1.2. *Definitional issues and methodological caveats*

The definition of art and culture has obviously been broadened as cultural policies have developed. The anthropological definition of culture has become more and more legitimate as public support has taken into consideration the local community level and its whole apparatus (amateur activities, associations, so called socio-cultural activities) by setting up links between art, culture, leisure, schooling and social work. At the same time, one may note an opposing trend in cultural policies; the development of a discourse about culture as a real economic sector. A new form of “cultural accountancy” has in fact emerged that seeks to quantify the economic output of public spending on culture; of course, the wider the definition of culture, the more culture can claim to play an economic role, and the stronger the economic rationale of public support may appear to be, at least at first sight.

These changes also raise definitional issues concerning what the artistic occupations are and where the boundaries of the artistic sector lie. Research on British Census data [O’Brien and Feist (1995)] builds on an occupational as well as on a sectoral breakdown; the redefined categorizations cross both classifications. As a result, cultural work appears to spread across a number of professional occupations and industrial activities; among individuals involved in the cultural sector, 25 percent work in the cultural industries in cultural occupations, 40 percent have cultural occupations outside the cultural industries, and 35 percent work in the cultural industries in non-cultural occupations.

Almost every research report on artistic occupations opens with a list of the limitations and discrepancies of Census and non-Census data including: problems of defining who are professional artists and how their occupation is determined; the delimitation of artistic fields, and the inclusion or exclusion of peripheral specialties within a field; the variations in job classifications and the periodic addition of new occupations to the artists’ subset in the Census classification; and the lack of any serious treatment of multiple jobholding. Regarding the tricky issue of the comparative merits of survey versus Census data, one need only mention the primary source of most variations, as noted above: the definition of the artist. The Census uses a parsimonious classification rule, which narrowly interprets the “chief job activity or business last week”. Two serious problems confront researchers using the Census source [Wassall and Alper (1990)]. First, since many persons who identify themselves as artists are multiple jobholders, their labor market behavior (earnings, working time) cannot be attributed solely to their artistic involvement. Second, those who earn a living mainly in non-artistic jobs, and yet identify themselves as artists, are reported as members of non-artistic occupations.<sup>2</sup> Surveys, by contrast, generally use one or several criteria [Frey and Pommerehne (1989)] and may categorize various activities as art in accordance with the particular interests of

<sup>2</sup> One should add that measurement of the unemployment level in the arts is consequently disputable, since artists switch temporarily to work in different occupations when unable to make a living in their primary vocational field, without stopping producing art works; they would therefore not be classified as unemployed in their artistic occupation if they are primarily engaged in non-artistic work during the Census week.

the researcher. The most controversial of these criteria is, of course, that of subjective self-definition as an artist; this criterion encapsulates a temporal dimension of occupational commitment, since artists may cycle between several jobs and yet continue to think of themselves as artists.

As stated by Adler<sup>3</sup> “a study of artists in a society in which occupational membership is (fortunately) not defined or restricted by a guild, an academy or a state system of licensing can neither comfortably ignore problems of occupational definition nor resolve them”. Indeed, although some of the most remarkable studies by sociologists, economists or art historians on art labor markets and careers have been historical ones,<sup>4</sup> the definition of the artist as well as the orderly course of an artistic career appear today to be dependent variables in the operation and evolution of highly competitive and contestable labor and product markets, interacting or not with state intervention.

Such theoretical and methodological issues are by no means new in social science; labels, taxonomies and classification systems are core issues in interactionist and constructivist sociological theories. Sociologists deal with these matters more cautiously than do economists; while the former run the risk of questioning endlessly the significance of any quantitative measurement, the latter run that of taking for granted that Census data (almost the only source they use) will lead to strong results by virtue of sophisticated econometrics, once the obvious limitations from which the data suffer have been acknowledged.

Despite all these discrepancies, a review of a number of recent studies will allow us to highlight the key issues for a comprehensive approach to artistic labor markets.

## 2. Employment status and careers

The steady increase in the number of artists across all art sectors during the last three decades appears to be driven by the rapid increase of independent, self-managed work, with increasing numbers of artists now to be found in the sectors where self-employed practitioners work such as creative writing, the visual arts and the crafts, and by the rise of contingent work wherever salaried employment relationships prevail. Unfixed-term employment represents a rapidly declining share of the cultural work force today. Only long-lasting organizations such as symphony orchestras, opera houses, and visual art and music schools may hire a large part of their artistic personnel on long-term contracts [Towse (1993, 1996)].

Among the salaried artists who work on a long-term basis, musicians and their careers meet a rather well-patterned job system that has been carefully studied.<sup>5</sup> Bureaucratic careers can be found in permanent orchestras with positions ranging on a well-defined

<sup>3</sup> Quoted in Alper et al. (1996).

<sup>4</sup> For example, White and White (1965), Montias (1982), Ehrlich (1985), Warnke (1985).

<sup>5</sup> See, for example, Westby (1960), Faulkner (1973), Allmendinger, Hackman and Lehman (1994), Allmendinger and Hackman (1996).

scale of status; a majority of the orchestral players become anchored in their organization, experience little or no mobility and become committed to their role in a stable work setting. Advancement on the job ladder is limited since top ranks are filled mainly through external recruiting, so that the mechanics of vacancy chains operate rather poorly. Individual career opportunities develop through mobility within a stratified set of organizations ranked on a hierarchy of prestige, musical excellence, caliber of musicianship, working conditions and operating budgets, either towards similar or higher positions in higher ranking orchestras or towards higher status positions in lower ranked orchestras. Such moves are few in a professional lifetime; as described by Westby (1960), each musician behaves like his or her own employment agency, compiles an inventory of probable and possible jobs, gets information about the approximate ages, professional histories and abilities of the current holders of the most desirable jobs, so as to be prepared for an opportunity that may appear only once in a lifetime. The curvilinear profile of such career mobility means that the artist has to move early to reach the peak of this organizational set, and that chances of mobility diminish rather quickly after an age of 30 or 35, at least with respect to the top-level tier of prestigious organizations. Publishing houses [Powell (1985)] and architecture firms [Blau (1984); Champy (1998)] are additional examples of permanent organizations that combine constraining hierarchies of jobs and career development through lateral mobility. However an increasing proportion of salaried cultural workers now work on a short-term contractual basis. Proportions may vary with national contexts and occupations, but trends are similar and exhibit the search for increasing flexibility and the minimization of fixed costs in the arts.

The search for flexibility is a core feature of artistic work, due to the “high rate of change over time of the content of activities”, according to Stinchcombe’s (1968) phrasing. This occurs for at least three reasons:

- artistic products are often designed as prototypes and their market value depends on their originality and on a more or less pronounced differentiation;
- the combination of activities needed to produce a movie, play or opera involves a large number of different artistic occupations and crafts, and each participant shifts to a new project just hours, days or weeks after the initial one, with new requirements and challenges; and
- consumer versatility and taste for novelty give social and economic value to newness and originality to the extent that these are more or less radically unpredictable. Uncertainty is the true condition of the breakthrough innovation that opens up to its author a new (temporary) monopoly; it is also the threat contained in the destructive aspect of every true innovation. Tastes are subject to unpredictable shifts, especially in the most speculative art markets such as popular music, hyped contemporary painting, blockbuster novels, mass audience designed movies and serials.

Flexibility can be attained through several avenues: a system of performance contracts, a system for transmitting information about the performance capacities of people, and a minimization of overhead costs [Stinchcombe (1968)]. For each project – film,

opera or theater performance, musical show, gallery exhibition – new teams are formed and then dispersed, or parts of the production process are subcontracted afresh. Networks help to build stable relationships that are needed to lower transaction costs; hiring procedures very often operate through patronage and trustworthy ties among peers that rapidly convey reliable information about skills and talents; increasingly hiring also operates through the development of a brokerage system that enhances the role of talent agencies in mediating the labor market for contingent employment [Bielby and Bielby (1999)]. As a result, vertical disintegration in the production and distribution of performances and products has increased. Firms minimize their risks by using contractual relationships which transmit the market uncertainty down the hierarchy of control to subcontractors and ultimately to individual workers.

Consequently, expansion of artistic worlds leads to a rapid increase of the population of employers and small organizations. In the record and motion picture industries, for instance, although oligopolistic market control by major companies remains a striking feature mainly through the control of distribution and finance [Aksoy and Robins (1992) and Storper's reply (1993)], a vertical disintegration scheme at the production level has occurred resulting in an increasing number of independent film producers [Christopherson and Storper (1989); Storper (1989)], record companies [Burke (1997)], and publishing houses [Boin and Bouvaist (1989)]. In the performing arts, the expansion of the non-profit sector and the increase in public support have favored the multiplication of dance companies [Sussmann (1984)], theater groups [Menger (1997)] and music ensembles. Even if demographic trends concerning the rise and fall of organizations differ across the various arts scenes,<sup>6</sup> on the whole the expansion of the “craft-administered” and “flexibly specialized” production sector, with its growing product differentiation, has brought temporary organizations or small organizations to hire performing artists and craft workers almost only on a short-term basis.

Thus performing artists act like independent contractors and cycle between employers and between work and unemployment spells. Although asymmetrical, the relationship between the employer and the freelancer is that of a matching process where both sides build a career interdependently, as carefully demonstrated by Faulkner in his study of the Hollywood job system.<sup>7</sup> Artists as well as entrepreneurs accumulate a history of results and their performance ratings translate into reputations and into distinct industry identities. Careers are two-sided affairs, with entrepreneurs making distinctions among qualified artists, and artists (directors, screenwriters, composers, etc.) making distinctions among film productions. Careers advance incrementally through recurrent and non-recurrent matches; artists learn how to spread their occupational risks by forming career portfolios, i.e. by mixing one-shot ties which are the normal feature of a loosely coupled hiring system and recurrent “bread and butter” accounts with a few producers. Faulkner shows that such a spreading of accounts allows the artist to hedge his or her

<sup>6</sup> For an extreme example, see the case of dance music [Hesmondhalgh (1998)].

<sup>7</sup> See Faulkner (1983) and Faulkner and Anderson (1987).



bets, to get information about a wider environment and to accumulate credits in a human capital investment program through a variety of work, stylistic diversification and adaptation to changing teams. As cumulative productivity profiles greatly differ, distinct matching proclivities segment the labor market; team matchings are neatly stratified in equivalent classes of market agents. Yet given the high variance in activities and the volatility of the cultural industries, career advancement and attainment are never secured.

In the creative arts, self-employment has been for long the prevailing work status. Self-employed artists' careers display most of the attributes of the entrepreneurial career form: the capacity to create valued output through the production of works for sale; the motivation for deep commitment and high productivity associated with their occupational independence (deriving from the capacity to control their own work, a strong sense of personal achievement through the production of tangible outputs and the ability to set their own pace); and a high degree of risk-taking, as shown by the highly skewed distribution and high variability of earnings [Alper et al. (1996)]. Thus, as stressed by Freidson (1986a), self-employment may bring with it only an illusory independence and autonomy; the freelancers who fail to move into the inner circles of successful colleagues get locked in a precarious situation. Being neither a stage process nor a simple bargaining process [Abbott (1990)], career trajectories under a self-employed or a contingent work status combine traits from professional as well as from entrepreneurial careers as defined by Kanter (1989). Artists rely on skills as well as opportunities to take on evermore challenging assignments that bring them greater knowledge and more rewards; they have an external market value based on reputation; they exhibit less loyalty to particular organizations than to the professional community; and they may manage their working life much as property owners do when spreading their risks. Indeed flexibility requirements and career concerns lead individuals and organizations to combine different contractual forms and use many opportunities within a whole range of contractual arrangements; for example, musicians in orchestras can also be hired as freelancers for some studio recording jobs and hold a teaching position in a conservatory, so that the employment-status distinction is somewhat blurred at the individual level.

Thus artists, even if operating as single input firms, may behave like entrepreneurs managing small businesses and work portfolios, and their labor market may be compared to a network of small ad hoc firms trading along matching processes from one project to the other. The analogy with small firms may be taken one step further when multiple jobholding behavior and role versatility are brought into the picture, as shown below. The large number of small artistic organizations and their high rate of turnover may be explained that way, since composers [Burke (1997)], choreographers [Sussmann (1984)] and stage directors [Menger (1997)] can easily set up companies or fringe firms by relying on a portfolio of resources and multiple roles. Although brokerage has emerged as a major device for mediating labor allocation and matching processes in a highly fragmented labor market [DiMaggio (1977); Bielby and Bielby (1999)], artists may be induced to exercise supervisory or managerial skills and in so doing to blur the line between management and labor [Christopherson (1996)]. In that respect, the

artist's earnings, like those of any self-employed worker, depend not only on her skill, talent and effort, but also on how well she performs the managerial and entrepreneurial functions [Aronson (1991)].

According to Weick's notion of self-designing organizations [Weick (1979); Weick and Berlinger (1989)], careers in such a labor system are subjectively patterned since they are committed to impermanence, to cumulative learning and exploration rather than tied to external career markers. Regarding the dynamics of personal growth and achievement, one striking feature of careers in the arts is their temporal aspect: to take only each end of a working life in the arts, precocity often plays a significant role, not only as a mythical feature of the "self-generating genius" topic described by Kris and Kurz (1987), but also a symptom of the ambiguity of the transition from training to work, since many creative artists and performers produce serious work and get credits before their formal training is complete [Menger (1997)]. Conversely, late starters are particularly prevalent among writers [Throsby and Hollister (2003)] and the increasing occupational flexibility of careers also leads to late entry for a second career, whether this be corresponding to a deferred vocational choice or to reconversion following redundancy, as is the case in the crafts sector surveyed by Knott (1994). Of course, self-employment status typically allows for such switches.

The span of a career varies greatly with the type of art (e.g., dance vs. creative writing), with the subsector of each art world (classical dance vs. contemporary dance), with the occupation in it (performing vs. creative work), and with the organizational and market features of each world.<sup>8</sup> Only sharply contrasting examples may be cited here: a conductor's career may extend until near the end of his life with almost no time for retirement, but classical dancers have career schedules constrained by severe physical requirements [Baumol, Jeffri and Throsby (2004); Rannou and Roharik (2006)]. In the high arts sphere, reputation may be a factor explaining exceptional longevity, enhancing the sense of achievement well beyond the average working-life terms [Anzieu (1981)]; furthermore the reputation capital may be converted into an artistic and economic rent, since the famous artist faces an inelastic demand for his praised work [Moulin (1987)]. By contrast, skyrocketing success in the mass-market arts and entertainment industries is subject to sudden shifts in market demand towards new competitors, and is characterized by highly volatile reputations.

### 3. The rationales of occupational choice and risky careers

In most advanced countries, census data provide quite similar pictures about artists' earnings: mean annual earnings appear to be less than those in occupational groups which require similar levels of professional training and qualification. Filer (1986), in a

<sup>8</sup> For empirical research results on careers, persistence in occupations, quit determinants, and transition profiles, see Chapter 23 by Alper and Wassall in this volume.

provocative paper, claimed to have refuted the “myth of the starving artist”, estimating the income penalty in artistic occupations to be less than 10 percent. But his study did not distinguish between arts and non-arts sources of income nor between income from creative activity and that from arts-related work. Moreover, the income gap estimated by Filer varied greatly among the different artistic groups (–69 percent for dancers, +58 percent for actors and directors).

In short, as summarized by [Throsby \(1994b\)](#) and by [Alper and Wassall \(2006, Chapter 23 in this volume\)](#) in their review of numerous studies, artists actually appear to suffer from significant income penalties, to have more variable income both across time for an individual artist and across artists at a given point in time, and to get lower returns from their educational investments than is the case in other comparable occupations. Although data based on similar sources and similar methodological design may be difficult to obtain for a careful comparison of each category of artists’ incomes over time, the distributional evidence remains the same: the skewed distribution of artists’ income is strongly biased towards the lower end of the range and artists as a group experience huge income inequalities. Nevertheless artists are not deterred from entering such an occupation in growing numbers, nor is there as much withdrawal from artistic careers as would be expected.

Are artists irresistibly committed to a labor of love, or are they true risk-lovers, or perhaps “rational fools”, to use Sen’s phrase [[Sen \(1976\)](#)]? The “labor of love” argument [[Freidson \(1990\)](#)] insists that occupational commitment and achievement in the arts cannot be matched to the monetary considerations of a market economy of exchange; they should better be conceived as skilled and sustained activities that entail a transfer value and that artists carry out by making a living in host occupations such as teaching. Artists’ notion of their “calling”, analyzed by [Kris and Kurz \(1987\)](#) as an historically recurring feature of artistic biographical narrative, calls to mind the “inner drive” reported by [Jeffri \(1991\)](#) and by [Throsby \(1994a\)](#) as the foremost criterion of professionalism according to US visual artists. The ideology inherited from the “art for art’s sake” era may even reverse the meaning of success and failure, so that only recognition by the peer group matters, at least in high art worlds [[Bourdieu \(1992\)](#)]. One way to deal with this ideological dimension is to turn it into an inherent cultural trait – a kind of occupational characteristic that goes along with artistic life or, to be more precise, that blurs the boundaries between occupation and private life. However, once this trait is regarded as belonging to the initial socialization process of the artist via a very early manifestation of ability and taste for the arts, such an explanation turns out to be highly deterministic and ultimately tautological; artists are committed to their art and linked to their community of fellow artists whatever degree of success in the market they may meet. Inescapable commitment results in a highly inelastic labor supply function.

The second argument is that of occupational choice under uncertainty: artists may be risk-lovers (whatever origin one may assign to this preference), or they may be induced to take risks by a probabilistic miscalculation. Occupations where enormous rewards are concentrated in the hands of a small number of practitioners while the majority of entrants may do poorly entail a high degree of uncertainty; entry into these fields

is like a lottery where players overestimate their chances, as has been emphasized by Alfred Marshall (1947). The analogy with a lottery is not entirely appropriate; while it is helpful to think of the skewed distribution of incomes as a matrix of payoffs, it would be misleading to suggest that success is purely random and has nothing to do with individual abilities.

A third, less deterministic view may be offered that substantiates an occupational choice dimension without overshadowing the characteristics either of work or of workers. Rewards in artistic jobs are of two sorts: monetary and non-monetary, the latter being “psychic income” flows which have in fact been regarded for a long time as an essential dimension of work. Analytically speaking, every job can be regarded as a bundle of characteristics; wage differentials compensate for more or less attractive work and equalize among workers the total monetary and non-monetary advantages or disadvantages. This theory of equalizing differences [Rosen (1986)], which goes back to Adam Smith, seeks to explain the diversity of characteristics of work and workers by giving central consideration to individual preferences and choice, provided that there is perfect information on both sides of the market. Artistic work can be considered as highly attractive along a set of measurable dimensions of job satisfaction that include the variety of the work, a high level of personal autonomy in using one’s own initiative, the opportunities to use a wide range of abilities and to feel self-actualized at work, an idiosyncratic way of life, a strong sense of community, a low level of routine, and a high degree of social recognition for successful artists. All these benefits have a shadow price, which may be compensated for by a lower income than would be expected from less amenable jobs.<sup>9</sup>

The benefits derived from non-monetary income are, however, not of a uniform magnitude; an analysis in terms of equalizing differences requires that we adjust the total amount of these benefits according to the job, the level of professional achievement, and the conditions which prevail for those in the profession who, still waiting for success, are forced to take on secondary jobs. Comparisons between artists salaried and independent artists [Fohrbeck and Wiesand (1975); Taylor (1987)] reveal, for example, that the latter obtain higher levels of non-monetary satisfaction, but have lower average incomes, higher levels of job insecurity, higher rates of unemployment and greater variance in individual incomes around the mean. On the other hand some studies reject to a great extent the presence of any compensating “psychic income”: the emblematic case of orchestral musicians illustrates the counter-mythology of the artist subjected to the constraints of an organization, resigned to a humdrum and narrowly-specialized job that is very distant from what long years of apprenticeship oriented towards individual accomplishment in a soloist career had led him to expect [Arian (1971)].

Conversely, in contingent work the risk of unemployment is pervasive and insurance devices through long-term contractual relationships are by definition missing. The typical worker will view the risk of unemployment as something that must be compensated

<sup>9</sup> It should be noted that in strong contrast to the ideological argument, especially to its deterministic aspect, people discover what a non-routine job really is only by experiencing it.

for by a higher hourly wage. Such compensation for uncertain labor prospects is in fact observed in the performing arts since intermittent artists and workers earn higher hourly wages than those employed on a long-term basis [Debeauvais et al. (1997)]. The wage premium is paid by employers in order to draw on a reserve army of underemployed individuals whose availability has to be secured; a loss of flexibility in employment decisions would be more costly for firms. Yet this compensating differential scheme operates only imperfectly, since hourly wages are no higher for seriously underemployed workers than for their more successful colleagues.

Compensating wage differentials therefore play their role mainly at the industry level. Individual differences in hiring probabilities are, by contrast, not subject to compensation, leading to another kind of risk. The mechanics of freelance and contingent employment are such that accumulation of hiring acts as a reputation signal in a self-reinforcing process: hiring calls for more hiring. Thus as the intermittent working system expands, at any given time the number of job candidates increasingly exceeds the supply of full-time jobs. In their extensive study on the vertical disintegration and flexible specialization trend in the Hollywood film industry and on its effects on the labor market, Christopherson and Storper (1989) showed that through subcontracting, financing and distribution of independent producers, utilization of less costly production methods and expansion of auxiliary markets, the demand for short-term contract workers increased. They went on to demonstrate that the aggregate quantity of work available increases far less rapidly than the pool of individuals employed intermittently, generating a growing competition and resulting in a decreasing average participation in production. Thus, when production undergoes a process of increasing vertical disintegration, employment instability and labor market segmentation develop; since job allocation takes place on an individual basis and involves on-the-job accumulation of skills and reputation, experienced and network-building artists and workers are frequently hired and face less discontinuous employment than beginners and individuals only loosely connected with the most active entrepreneurs. Thus differences in annual earnings of workers may reflect differences in hours worked more than in wage rates.

Research on the French performing arts labor market highlights these mechanics of work contingency [Menger (2003)]. The French labor market for the performing arts has constantly expanded over the period 1986–2002. However, the supply of work (the number of artists working) has evolved at rates of increase much higher than the demand-side trend (the number of hirings, the number of worked days declared and the total amount of earnings). As a result, the median amount of working time and earnings per artist decreased over the period, although the number of hirings increased; individual intermittent work was increasingly fragmented in shorter hirings, and competition turned out to become fiercer among the growing numbers of artists sharing the less rapidly growing “work pie”. People have been partly compensated for the increasing risk that goes along with the shortening of individual hirings, since hourly wages have been increasing faster than in other sectors. On the whole, however, the decrease of median earnings over the period indicates that employers do not insure the artists they choose to hire under such a working scheme against the consequences of the unbalanced

growth of that labor market. Employers and consumers may benefit from the increasing variety of talents supplying their work, but at the expense of increasing variability in individual working arrangements, both across the workforce and during the career of each artist.

It should be also stressed that the non-monetary dimensions of work contingency might vary with a freelancer's age [Spilerman (1977)]. Artists offer many examples of a "career-line vulnerability to aging"; as they get older, freelancers such as actors appear to be increasingly sensitive to job insecurity and to the steady strain of searching for jobs, of gathering information about new projects and of maneuvering repeatedly to remain visible in a highly competitive labor market [Laplante (1990); Menger (1997)]. Orchestral musicians and dancers also experience well-patterned sequences of job change over their life cycle; the upward mobility chances of the former decrease quite abruptly after about age 35, inducing them to adjust their occupational commitment [Faulkner (1973); Allmendinger, Hackman and Lehman (1994)], while the latter have to plan their transition to a new career at about the same age [Federico (1983); Baumol, Jeffri and Throsby (2004)].

#### **4. Talent, tournaments and the manufacturing of inequalities**

The "equalizing differences" argument is attractive for its elegant parsimony. Artists who remain in artistic occupations despite low and uncertain earnings gain something else that has to be taken into account in order to preserve the rational occupational choice frame; the additional income flow that one would expect to draw from another occupation has been exchanged for psychic goods. However such an argument formulates its notion of the compensating wage premium with respect only to the differences in average income levels across occupations standardized for a number of individual income-related characteristics (mainly education, experience, age, sex and ethnicity, location of residence and of work). From a distributional perspective, artistic occupations show a strikingly high variance in income. Factors behind this skewedness include talent, the formation of teams and the existence of tournaments in the arts.

##### *4.1. Talent*

Stinchcombe (1963) distinguishes between talent as a complementary factor of production and talent as a nearly additive factor. The former is found in firms, activities and positions (e.g., scientific research, soloist performances in classical music concerts and lyric productions) where output value may benefit more than proportionately from inter-individual differences in levels of ability; accordingly, earnings inequalities are high. By contrast, the distribution of rewards is less skewed and seniority a more important factor where individual performance has a less dramatic impact on the value of the total pro-

duction, as in a symphony orchestra.<sup>10</sup> Moreover, in the first case small differences in talent can become magnified into wide earnings differences, as shown by Rosen (1981) in his superstar model;<sup>11</sup> on the demand side lesser quality is a poor substitute for greater quality, so that preferences are strongly biased towards the latter, while on the supply side, due to joint consumption technology (that of mass production and the distribution of art and entertainment through records, books, TV, radio, etc.), the marginal costs of production do not rise in proportion to the size of a seller's market, but profits do. This is all the more so as media technologies and the internationalization of markets expand the scope of talent valorization. Such a model is consistent with the distribution of incomes observed in the industries relying on scale economies of joint consumption.<sup>12</sup> However, it has been objected that the basic assumption that small differences in talent may lead to huge return differentials requires a measurement of talent and quality other than income [Hamlen (1991, 1994)]. As that measurement is impossible to standardize in those artistic fields in which creativity runs against widely accepted canons (as opposed to the calibration of performances of a standard repertoire of works by unequally skilled performers), the explanatory power of the Rosen model underscores why the process of valuation of art and artists is indeed subject to considerable inflexibilities, asymmetries and imperfections.

Creative artists and their works are usually ranked according to notions of talent rather than skill. Artistic creation is built on a distinctive property, that of fine-grained differentiation of its products due to highly individualized strivings for originality and novelty. But how can talent be measured if its embodiments come to life under the rule of infinite differentiation? And should we accept in the name of creative talent all aesthetic infringements, especially those that break repeatedly with conventions, traditions and norms? The multidimensional nature of differentiation encourages recognizing many different embodiments of originality as true manifestations of creative talent. Yet critics, experts and consumers never cease making comparisons by ranking filmmakers, visual artists, writers, composers or actors. Both the market sellers, experts, critics, and eventually the end consumers sort and organize in a hierarchy those products of individual creativity which the criterion of originality by itself would tend at first sight merely to juxtapose. This occurs through multiple comparisons, disputes, controversies, strategic maneuvers, affiliated evaluations and marketing policies that turn horizontal differentiation into a vertical and inegalitarian one, based on more or less overtly publicized market scores and reputation ratings.

<sup>10</sup> On the additively separable vs. multiplicative production function, see also Caves' (2000) and Seaman's (2003) comparative study of cultural and sport economics.

<sup>11</sup> See further in Chapter 25 by Adler in this volume.

<sup>12</sup> See, for example, Menger (1997) for differences in actors' earnings distribution in theater vs. audiovisual and cinema industry.



#### 4.2. Teams

Another main dimension of this manufacturing of spectacular inequalities is the matching of individual skills and abilities within a world seeking to mine increasing productivity benefits from new designs of teamwork. The flexible organizational architectures most prevalent in this world (networks, project-based organization and vertically-disintegrated systems of production) favor structuring teams by co-opting professionals of similar reputation or quality – in other words by selective matching. A successful career means upward mobility within a stratified world of network collaborations by optimally mixing recurrent and unique ones [Faulkner (1983); Baker and Faulkner (1991)]. Theoretically speaking, the selective matching process between similarly high-skilled agents or talented creative professionals within team projects combines Rosen's analysis of multiplicative quality effects with Gary Becker's analysis of matching in marriage markets, as proposed by Kremer (1993) in his O-Ring theory of economic development. To the Stinchcombe distinction between the complementary vs. additive dimension of talent as a production factor, this model adds the notion of increasing returns to skill according to a non-standard production function, once the workforce of a firm or of a team project is considered as a whole.

#### 4.3. Tournaments

In the art world, as in the sport world, highly unequal distributions of both monetary and non-monetary rewards are not only tolerated but demanded and even celebrated through all sorts of devices: celebrity tournaments, prizes, Oscars, awards, competitions of all sorts, publicity for highly priced artworks, lists of best selling novels and records, media coverage of super incomes, etc. Even though individual dedication to creative work is supposed to belong to the realm of intrinsic motivation, the business of fame turns inequality and hierarchy into subjects for fascination and admiration. These labor markets are built on the most astounding defense of inter-individual competition. Indeed artists' talents and their outputs' quality may be hardly judged by standardized, cardinal measurement devices; as stated by Lazear and Rosen (1981), competitive lotteries are here superior to more familiar compensation schemes.

Yet critical expertise is fallible, a matter that is less harmful for the critic than for the misranked artist [Ginsburgh (2003)]; moreover there are countless cases of collusive maneuvers, critical herd behavior, or even illegal practices such as payola [Coase (1979); Caves (2000)]. Actually the valuation process that produces quality ratings and translates them into rank positions is a noisy process, and the sorting procedure of the best has its long-lived pathologies. Artistic markets also generate a causally reversed rating process due to a cumulative signaling advantage, fairly similar to the Mathew effect observed by Merton in science. Since the market for artistic products and performances is an imperfectly competitive one, problems arise as to how consumers can know and appraise many characteristics of so many widely differentiated goods. Employers have search and information costs, as do consumers. Both may minimize their search costs by



using price or the artist's reputation as an index of quality. An artist's established reputation is less elastic to perceived and appraised quality than the competitive structure with its tournaments and sorting rankings would suggest. The vocabulary of reputation also brings to light the fact that the appraisal of art and artists varies with the organizational traits of each art world, since it reflects the cooperative and competitive activities of the various members of each world. Several dimensions of appraisal exist, of which the spot market value of the outcome is only one. Deferred financial success occurs especially in art markets where the appraisal is initially undertaken by a narrow community of experts and learned consumers, and where a capital of recognition may be accumulated that is eventually converted into an increasing share of demand, which may provide the most famous artists with a slowly increasing flow of earnings [Bourdieu (1992)].

## 5. The excess supply of artists

### 5.1. Causes of excess supply

Sociologists, economists or historians dealing with artistic labor markets have almost always referred to the oversupply of artists. For example, in their analysis of the economics of musical composition in Mozart's Vienna, Baumol and Baumol (1994) explain the "profusion of composers" by the merging of two forces, that of the still powerful imperial patronage and that of the rise of the free market, a process entailing an additive effect on employment opportunities and on the attraction into the musical profession of many "who would otherwise have sought to earn a living elsewhere" (p. 73). In the first half of the nineteenth century the glut of novelists and poets in Paris led to Parisian bohemianism and accounted for the success of the "art for art's sake" ideology, which acted as a compensating device for the subordination of the artist to impersonal market forces [Graña (1964)]. In several other European countries, literary proletariats were similarly spawned by the mid-century publishing boom.

The Impressionists' revolution took place in a Parisian art world whose institutional apparatus – the Academic system – was collapsing under pressure from the greatly expanded number of professional painters. White and White (1965) show how control was lost over the flow of recruits through art schools, the flow of paintings produced, and the careers of the painters; a free market took over to launch innovative artists and movements on a more flexible and also much riskier basis of open competition involving dealers, critics, painters and buyers. Supply was no more to be regulated, hence oversupply became a permanent feature of that market. Furthermore, in Berlin and Munich at the turn of the century the art market was similarly overcrowded with painters competing for recognition and success; periodic panics about the glut and the high rate of unemployment didn't deter students from entering art schools in growing numbers [Lenman (1989)]. A further example comes from the minute study of the music profession in Britain carried out by Cyril Ehrlich (1985) who reported substantial evidence of a glut at the turn of the nineteenth century, at the end of a 60-year period during which

musicians had become one of the fastest growing professional groups; he showed how musicians, aside from lamenting over the damnable flood, tried to react to the pressures of relentless competition and its consequences (very low fees and depressed incomes, underdealing practices, etc.) by establishing professional associations and trade unions, despite increasing segmentation among the workforce.

Summarizing the evidence at hand, [Lenman \(1989\)](#) states that

... the problem of surplus artists was part of a much broader, international cultural phenomenon. Between the 1860s and 1914, for example, steeply rising enrollment in German higher education led to periodic panics about overcrowding and unemployment. In several European countries, literary proletariats were spawned by the mid-century publishing boom. Music and theatre were overflowing with excess labour; in Britain in 1891 there were nearly twice as many musicians as bank clerks, and extreme variations in status and pay. Though the market free-for-all enhanced the importance of dealers, agents, professional organizations and other stabilizing elements, it also created a reserve army of starving music-teachers, hack authors and painters forced into all kinds of low-grade and shady occupations (p. 131).

In each of the cases mentioned above, a similar array of factors is invoked: a rising level of demand (enhanced by factors such as urbanization, increasing educational level, growing incomes, more leisure time, public support), changes in the commercialization of art that bring market principles of organization and bargaining into harmony with the stream of artistic innovations, and technological innovations affecting the transmission and distribution of art. Unlike short-term fluctuations that may be provoked by fads and fashions, long-run shifts causing an increase in private and/or public demand trigger an expansion in training facilities, and more artists appear. But, as Ehrlich shows in the case of musicians, inflexibilities may dramatically hinder the equilibration process if demand turns down, as in the case of the briefly flourishing demand for musicians in cinemas which collapsed with the coming of talkies. Existing practitioners are trapped in a disintegrating market while new aspirants continue to flood in; the training system may play an unintended role in the self-congesting spiral of oversupply, since teaching positions and kindred activities in non-profit art organizations shelter artists from occupational risks.

Innovations in artistic production resulting from the interaction between new techniques, aesthetic shifts and market transformations have often been studied in respect to their impact on labor supply. Some of these innovations tend to lower or to modify the usual skill requirements, and/or the quantity of input factors in the production process, resulting in an increase in artistic productivity, a growing competition among artists and a declining control over entry and professional practice through the traditional devices of the professionalization system.<sup>13</sup> Also technical innovations such as

<sup>13</sup> Among numerous possible examples we may cite the new methods of production of paintings in seventeenth-century Holland [[Montias \(1996\)](#)]; the deskilling process at stake in many avant-garde innova-

motion pictures, radio, television, records and other recent changes increase the extent of scale economies in artistic and entertainment activities [Rosen (1981)]; as the market supply of works and services grows, the scope of each performer's audience gets larger, and more numerous artists are induced to enter the labor market, though some occupational trades and niches of specialization may disappear. Even if this results in a greater concentration of rewards among the most talented who can operate on an international scale, the lure of enormous rewards and the associated social recognition may favor gambling behavior, as success seems like a lottery game in a more speculative market of talents.

Organizational flexibility in the arts plays a major causal role in nurturing steady over-supply. For record companies or book publishers,<sup>14</sup> overproduction of new items, along with the allocation of numerous personnel to boundary-spanning roles and co-option of mass-media gatekeepers, is a rational organizational response to an environment of low capital investments and demand uncertainty, especially in the most speculative and entrepreneurial segments of the market; because of a strategy of differential promotion of the numerous items released, the firm's attention eventually focuses itself on a small proportion of them. Since "nobody knows", too many contestants are induced to enter the success race [Caves (2000)]. The same holds true for employment relationships in the performing arts. Employers in project-based performing arts organizations seek to draw from a large pool of artists and personnel in order to reduce overheads and build efficient and well-matched teams, gaining from the variety of talents and skills at hand.

## 5.2. Agglomeration and congestion

Excess supply of artists is still more than ever evident today. Its spectacular manifestations occur exclusively in the main cities where artists and cultural producers and employers agglomerate [Menger (1993); Scott (2000)]. An effective way to overcome the complexities of a vertically disintegrated and highly flexible production process is indeed to rely on spatial concentration. Especially dense transactional relationships between production units have geographically-sensitive cost structures. The greater the costs per transaction, the greater the probability that firms will agglomerate in order to benefit from external economies of scale [Hall (1970); Storper and Walker (1989); Glaeser (1998); Quingley (1998)]. This leads to an original scheme of competition between the firms. The distinction between short contractual arrangements (at firm level) and employment processes (at industry level) is blurred by the multi-sided activities of each worker as well as by the dense formal or informal relations between employers. Indeed, artistic production is based on three components:

tions in visual arts [Moulin (1992)]; the pop music revolution [Peacock and Weir (1975)]; and the success of dance music [Hesmondhalgh (1996)], which can be partly explained as the result of the widespread availability of production technology, the transformation of the record industry, shifts in authorship and the segmentation of market demand.

<sup>14</sup> As highlighted in Hirsch's pioneering paper [Hirsch (1972)] and by Coser, Kadushin and Powell (1982).

- a nexus of ties between firms involved in the different parts of the production process and between the many employers who draw from the artistic labor pool;
- an original way of processing information through this network in order to minimize the costs and length of sorting and hiring operations; and
- conventional industry-wide negotiations and arrangements regarding wage and reward schemes as well as the mitigation of risky employment prospects.

Employers compete for contracting with the most profitable talents, but they all need to have access to a reserve army of artists; they are of course better off if the major part of the costs of securing pools of employable artists fall on these mechanisms.

For these reasons artistic activities show a very high level of spatial concentration in a few locations or even in one dominant city in each country. A threshold or critical-mass effect exists both on the supply and the demand side, as suggested by Blau's study of the cultural organizations in the largest US cities [Blau (1989)]. The relation between increase in the city population and increase in cultural supply (artists, organizations) is linear in the case of popular culture, but multiplicative for elite arts institutions. In other words, popular culture depends directly on market forces and consumer sovereignty, when the less popular high culture needs a larger pool of potential consumers to develop. It is also remarkable that even in the presence of an actively decentralizing cultural policy, as in France over the last two decades, the concentration of artists and art professionals did not significantly decline. The Parisian case is striking: during the 1980s, the population of artists and professionals involved in cultural production expanded rather rapidly in France (+55 percent between 1982 and 1991) but the share of artists living and working in Paris and the Parisian region also increased (from 45.8 to 54.1 percent).<sup>15</sup>

Artistic supply also has its seasonal congestion peaks. The dramatic example of the 19th Century annual Salon exhibition in Paris has been for a long time a prime symbol of a congestion phenomenon in the arts, with exhibition halls full of paintings from the bottom to the ceiling.<sup>16</sup> Today we see bookstores overwhelmed by crowds of debutant writers' novels, especially at the start of the literary calendar – how often do critics lament over the crazy publishing policy that releases hundreds of such novels over a very short span of time during the literary prizes peak period in Paris. Or take the example of the simultaneous release of many big-budget movies on the same weekends, discussed by Camerer and Lovallo (1999) in their study on overconfidence and excess entry. One may also mention the ever more numerous classical musical contests with queues of contestants trying to win a prize and to attract critical and public attention, although in this case filtering procedures<sup>17</sup> are quite strict, unlike the situations mentioned by Camerer and Lovallo where the criterion for success is more vague and ambiguity permits excess optimism, letting people or firms overcompete.

<sup>15</sup> See further in Menger (1993).

<sup>16</sup> As satirically drawn and painted by Daumier.

<sup>17</sup> Note, however, that these procedures are subject to significant evaluations biases such as the order of appearance at the competition, as studied by Ginsburgh and van Ours (2003).

Excess supply of artists, insofar as it is causally related to the lottery-like structure of occupational prospects in the arts, has been subject to interpretative controversies among economists. Frank and Cook (1995) call into question the way the winner-take-all markets operate, with their damaging features; the payoff structure generates a spiral of individual and social occupational waste, since it leads both to increasing inequalities in monetary and non-monetary rewards, overcrowding in markets, and occupations prone to an overestimation of one's chance of success. As a result when excess numbers of contestants are induced to invest in performance enhancement in order to raise their individual odds of winning, these investments will be mutually offsetting and socially inefficient; end consumers may get more valuable products but the social costs are excessive. By contrast, Cowen (2000) argues that the superstar effect is welfare-improving (consumers get better performances) even if it leads to increased income inequality, but that it should not be overstressed. Indeed, fame is a positive-sum game, not a negative nor a zero-sum one. Instead of an unambiguously increasing concentration of rewards, Cowen states that countervailing forces operate such as a convergence of quality that limits the ability of the very best stars to dominate the market for long, or more radically the elastic supply of fame, so that when demand for fame increases, the number of prizes, rewards etc. rises too.

### 5.3. *Monopolistic competition*

Common traits of monopolistic competition include high product differentiation (works and performances), a large variation in consumer preferences and excess capacity of production [Lancaster (1979)]. The application of a monopolistic competition model to the market for art works is well known. In regard to artists, a temporary monopolistic position in the market may be provided by an artist's reputation, as long as her skills and talents are in demand. As cultural industries develop, several markets get related and monopolistic return to reputation may be increased according to different pricing schemes. A striking example is provided by Krueger's (2005) study of the market for rock concerts. Uniqueness of sound and style of a rock band, when successfully meeting the consumer demand, provides it with a monopoly power. Until the late 1990s, famous rock bands were able to exploit the two complementary markets of concerts and records. Krueger assumes that when greater concert attendance correlates with greater artists' record sales, artists may be induced to price their tickets below the profit-maximizing price for concerts alone. When new technology allowed many potential customers to download music, the link between the two markets weakened. Krueger's data and statistical study show a sharp increase in the average concert ticket prices and in price dispersion from 1996 to 2003; star bands were able to compensate for their income loss from declining record sales by maximizing their monopoly profit from concert attendance.

In theory, the supply of artistic talent is infinitely differentiated; every artist claims to be endowed with unique skills and to supply original achievements. One may therefore speculate about generalizing the monopolistic competition model to the entire work-

force of creative and performing artists. The level of reputation, the amount of work, and the degree of occupational success vary considerably among artists; shall we consider as monopolistic suppliers of their own work not only those artists who face a rather inelastic demand curve due to their temporary or established fame, but also those in the lower segments of each artistic occupation who are induced to act as monopolistic suppliers (i.e. to trade their human capital and work as freelancers) although their market value and market power are very poor? In the latter case, young aspiring artists, or moderately or poorly successful ones, may claim to be endowed with a unique set of skills, talents and abilities, yet do not benefit from the rent attached to its seemingly monopolistic supply.

Consider also creative artists whose compositions, paintings, manuscripts and screenplays are put up for sale by gallerists, publishers, talent agencies and so on. Because most of them are self-employed, it would seem meaningless simply to equate fewer working hours with unemployment spells or underemployment levels. Their income does not derive from a quantity of working time at a given wage rate [Frey and Pommerehne (1989)]. Creative artists decide whether or not to continue to work in their chosen field according to their income and to the stream of their expected earnings. If their income is low, because of low demand for their work, a simple increase in production through more work may have no effect and an increasing supply of the works for sale at lower prices may not trigger an equilibration process, since the price acts as a signal of quality and a decrease in the pricing of a contemporary artist will promptly be interpreted negatively. Since these artists can make their own work opportunities, oversupply of the works they produce cannot be defined at any given price; this explains why so many creative artists, though working hard and being fully committed, may suffer from low or very low income levels, and develop a sense of null or even negative correlation between effort and earnings in their vocational trade.<sup>18</sup>

Caves (2000) views creative products as a mixture of vertical and horizontal differentiation. This may help solving the intriguing issue of the competitive nature of artistic labor markets. Horizontal differentiation stems from the fact that artists and their work differ from one another in many ways; vertical differentiation refers to the rankings of artists according to their skills and level of talent, and to the quality and originality of their products and performances. The mixture of both types of differentiation may blur the ranking process and generate evaluations that are disputable and volatile. Would excess supply be less pathological in the arts if artists worked more uniformly like scientists? In the sciences, the winner-take-all structure of competition takes an extreme form wherever the priority race in discovery prevails. There is only one winner in each race, and inequality with regard to scientific productivity and to the awarding of priority is considerable and increases over the careers of a cohort of scientists, due

<sup>18</sup> As reported by Moulin (1992) in her extensive study on visual artists.

at least to some state dependency and cumulative advantage process.<sup>19</sup> The priority system obviously creates an excessive attraction to scientists in certain extremely well-rewarded contests. Note however that this portrayal of scientific competition mirrors the vertical differentiation process of creative destruction of knowledge when innovation proceeds along predictable paths of accumulation. Yet the creative game in science also has a serendipity dimension and benefits from hybridization of different specialized knowledge. Since every scientific field grows by splitting itself into increasingly fine-grained subspecialties where the intensity of competition varies according to the prominence and reward-generating power of each field, room is left for more horizontal differentiation and its subsequent monopolistic competition aspects as well as for cross-disciplinary innovations.

Vertical differentiation in the arts goes closer to that found in the sciences whenever a dialectic process of discovery, obsolescence, destructive creation of styles and languages takes place, especially in the fine arts and within periods of time strongly committed to the *avant-garde*. Under such conditions, priority becomes a key value and generates a strong opposition between an elite group of innovators, groups of followers, and crowds of more or less instantly displaced artists producing old-fashioned art whose market shrinks and eventually vanishes. However it is well known that such a priority-driven competition scheme fits only parts of the artistic production landscape. Caves' analysis very aptly captures the complexities of the valuation process within a context of mixed production:

The process of distinguishing significant innovation from everyday creativity varies among creative activities due to the filtering porosity of their filters. In a creative activity with tight and clearly articulated standards of performance, critical presumption is loaded against the acceptance of novelty as a valid and desirable innovation. An innovation must either carry the credential of manifest face value, or eke out slow victory in localized skirmishes between novel and traditionally accepted creative goods. On the other hand, an art realm that welcomes any novelty as a noteworthy innovation necessarily lacks consensus on any critical paradigm. Critical rankings lose their value for consumers to calibrate and rationally order their selections among creative wares. A list/B list rankings are impaired for contributing to the efficient organization of complex creative activities. It is the old issue of liberty versus license. A creative activity that yields the most value for its participants as a group will need a suitable compromise between tight critical standards that resist those innovations that will ultimately overrun the "establishment", and loose standards that yield no stable valuations or points of reference to guide either artists' training or consumers' investments in cultural consumption capital [Caves (2000, pp. 202–203)].

<sup>19</sup> See Stephan (1996) for a review of the literature.



## 6. Remedies to excess supply? Creative prospects and chances of self-actualization

Things would be simple if artists could form correct expectations about their chances of success or at least about their odds of decent living within the occupational sphere they choose to enter. Competition would seem to be less wasteful, failures and occupation switching less frequent if not marginal, misallocation of talents due to excessive lure of stardom or of self-achievement promises wouldn't hamper the development of other occupational worlds that might be short of such diverted abilities, training systems wouldn't favor wasted investments, risky occupational prospects with their possible generous unemployment insurance wouldn't claim increasing public support at the expense of other economic sectors, and competition might gain in fairness, since artists would have enough time to prove themselves.

Such an argument has a strong and a weak version. The *strong* version is that of an optimal allocation of talents according to an "optimal division of labor" scheme. Such a world can be found in functionalist models of society originating in early Durkheimian sociology, and also in some economic models of welfare analysis of work such as Lancaster's model of optimal division of labor which matches people to the occupation where their abilities and skills are optimally employed. Required are: a set of well-defined standards of performance and creative content on which critical evaluation and competition may be based; the optimization of the educational and training system that detects abilities and provides individuals with the best-fitting skills; and the use of optimal reward schemes that deter people from choosing what happens to their first-best occupational fate.<sup>20</sup> This view could hardly apply to art. In its sheer essence, art has been celebrated and valued as the symbol of creative, innovative and non-routine work. Creative and non-routine work provides psychic and social gratification proportional to the degree of uncertainty of success. The more the work is non-routine, the less one can be certain about the immediate or long-term chances of individual achievement. To be sure, artistic work also entails routine aspects both in relative and absolute terms.<sup>21</sup> Yet as the non-routine dimension points to the challenging and intrinsically rewarding side of work, it gives artistic creative work its differential value. Performance in non-routine activities does not depend on skills that could be easily objectified, transmitted and certified in the training system. Indeed, the impact of schooling on earnings is typically

<sup>20</sup> The two Lancastrian propositions in his welfare economics of work are, first, the optimal match proposition: "Among all the allocations of persons to jobs which satisfy the requirement that every person holding a job has the skills needed to perform it, there is an optimal allocation", and second, the optimal division of labor proposition: "If the number and specification of jobs can be varied, there is an output-maximizing division of labor which gives the greatest output from the skills available in the population and an optimal division of labor which gives the greatest overall welfare from work and consumption. The output-maximizing and optimal divisions of labor need not be the same" [Lancaster (1979, pp. 326–327)].

<sup>21</sup> Various artistic occupations and individual achievements may be ranked according to how routine or non-routine the work is, and no artist could every time reconstruct afresh his own frame of activity and no collective work could be achieved if conventions didn't exist as stabilizing forces [Becker (1982)].



smaller for artists than it is either for all workers or for managers, professionals and technical workers [Filer (1986, 1990)]. Insofar as non-routine activity refers to a wide range of changing and challenging work situations, it therefore implies that abilities may be revealed and skills acquired only progressively through a process of learning-by-doing which is highly informative and which cannot be perfectly anticipated *ab initio*. Even if one were to assume that innate abilities command success much more than formal training, talent could express itself only by coping with work situations that reveal the multiple characteristics of what artistic achievement really is.

A *weak* form of the optimal division of labor argument is the optimal trial-and-error process. If talent can be detected more rapidly, then quit rates in artistic professions will be much higher and turnover rates will help to form more realistic expectations about one's chances. And if artists behave rationally, expected risky occupational outcomes should be experienced in a way quite similar to that predicted by the theory of option pricing in finance. An optimal sequential decision scheme orders occupational alternatives with respect to risk; it is rational to choose the job with the greater risk first and to switch to a less risky alternative if the outcome turns out to be unfavorable. In the spirit of Johnson (1978) and Jovanovic (1979), MacDonald (1988) applies his theory of job assignment and sequential accumulation of information [MacDonald (1982)] to artistic careers; he conceives of the stochastic dynamic superstars model as a model of occupational choice where performers accumulate information on the likely outcome of a subsequent performance and learn to rate their job match.

Indeed, in a realistic dynamic occupational choice model, informational considerations are brought in, as in the job-matching approach of occupational choice [Miller (1984)] that seems to fit rather well with the results of surveys on the careers of freelancers. Job applicants learn only gradually how well they are suited for a particular artistic occupation and to what extent they can expect to meet success in it. It is a costly and rewarding "trial and error" process; one becomes more and more informed about the various facets of the occupation and about one's own abilities through doing the job. Workers accumulate skills through experience and learning-by-doing. As non-routine work implies a steady human capital investment, it takes place in a matching process where jobs are "tied packages of work and learning" [Rosen (1986)] and are ranked along their varying learning potentials, as shown in Faulkner's (1983) research on freelance composers in Hollywood. The attractiveness of artistic jobs can therefore partly derive from their high learning potential, at least as long as the work is non-routine enough. In fact, many artistic occupations provide this kind of information only through a learning-by-doing process, either because formal training is not strictly required to enter the professional community and to succeed (in some artistic occupations like that of writer, formal training seldom exists), or because formal training doesn't act as an efficient means for selecting talents and for screening abilities. This is probably why so many artists think of themselves as self-taught, even in occupations where formal training plays a genuine role. For example, most actors, while satisfied with the technical aspects of their training, are nonetheless critical of the lack of preparation; information about one's abilities and chances of successful professionalization is mainly

acquired in the course of practicing [Jackson et al. (1994); Menger (1997)]. Yet the learning and information acquisition process is costly; jobs where one can benefit more from learning-by-doing are on average less well-paid initially than jobs where applicants can be selected on the basis of university degrees or through other immediate skill certifications.

However, application of this job-matching model to artistic occupations raises two issues. First, it may be asked how much information one needs before being able to assess the quality of one's job match, considering that occupational practice acquires so many different forms, and takes place in so many environments and in relation with many diverse employers and patrons. In addition, this high variability in practice probably influences the artist's behavior regarding risk-taking. In some respects, each work experience in the performing arts such as theater or movie production is unique, and each team of artists and technicians is different; there seems to be no end to the learning process and to the assessment of one's talent and no situation seems really crucial when one has to decide how far to go in such a career. This could explain why many artists maintain for so long the hope that they will eventually become famous, even after death. Romantic writers and poets invented a psychological and ideological device for fighting against short-term disenchantment: the "loser is eventually the winner" game [Sartre (1971); Bénichou (1985)], which designs the consolatory counterpart – the "loser-take-all" society – of the world depicted by Frank and Cook (1995). More generally, under a highly flexible working scheme the explanation of oversupply by the "risking-and-learning" model is especially appealing. Where information about the quality of the individual occupational match is delivered only through on-the-job experiences which are subject to contingency and discontinuity, aspirants are not screened at entry; experiencing the more or less unbounded autonomy of creative work may raise the non-monetary valuation of work and lower the opportunity costs of the choice to stay in the trade, even if facing low success. This helps explaining why barriers at the exit may rise although dropping at the entry.

Second, once multiple jobholding is taken into account, risk diversification considerations may suggest an enlarged definition of occupational choice, where several related jobs provide switching opportunities that may result in a cycling pattern between various kindred activities instead of building irreversible sequences of choices. In this case, an interesting way to test the assumption that, against the standard economic view, workers may derive satisfaction from the process of work itself and not just from the income it earns, is to study whether artists turn down better-paid jobs in order to pursue their vocational work. In estimating labor supply functions for Australian artists with arts and non-arts wage rates as explanatory variables, Throsby (1992) showed that artists supply the non-arts labor market only up to the point where an adequate return is received to support their primary artistic work. Similar results have been obtained by Alper and Galligan (1999). It follows that the notion of oversupply due to labor rationing turns out to be questionable [Killingsworth (1983)] when the standard model of conventional income and substitution effects on leisure has little relevance and when the causal link between oversupply and rationing refers to a disequilibrium in only one of the labor mar-

kets the artists supply, that of their principal vocational work. When multiple jobholders cycle between rationed and less- or unconstrained job markets, underemployment or oversupply may be hard to specify, provided that under such a steady “management of risk” scheme, work remains more attractive than occupational alternatives outside the arts sphere.

Another view of the “excess supply” disease, quite the opposite of the functionalist one, assumes that competitive markets not only magnify and exploit unequal individual endowments of artistic ability and talent, but that the market organization of the arts neglects to recognize the sheer potential of creativity of everyone. In that respect it fails to implement the optimal level of innovation and creativity, both from an individual and from a social standpoint. This is the Marxian view of the self-achievement ideal which could be implemented in a society that overcame the alienating division of labor organization of work. Such a conception has its roots in a powerful tradition, the expressivity model of praxis that can be traced back at least to Hegelian philosophy [Habermas (1988)], if not to the Aristotelian conception of action. In this model, self-actualization through creative work entails a basic distinction between labor as a routine and alienating activity, and work as a non-routine pursuit. This distinction plays a major role in Arendt’s (1959) theory of work as a non-utilitarian kind of lasting human achievement, as well as in Freidson’s (1986a, 1986b, 1990) view that artistic professions present a challenge to conventional conceptions about vocation and labor. Hirschman’s (1986) classification of different kinds of work in terms of the varying predictability of their intended outcome brings to light the non-instrumental nature of the artist’s effort; in the uncertain course of creative action, the strenuous overcoming of obstacles takes place through alternations of tension and the anticipated savoring of the future result. Therefore self-actualization through work, which makes artistic activity so attractive, occurs only if the outcome is unpredictable enough; the possibilities of personal invention are wide open, and at the same time the artist is never sure that she will express herself in her work as she expected to.

Clearly the expressivity model catches the link between creativity and unpredictability of work, but it suffers from an internal contradiction when it precludes the possibility of failure, as in the Marxian conception of self-actualization. In Marx’s view, everyone is endowed with the same abilities; the considerable variation in artists’ reputations and incomes is mainly ascribed to strongly biased consumer preferences shaped by market forces rather than to talent differences. In fact, artists quite commonly conceive of softer market competition as a means of favoring self-actualization of everyone’s creative endowments. A non-rationed labor market would require either sufficient homogeneity on the supply side or a quality inelasticity of demand high enough so that the substitutability of artists and of works would allow for market equilibrium. Yet both supply and demand of creativity are based on inter-individual comparison and competition.

The risk of failure is a built-in characteristic of artistic undertakings. Moreover, failure or success does not merely depend on the creators’ own appraisal of their work, unless their art world forms a community of producers who have no interest in others’ production or in anyone’s consumption [Elster (1985)]. Individuation through creative

work, which greatly accounts for the admiration of artists, requires that others have an interest in one's work [Cohen (2000)] and consequently that some competitive comparison occurs. True, artistic individualism has been recognized as a sign of a trend towards autonomization of the artistic sphere and towards professionalization of its members, according to Max Weber's theory. However, artistic individualism could hardly be equated with an intrinsic, competition-free striving towards self-expression and self-actualization. Thus individualism, apart from characterizing a lifestyle and referring to a loosely structured occupational community, may signal the tension between a strong sense of personal achievement experienced in absolute terms, and the way one's creative work unavoidably involves relative comparison with others. As experimentally shown by Camerer and Lovallo, overconfidence and optimistic excess entry in a business may be due to the fact that people neglect the reference group of competitors, each one estimating they are skilled enough to succeed; "neglecting the increased level of competition is like the neglect of adverse selection which leads to the 'winner's curse' in bidding" [Camerer and Lovallo (1999, p. 307)]. Relative skill perception may entail miscalculation of one's chances especially when the skill requirements are underspecified, when the performance feedback needed to adjust one's level of aspiration is fairly noisy, and when the employment system magnifies heterogeneity among the workforce.

This highlights the complexities of uncertainty as a double-sided incentive. One side is the non-routine and non-instrumental aspect of work that makes it attractive; since work experiences have to be constantly renewed in order to be attractive, a strong sense of challenge, and of competition against oneself, nurtures the artist's quest for individual achievement. On the other side, given that the production of creative work has been increasingly market-driven since the nineteenth century, the quest for systematic originality and innovation brings to light the strategic dimension of uncertainty, that of inter-individual competition. Ironically enough, the non-monetary value of expressive self-actualization through creative work is the genuine fuel of market competition.

Uncertainty management techniques and learning processes related to the core of artistic invention on one hand, and to the market structure of competition among widely differentiated talents and products on the other, provide economic research with insights into the behavioral type of the artist. The artist may be portrayed neither as a conventional rational actor well-equipped to survive in an ever more competitive market, nor as a myopic one induced to take occupational risks only because she forms probabilistic miscalculations of her chances of success or because she was programmed by her initial socialization to enter an artistic occupation. Rather, she may be portrayed as an imperfect Bayesian actor gathering information; learning by doing; revising her skills, expectations and conception of herself; building networks in order to widen her range of experiences; and acting without knowing her initial endowment of ability and talent or what she may be able to express over the course of her loosely patterned career. Insofar as she acts as a monopolistic supplier, the artist tries to expand the control over her own work and over the market of the goods or services she provides. However this outward-oriented goal, driven by the competitive pressure in the market for the arts and entertainment services, would be meaningless were it not matching the inward-oriented

goal of self-discovery and self-actualization, a goal that may be pursued only as long as the variety of work experiences and challenges is optimal and if the balance between invention, security at work and temporary routine exploitation of innovation is secured.

## 7. Managing the risks of the trade

Studies of artistic occupations show how artists can be induced to face the constraints of a rationed labor market and how they learn to manage risky careers. Pioneering empirical research by [Baumol and Bowen \(1966\)](#) found that artists may improve their economic situation in three main ways which are not incompatible and may be combined: artists can be supported by private sources (working spouse, family or friends) or by public sources (subsidies, grants and commissions from the state, sponsorship from foundations or corporations, and other transfer income from social and unemployment insurance); they can work in cooperative-like associations by pooling and sharing their income and by designing a sort of mutual insurance scheme; and finally they can hold multiple jobs. Most studies, both in sociology and in economics, have focused on this last means, since apart from being widespread and becoming more so, it brings into light a puzzling feature of the artistic labor market: that of the diversification of risk through one's own human capital and labor, which seems a much more unusual phenomenon than risk management in the financial sphere. In fact, it makes artists resemble entrepreneurs since, just as property owners spread their risk by putting bits of their property into a large number of concerns, multiple jobholders put bits of their efforts into different jobs [[Drèze \(1979\)](#)].

Multiple jobholding shows a general upward trend, and artistic workers in many countries rank among the highest in the percentage of all workers who have secondary jobs; in addition, artistic occupations rank at the top in the percentage of all jobs held as secondary jobs. In the US, for example, almost every artistic occupation appears among the 25 occupations employing the largest proportions of their workers through a secondary job [[Amirault \(1997\)](#)]. In their survey of 3000 New England artists, [Wassall, Alper and Davison \(1983\)](#) found that only 24 percent of artists did not hold a non-artistic job. In their 2000 study on moonlighting in the arts, [Alper and Wassall \(2000\)](#) provide an extensive review of the multiple-jobholding behavior of American artists from 1970 to 1998 based on the monthly Current Population Survey date files. As their calculations show, rates of moonlighting by artists have increased over the years much faster than in the overall labor force or than in the category of professional workers whose personal characteristics are closer to those of artists; performing artists experience the highest rates of moonlighting; educational level is the only personal factor strongly positively related to multiple-jobholding behavior.

As shown by [Throsby \(1992, 1994a, 1996\)](#) in his studies on artists' income and labor supply, not only must economic studies recognize the arts/non-arts earnings distinction as providing a more complete picture of artists' income sources, but that simple

dichotomy in itself does not go far enough. In order to capture the full range of relationships between labor supply and earnings experienced by artists, a three-way division of working time and earnings is essential<sup>22</sup> between:

- the creative activity itself, which corresponds to the primary creative labor and the tasks associated to the preparation of the artistic product (thinking, dreaming, searching for materials, rehearsing, practicing);
- arts-related work, which includes the various activities within the particular art-world that do not contribute directly to producing the artistic product, but still rely on the skills and qualifications possessed by the professional artist – common examples of such work are teaching activities and management tasks in artistic organizations; and
- non-arts work, which may differ considerably among individuals, among artforms and over the individual life-cycle in an artistic career; for example, recent US Census and survey data report that while a majority of authors (as primary occupation) hold secondary jobs in other professional occupations and especially in educational fields, actors' and singers' secondary jobs are mainly in sales, clerical or service jobs, i.e. jobs with a history of low pay and poor benefits [Alper et al. (1996)].

The range of various jobs may be compared to a portfolio of financial assets.<sup>23</sup> This way of handling uncertainty has already been evoked above in the case of freelancers, who may insure themselves against downswings on the employer side as well as strengthen their position by building a career portfolio that is mixed with tightly and loosely coupled work associations. With sectoral diversification of hirings, artists may also be financially better off and have greater career continuity in a disintegrated labor market. Holding other jobs outside one's vocational field of activity corresponds to a better known scheme of occupational risk diversification. Instead of thinking statically in the terms of the old dilemma – freedom or alienation – the portfolio model of occupational risk management offers insights for the dynamic study of how artists cope with uncertainty throughout their careers and allows us to maintain the centrality of choice of career path.

Sources of income are much more dispersed at the beginning of an artist's professional life, coming under greater control when the artist's reputation grows and when his ability to select among different opportunities allows him to reach a more careful balance between constraints and fulfilling commitments. However, in focusing on the combination of insecure and secure sources of income, the "diversification of risk" approach fails to deal with the characteristics of different kinds of work, assuming that a secondary job doesn't provide the artist with anything else except income. As a result another complementary dimension of multiple job-holding is overshadowed, concerning the relationship between creative work and related artistic work, as described in

<sup>22</sup> See Throsby (1996), Menger (1997), Paradeise (1998).

<sup>23</sup> Faulkner (1983) and Menger (1989); for a more committed, prophesying view on the rise and desirable future of portfolio work in our society, see Handy (1989).

the “role versatility” scheme [Nash (1955)]. In certain art worlds such as that of “serious” music, high technical skill requirements act as a selective barrier to entry as well as an integrating device among the professionals employed in the various occupational roles (composer, performer, conductor, publisher and so forth) whose differentiation has increased with the professionalization process. Through role versatility, the composer may reduce the financial risk in his creative activity but also extend his control over the distribution process of his music, facilitate his interaction and communication with the other roles, and increase his prestige among his peers. Roles simultaneously or successively played are thought of in terms of positions in various spheres, as in Abbott and Hrycak’s (1990) study on eighteenth century German composers, or as in Baker and Faulkner’s (1991) study which examines the shifting combinatorial patterns in Hollywood filmmaking and sees roles (e.g., producer/director/screenwriter) as resources to enact positions in evolving organizational settings. More generally, organizational or aesthetic innovations induce role combinations and hybridizations and transform both the content of cooperative activities and the extent of control over new market resources.<sup>24</sup>

Wherever practice needs a specific training, the center of the artistic role constellation is traditionally the teaching role, the most frequent “pool” profession [Abbott (1988)] or “host occupation” [Freidson (1986b)] for creative artists. This teaching position in the arts has been compared by Baumol and by Freidson to the role of teaching in academic life, which hosts and supports research activities; this might explain why creative artists so often consider themselves researchers. The paradox of artists whose educational profile as a group is close to that of managerial and professional occupational categories but has far less impact on their earnings can also be solved. Throsby (1996) shows that relationships between arts income and art training may be strong for arts-related activities such as teaching whereas income from primary creative practice is more influenced by on-the-job experience. Human capital and role versatility considerations militate for arts-related rather than non-arts jobs, and portfolio choice considerations tend to favor supplementary jobs that are stable and salaried, such as teaching. White (1993) suggests that the artist as teacher combines two opposing forms of career, one (teacher) that represents the image of the traditional career since it entails seniority and some order and sense of cumulation from training, and another (the artist as genius) that is built on originality and conveys a sense of destructive creation. That paradoxical role combination is especially striking in avant-garde music [Menger (1983)] and in the visual arts [Moulin (1992)].

<sup>24</sup> See Moulin (1992) on the case of the entrepreneurial artists who work as performers and producers of services in the contemporary visual art market; Christopherson (1996) on the emergence of entrepreneurial filmmakers whose managerial skills blur the lines between management and labor; Kealy (1979) on the emergence of the hybrid ‘artist mixer’ in rock music; and Hesmondhalgh (1996) on the entrepreneurial strategies of sound mixers and DJs in the dance music record sector.



## 8. Collective action and public support in occupational risk management

Risky employment prospects in the arts may also be mitigated via co-operative and collective action [Peacock and Weir (1975)]. Artists may share the occupational risk by pooling their resources, as in the case of groups of visual artists [Simpson (1981)] who provide each of their members with mutual support, or the main symphony orchestras in London which operate on a self-managed organizational basis, with musicians being shareholders of their own company and augmenting that position with freelance hirings elsewhere [Peacock (1970)]. Most small organizations in the live performing arts (dance companies, chamber orchestras, baroque and contemporary music ensemble, etc.) work on this co-operative basis.

Studies on the collective action of unions in the arts are far fewer than those devoted to state and public support for the arts.<sup>25</sup> One common feature of unions' action concerns income transfers and redistributions that may allow workers to adapt to more flexible and more unbalanced artistic labor markets. Apart from traditional direct and indirect forms of public support to artists that are prevalent in European countries [Mitchell (1992)] and mainly intended for self-employed creative artists, collective action regarding artistic labor markets relates to the funding of non-profit organizations such as performing companies employing artists, as well as dealing with the impact of increasing flexibility. Paul and Kleingartner (1994) show that in the US Film and TV industries, the actors', writers' and directors' unions, unlike craft unions, have expanded in spite of the introduction of highly flexible production. A three-tier compensation allows artists both to be covered on an egalitarian basis (through minimum pay rates), to allow those whose market value exceeds union scale to negotiate additional compensation, and to get additional payments (residuals) for the re-use of the films and TV programs to which they have contributed. This last device can hardly be underestimated, although the importance of residuals to total compensation varies greatly among labor market segments. In 1988, according to the American Screen Actors Guild statistics (quoted in Paul and Kleingartner), the total residual compensation from all markets was almost equal to total initial compensation. Not only a compensation mechanism, the residual's role is also that of softening the impact of work contingency and frequent periods of unemployment by generating a passive income stream. As film and audiovisual markets expand and flexibility increases, residuals have become the focus of individual contract negotiations and collective labor relation bargainings in that sector.

In many countries access to unemployment insurance compensation is beyond the reach of freelancers who are numerous in artistic professions. In some countries however, freelancing can be equated with a wage-earning position and is eligible to unemployment compensation. Where benefits are paid by the state, they may be used as a deliberate means of public policy in support of the arts. When unemployment insurance expenses come under a self-administered fund aimed at compensating every worker in

<sup>25</sup> See Gray and Seeber (1996).



the economy for her unemployment spells, a cross-subsidization between sectors may play the supporting role. For those freelancing artists and cultural workers who are eligible for unemployment benefits, the resulting combination of security and of autonomy at work may perform two different functions:

- that of providing earnings replacements which reduce the compensating pay differential associated with the risk of unemployment and the uncertainty about lifetime earnings – in fact the position on the contingent labor market may be optimized so that each individual permanently combines fees and unemployment benefits; and
- that of subsidizing non-working time which can be used as leisure time, or as training time for a future demanding job, or as a searching spell for new jobs.

In the former function, unemployment is seen as a constraint on individual behavior via a labor-demand explanation, whereas in a labor-supply explanation unemployment can be interpreted as the outcome of a worker's choice with regard to job search. Ambiguity also arises from the way uncertainty itself is interpreted; as observed by Drèze (1979, p. 349) "in the case of the self-employed, the distinction between endogenous and exogenous economic uncertainties is not always clear-cut". Consequently, insurance against career failures does raise questions.

The French unemployment insurance system put in place for artists and craft workers in the performing arts provides a striking illustration of these interwoven functions [Menger and Gurgand (1996)]. A generous compensation scheme had been designed to fit the requirements of contractual flexibility but it ran into financial problems because compensated unemployment grew more rapidly than paid work. Work has indeed been allocated quite exclusively in the form of contingent jobs and short-term hirings which typically spread the available work among a growing number of agents. Employers have been able to hire personnel at lower cost, to sort out the most talented, to build well-matched teams and to draw at will from a considerable reserve army of underemployed workers. Thus the performing arts sector did expand by having an increasing part of the income required to attract workers paid through unemployment insurance allowances. In aggregate, the amount of unemployment insurance benefits paid to those workers today in France represents more than two-thirds of their total amount of wages and fees [Menger (2005)].

Actually substantial moral hazard is pervasive, induced by the insurer's inability to distinguish unpredictable exogenous constraints on the hours intermittent workers are able to sell in the market from the worker's choice with regard to job search or allocation of non-market time. In their survey on unemployment insurance issues, Topel and Welch (1980) noted that "to the extent that workers take future unemployment benefits into account when evaluating a job offer, this effect must be ambiguous. While workers will certainly be more selective with respect to job offers if benefits are increased, the value of any particular job must be comprised of both income from working and benefit income from contingent unemployment. The increase in benefits will allow firms to offer the same value of an employment contract with a lower wage" (p. 354).<sup>26</sup> Much

<sup>26</sup> See also Atkinson and Micklewright (1991).

less ambiguous are the effects on employers, who are able to exploit asymmetrical information about their work and job allocation agenda in order to include entitlement to unemployment insurance benefits in their wage bargaining with their contingent employees. Firms may also collude with their employees by hiring them repeatedly for short periods in order to secure a kind of internal labor market without bearing the full cost of long-term relationships. It seems quite obvious that the implementation of an experience rating formula under which an employer's unemployment insurance tax rate depends upon the stability of employment he provides is the only way to make employers responsible for the impact of their hiring decisions on the fund's finances.<sup>27</sup> Yet, in so doing, the state and the local authorities would be asked to provide the non-profit performing arts sector with an extra amount of subsidies heretofore passed on to the insurance fund, and to acknowledge the actual supporting role that it plays [Menger (2005)].

Public support for the arts has increasingly supplied additional means of income diversification and career enhancement opportunities to artists by funding artistic and art-related jobs, by financing or securing compensation systems like those described above, by enforcing property rights and by adapting them to new technological and market conditions. However, cultural policies are at odds with the way firms and entrepreneurs take advantage of the attractiveness of artistic occupations and of mistaken expectations. Increasing flexibility can be associated with higher rates of artistic innovation or at least with increasing differentiation in production. Yet it transfers more and more of the business risk down onto the artists. Cultural policies cannot disentangle the individual risks (those of discontinuous employment prospects and uncertain course of a freelance career) and the social risks (those of having innovations underrated or overlooked and of experiencing a suboptimal cultural development). Moreover, cultural policies and non-profit organizations have mostly developed while the contemporary artistic scene has become more contestable. The valuation process is subject to more volatility, leaving more room for speculative bets and for joint action by several categories of actors (curators, critics, dealers, experts and boundary spanners of different sorts) to promote artistic movements, innovations and fashions. Therefore public support needs to be driven by the same uncertainty principle that underlies the market competition for successful innovation.

Let us consider the ever-increasing number of pieces of art and culture that are consecrated and offered for public admiration in museums, concert programs, books, and audiovisual or computerized archives. These pieces act as permanent reminders; they have emerged from a large stock of works whose significance needed time to be correctly appraised and sorted out, after several rounds of celebrity tournaments and valuation proofs. This process legitimates a transfer of the title and merit of celebrated

<sup>27</sup> As it is well known, without an experience rating scheme, the structure of the unemployment insurance financing subsidizes the benefit payments of some industries at the expense of others, provided that the aggregate unemployment insurance system does balance.

artists of the past onto their contemporary heirs, whether the latter are known or unknown at this time. The definition of art as a public “merit-good” [Musgrave (1959); Netzer (1978)] may catch that mix of elitism and democratization of genius, i.e. the contrast between the dramatic super selection of a few hundreds of world famous masters in each art world and the crowds of candidates to fame which feel entitled to ask for accountable decisions of public support [Menger (2003)].

Uncertainty again plays a double-sided role here. According to DiMaggio (1986), the uncertainty principle which is at the core of the evaluation of any work bears on public and collective choices, both from an intra- and intergenerational point of view. In the long term, uncertainty regarding an artist’s or an artwork’s value, as it vanishes over time, turns into an extremely skewed distribution of fame and success. Yet in the short term, any public support to the arts has to balance efficiency against equity considerations, both by giving the best-rated artists the largest opportunities to develop their talent according to a market competition structure, and by mitigating the market proclivities by sheltering potential but uncertain talents from the “winner-take-all” market structure effects. Thus the uncertainty principle provides a true rationale for the public support of a large number and variety of artists. It can be claimed that it is in the interests of society at large to nurture an oversupply of artists so as to have the best possible choice of talented artists. Indeed, as pointed out by Nisbett and Ross (1980), people sometimes may require overly optimistic subjective probabilities to goad them into effective action. The social benefits of probabilistic mistakes are supposed to be great enough even when the individual cost for an unsuccessful career is high. However, the claim for such positive externalities, though belonging to the rationales of public support to the arts, may trigger an endogenously unbalanced growth.

What kind of uncertainty is it that in the short term has to be managed through insurance devices? Is it exogenous or endogenous? Should a lack of jobs and an unsuccessful career be attributed to insufficient ability? Or are they due to insufficient demand for the kind of ability with which the artist is endowed? The answer lies in the fact that ability and talent themselves should be considered not only as an exogenous factor of market success but also as an endogenous factor shaped by competition through innovation. The more competition raises the rate of innovation or at least of differentiation between prototype-like works by exploiting and stimulating consumer demand for novelty, the more the sorting mechanism will be based on shifting specifications of marketable talent.

## **9. Art as a model for creativity-enhanced work in advanced societies?**

Both capitalism and anti-capitalism celebrate innovation, knowledge, learning and intrinsic motivation as the most powerful leverage to ensure growth and to overcome the alienating dimensions of labor division and routine careers. Marx saw the artist as possibly the highest embodiment of his disalienated worker. Today creativity is one of the key managerial requests to raise productivity levels through changes in work organization.

The economic theory of endogenous growth assigns a central role to idea generation, creativity and knowledge [Romer (1990); Aghion and Howitt (1998); Baumol (2002)]; firms and markets have to know how to draw from what is supposed to be the ultimate inexhaustible source of growth, human creativity, both in its specialized form (that of professional work in the creative activities of workers such as scientists, researchers, engineers, artists) and in its more mundane manifestations, that of everybody's intelligent behavior at work. The creative knowledge principle applies to work organization and management as well; according to these premises, art as a most celebrated realm of creative work should provide managers and workers with some insights as to what creative behavior and intrinsically motivated commitment to one's work can be, once the rhetoric of radical idiosyncrasy, irreducible originality and the undecipherable secret of creation in art has been set aside. Indeed, artists voluntarily supply the golden legend of creation, that of a subversive, anti-conformist, inspired behavior, rebelling against social conventions and commercial utilitarianism; in fact, however, they evolve daily within the economic settings most compliant to the demands of modern capitalism – extreme flexibility, autonomy, tolerance of inequality, innovative forms of teamwork. So the artist and the post-taylorist worker may be able to merge into the same figure, that of a creative professional. How far does that assertion fit the actual transformations of work organization?

The usual picture of organizational design and human resource management (HRM) fits the labor market segmentation theory by distinguishing between:

- secondary labor markets where flexibility, worker substitutability, skill transferability and fixed cost minimization through low-wage policy for low-skilled jobs are at their highest;
- internal labor markets developed by firms that emphasize low turnover and high productivity, bear costs of screening, trying out and training, and use optimal reward schemes based on long-term contracts and tenured jobs; and
- professional labor markets for highly skilled workers who enjoy weak attachment to a firm, even if incorporated, and considerable bargaining power due to high market value of their expertise and high transferability of their competencies [Baron and Kreps (1999)].

Artistic labor markets mix elements from the secondary spot market and from the professional market model. On one side, employment is more and more contingent, as for secondary labor markets, but on the other side individuals are skilled or highly skilled, and non-substitutability is a core value, as in the so-called professional market. This is especially true of the performing arts which appear to have been quite avant-garde in designing and experiencing the process of increasingly flexible labor markets.<sup>28</sup> Contractual work arrangements and organizational forms in the whole economy have been

<sup>28</sup> On Hollywood, see Storper (1989); on the diffusion of the Hollywoodian flexibility model, see Kanter (1995); and on the trend towards flexibility and its strongly contrasted effects see Smith (1997), Cappelli (1999), Kalleberg (2000).

repeatedly portrayed as evolving along a similar path. Contingency is increasing, although at first glance it may have completely different meanings at both ends of the labor market, low-quality and low-skilled work being contingent at the bottom and highly-skilled work being more often independent contracting and free-lancing at the higher end. At the same time, firms increasingly try to build internal professional markets or to secure networks of recurrent collaborations with independent highly-skilled contractors and service firms once having outsourced part of their previous operations. So contingency, networking and individualization of working ties take on contrasting meanings depending on one's labor market value.

How significant are these changes and how far could a comparison with the arts be drawn? Prophecy on the changes in employment, management and society has grown over the last decade, mainly under the headlines of flexible specialization, flatter hierarchies, networks of organizations, the learning firm, self-designing teamwork, creativity enhancement and the knowledge society [DiMaggio (2001)]. Taken together, are these changes cumulative and do they result in a coherent new architecture of work and organization? Or do they belong to the toolkit of management textbooks and gurus? Has unfixed long-term employment inside big companies been dramatically eroded? Or, as some scholars in Europe claim, is it still and well alive? And should the alternative between traditional employment and all sorts of contingent employment not be better conceived as a matter of cyclical management, depending on the condition of business?

Dimensions of change are of course far too numerous at individual, organizational and societal levels to be caught in a unique new formula designing a new working age with its new rising class. Yet repeatedly the future of work organization is prophesied with reference to the core values that artistic professions share with other "knowledge workers" – autonomy, responsibility, self-control in teamwork, extended range of competencies enhancing the sense of initiative, creativity-driven commitment to work, individualized reputation based on track records and team project organization of work. In order to get a better understanding of what is at stake, the following distinction may be useful.

First, creativity- and knowledge-oriented change in work can be discussed with reference to economic sectors where it occurs most extensively. For example, network forms of organization have been for long studied as a special trait of craft industries and cultural industries (publishing, film and record industries), and research now concentrates more and more on organizational models supplied by all sorts of knowledge-intensive activities such as scientific research, cultural production in general, design work, computer programming, software development and professional services. A sectoral distinction leads to a specification of the kind of worker concerned; as stated by Powell (1995), these are highly-skilled and talented people possessing fungible knowledge that is not limited to a specific task but applicable to a wide range of activities, relying also on know-how and tacit knowledge that is difficult to codify, and being less creative and less productive under hierarchical governance. In this perspective, work systems of artists and "knowledge workers" share enough common characteristics to allow for illuminating comparisons. They also allow for meaningful contrasts; after all, R&D in

science and technology are intended to be cumulative, while art is contradictorily oriented towards the preservation of masterpieces and past achievements as well towards frenetic search for originality and novelty, be these a source for long-lasting innovation or for ephemeral fads and fashions.

A second way to make a strong claim to creativity is to envision a new architecture of society involving the emergence and rise of a new class, as authors like Florida (2002) have proposed. In this case, the various changes mentioned earlier are said to have converged up to a critical point where a significant part of the labor force shares values, life style, work ethos and leisure habits. A creative class is made up of people sharing a number of similar characteristics (high skills, knowledge intensive activities, learning potential and several other traits of the creative worker), but above all sharing paradoxically the feature of idiosyncratic individualism. In so saying, Florida reverses the usual social map; the bohemian fringe becomes the social core. This view leads to expansion of the boundaries of creative people by including many occupations and professions without any sense of sorting them out according to the hierarchical structure of each occupation. And not surprisingly, the new social map is mainly a locational one; the places where bundles of creative workers live are made out of dense, relaxed, flexible, inter-individual relationships.

It is worth remembering how the founding fathers of social science at the turn of the nineteenth century contrasted two types of society and social organization: the *Gemeinschaft* versus *Gesellschaft à la* Tönnies, or the mechanic structure versus organic structure of society *à la* Durkheim, a distinction on which Burns and Stalker (1961) have built later to characterize innovation-friendly organizational designs. Durkheim's organic structure of society refers to a society where people, being more and more differentiated due to more sophisticated division of labor and increasingly specialized skill requirements, are at the same time induced to get in closer contact and in more dense relationships with each other, since they need to exchange more and more knowledge and information. Yet Durkheim saw the risk of conflicting interests, of growing individualistic hedonism that would undermine social order and economic efficiency – in a word, of growing anomy (social disorder) necessitating regulation and social governance. Quite the opposite view is taken when the celebration of everyone's creativity causes the highly unequal chances of self-actualization to disappear; the homogenization of infinitely differentiated creative workers or of individualistic loosely-tied workers operates horizontally only. The specific characteristics of each trade's organization, its internal and external competition structure and its vertical differentiations vanish inside the large pot of a communitarian creative life style.

Lastly, emphasis on creativity at work may be more general and may concern organizational changes with no exclusive reference to particular sectors or to specific occupational groups. In that case, a bundle of closely related factors increases the need for organizational change toward more efficient, more rewarding and more stimulating work settings; such factors include computerized systems of production and information, just-in-time practices and computerized inventory control systems, increasing variety of products, shortening of product life cycle, new standards of competition in

industries, etc. These factors have all a strong impact on the challenges workers have to face, challenges that include increasing variability, quality problem solving, flexible coordination, management of formal and informal communication networks, rising information needs and evolving skills. These new challenges are core elements of the high-performance work systems implemented by managers in the post-taylorist era of HRM innovations [Baron and Kreps (1999); Appelbaum et al. (2000)]. Of course, such systems claim to be in line with a long tradition of work design improvement intended to overcome the disincentive dimensions of alienating work fragmentation and mechanistic specialization, and intended to promote motivation and self-actualization at work. Central aspects of such work systems are autonomy, participation in decision, coordination and communication among employees, selective staffing and extensive on-the-job and formal training opportunities, self-directed work teams, more sophisticated compensation policies taking into account the bundle of the financial incentives, intrinsic motivation incentives and long-term relationships that secure mutual trust and promote stakeholder behavior. Furthermore a sense of ownership, so decisive in creative activity, should be also acknowledged, insofar as each skilled worker's personal contribution to the output can be tracked and brought to his or her personal credit.

Each of the three approaches to creativity in labor markets and organizations reviewed above – sectoral-occupational, class-remodelling, and organizational – is built on a similar ground; that of one common good, be it that of knowledge and creative learning-by-doing produced and shared through networks of firms, or that of positive identification with work inside the firm, or that of creativity as a shared ethos of work and life. In a sense, the performance of artistic work systems, as reviewed previously, falls into each of these rubrics. Yet, one crucial issue is missing in the broadened picture of creativity at work: that of the several dimensions of inequality magnified by the work system in the arts, which builds on networks, reputation, short term contracts, and highly individualized performance ratings. Ironically enough, although most of the artists and professionals in the cultural sector are ideologically left-oriented and prone to advocate egalitarianism in society, art worlds have developed an insuperable engine to rank artists by quality level and market value, to select and signal the best works out of an ocean of products through winner-take-all tournaments and endless competitive comparisons, to let the whirl of fads and fashions promote or eliminate aspiring superstars, to celebrate skyrocketing and ephemeral celebrity as well as to provide civilization with Pantheons of eternal values. This corresponds mainly to the distributive justice principle of equity, which prevails especially where performance varies significantly across individuals, and “according to which individuals ought to be rewarded commensurate with the outcomes they generate, factoring in the inputs – efforts, ability, and so on – they brought to bear in performing the task” [Baron and Kreps (1999, p. 107)].

Thus inequalities in earnings may be occupationally legitimate if several requirements are met. A first condition is that each individual contribution to the team work or to the end result be identified specifically, which presumes there is an economic and legal value associated with the “traceability” of each individual work performance. Credits, signatures and individualized reputations mean each worker can be distinctively

identified with the complete history of his jobs, assignments and realizations. This has to do with the sense of identification with the end product of one's work as distinct from an anonymous contribution to an output whose production is heavily fragmented and standardized. A further condition of legitimization is that a professional's performance can be judged in comparison to and in direct competition with other works and professionals, and that inter-individual comparisons must occur frequently enough to submit rankings and valuations to periodic re-assessment, so that a long-lasting monopolistic rent of talent remains the rarest exception. Lastly, unforeseeable streams of successes and failures should signal to new candidates that creative innovation remains a highly uncertain game; in that respect, uncertainty in its several dimensions acts as a veil of ignorance thrown over the ultimate causes of achievement and success, as these stem from an undecipherable mix of chance, talent and work.

## 10. Summary and conclusion

Artists supply the innovation engine in the arts with work mainly under contingent arrangements. Long-term artistic employment has been vanishing except in heavily subsidized and sponsored organizations like orchestras and opera houses. The population of small artistic and cultural organizations has been growing as fast as the number of artists. Firms compete increasingly under flexible production schemes that bring them close to the nexus of spot transactions. Do vertically disintegrated systems of production favor only loose employment relationships? Studies show how filtering mechanisms and selective matching processes generate transactional stability as well as labor force segmentation. Employers use reputations as screening devices and signals of employability; artists learn how to compose well-balanced sets of recurrent and non-recurrent hiring ties, in order to secure a living as well as to increase their human capital. Talent agencies mediate the contingent labor market and do increasingly broker artistic projects.

Considerable inequalities in amounts of work and earnings among artists are observed, caused by the skewed distribution of talent and by joint consumption technologies that turn small differences in talent into huge earnings differentials. It has been shown that they may also trace back to the way a disintegrated labor market operates, since both the allocation of piecemeal work based on reputational rankings and team formation based on selective matchings magnify the power of differences in talent and work opportunity to increase inequality. Allocation of work under a contingent employment scheme should not cause the kind of permanent excess supply of labor in the arts that has been noted for decades if the occupational commitment of artists were not combined with the management of business uncertainty through overproduction of infinitely differentiated goods and services. On the whole, artistic labor markets provide a textbook model of imperfect monopolistic competition.

Large parts of the business risk are transferred down onto the artistic and technical workforce in a highly flexible and disintegrated organizational setting. The study of



artistic careers illuminates how individuals learn to manage the risks of their trade: through multiple jobholding, occupational role versatility, portfolio diversification of employment ties, and transfer incomes from public support, social insurance and social security programs. Institutional arrangements regarding the legal status and financial support of artists may differ greatly between the countries and states, yet occupational risk management is a common and basic condition of economic survival and personal success; ironically enough, this shows how rationally artists behave, although artistic work may be highly idiosyncratic. Thus artists may be seen less like rational fools than like Bayesian actors.

## References

- Abbott, A. (1988). *The System of Professions*. The University of Chicago Press, Chicago.
- Abbott, A. (1990). "Conceptions of time and events in social science methods". *Historical Methods* 23, 140–150.
- Abbott, A., Hrycak, A. (1990). "Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers". *American Journal of Sociology* 96, 144–185.
- Aghion, P., Howitt, P. (1998). *Endogenous Growth Theory*. MIT Press, Cambridge, MA.
- Aksoy, A., Robins, K. (1992). "Hollywood for the 21st century: Global competition for critical mass in image markets". *Cambridge Journal of Economics* 16, 1–22.
- Allmendinger, J., Hackman, R. (1996). "Organizations in changing environments: The case of East German symphony orchestras". *Administrative Science Quarterly* 41, 337–369.
- Allmendinger, J., Hackman, R., Lehman, E. (1994). "Life and work in symphony orchestras". *Harvard Business School Report* 7.
- Alper, N., Galligan, A. (1999). "Recession to Renaissance: A comparison of Rhode Island artists, 1981 and 1997". *Journal of Arts Management, Law and Society* 29 (3), 178–203.
- Alper, N., Wassall, G. (2000). "More than once in a blue moon: Multiple jobholdings by American artists". National Endowment for the Arts, Research Division Report 40, Seven Locks Press, Santa Anna.
- Alper, N., Wassall, G. (2006). "Artists' careers and their labor markets". In: Ginsburgh, V., Thosby, D. (Eds.), *Handbook of the Economics of Art and Culture*, vol. 1. Elsevier, Amsterdam. Chapter 23, this volume.
- Alper, N., Wassall, G., Jeffri, J., Greenblatt, R., Kay, A., Butcher, S., Chartrand, H. (1990). "Artists in the work force: Employment and earnings 1970–1990". National Endowment for the Arts, Research Division Report 37, Seven Locks Press, Santa Anna.
- Amirault, T. (1997). "Characteristics of multiple jobholders, 1995". *Monthly Labor Review* 3, 9–15.
- Anzieu, D. (1981). *Le corps de l'oeuvre*. Gallimard, Paris.
- Appelbaum, E., Bailey, T., Berg, P., Kalleberg, A. (2000). *Manufacturing Advantage. Why High-Performance Work Systems Pay Off*. ILR and Cornell University Press, Ithaca.
- Arendt, H. (1959). *The Human Condition*. Doubleday Anchor Press, Garden City.
- Arian, E. (1971). *Bach, Beethoven and Bureaucracy*. The University of Alabama Press, Alabama.
- Aronson, R.L. (1991). *Self-Employment. A Labor Market Perspective*. ILR Press, Ithaca.
- Atkinson, A.B., Micklewright, J. (1991). "Unemployment compensation and labour market transitions: A critical review". *Journal of Economic Literature* 29, 1679–1727.
- Baker, W.E., Faulkner, R.R. (1991). "Role as resource in the Hollywood film industry". *American Journal of Sociology* 97, 279–309.
- Baron, J., Kreps, D. (1999). *Strategic Human Resources*. Wiley, New York.
- Baumol, W.J. (2002). *The Free-Market Innovation Machine*. Princeton University Press, Princeton.
- Baumol, W.J., Baumol, H. (1994). "On the economics of musical composition in Mozart's Vienna". In: Morris, J.M. (Ed.), *On Mozart*. Cambridge University Press, New York.

- Baumol, W.J., Bowen, W.G. (1966). *Performing Arts: The Economic Dilemma*. The Twentieth Century Fund, New York.
- Baumol, W.J., Jeffri, J., Throsby, D. (2004). "Making changes. Facilitating the transition of dancers to post-performance careers". Research Report, The aDVANCE Project, New York.
- Becker, H.S. (1982). *Art Worlds*. University of California Press, Berkeley.
- Bénichou, P. (1985). *Le sacre de l'écrivain*. Corti, Paris.
- Bielby, W.T., Bielby, D.D. (1999). "Organizational mediation of project-base labor markets: Talent agencies and the careers of screenwriters". *American Sociological Review* 64, 64–85.
- Blau, J.R. (1984). *Architects and Firms*. MIT Press, Cambridge, MA.
- Blau, J.R. (1989). *The Shape of Culture: A Study of Contemporary Cultural Patterns in the United States*. Cambridge University Press, Cambridge.
- Boin, J.G., Bouvaist, J.M. (1989). *Du printemps des éditeurs à l'âge de raison*. La Documentation Française, Paris.
- Bourdieu, P. (1992). *Les règles de l'art*. Seuil, Paris.
- Burke, A.E. (1997). "Small firm start-up by composers in the recording industry". *Small Business Economics* 9, 463–471.
- Burns, T., Stalker, G.M. (1961). *The Management of Innovation*. Oxford University Press, Oxford.
- Camerer, C., Lovo, D. (1999). "Overconfidence and excess entry: An experimental approach". *American Economic Review* 89, 306–318.
- Cappelli, P. (1999). *The New Deal at Work: Managing the Market-Driven Workforce*. Harvard Business School Press, Boston.
- Caves, R.E. (2000). *Creative Industries: Contracts Between Art and Commerce*. Harvard University Press, Cambridge, MA.
- Champy, F. (1998). *Les architectes et la commande publique*. PUF, Paris.
- Christopherson, S. (1996). "Flexibility and adaptation in industrial relations: The exceptional case of the US media entertainment industries". In: Gray, L., Seeber, R. (Eds.), *Under the Stars*. Cornell University Press, Ithaca.
- Christopherson, S., Storper, M. (1989). "The effects of flexible specialization on industrial politics and the labour market: The motion picture industry". *Industrial and Labour Relations Review* 42, 331–347.
- Coase, R.H. (1979). "Payola in radio and television broadcasting". *Journal of Law and Economics* 22, 269–328.
- Cohen, G.A. (2000). *Karl Marx's Theory of History*, new ed. Princeton University Press, Princeton. First ed. 1978.
- Coser, L.A., Kadushin, C., Powell, W.W. (1982). *Books. The Culture and Commerce of Publishing*. Basic Books, New York.
- Cowen, T. (2000). *What Price Fame?* Harvard University Press, Cambridge, MA.
- Debeauvais, R., Menger, P.-M., Rannou, J., Vari, S., Laplante, B. (1997). *Le spectacle vivant*. La Documentation Française, Paris.
- DiMaggio, P. (1977). "Market structures, the creative process and popular culture: Toward an organizational reinterpretation of mass-culture theory". *Journal of Popular Culture* 11, 436–452.
- DiMaggio, P. (1986). "Can culture survive the marketplace". In: DiMaggio, P. (Ed.), *Nonprofit Enterprise in the Arts*. Oxford University Press, New York.
- DiMaggio, P. (Ed.) (2001). *The Twenty-First-Century Firm. Changing Economic Organization in International Perspective*. Princeton University Press, Princeton.
- Drèze J. (1987). "Human capital and risk-bearing". Reprinted in: Drèze, J. (1987). *Essays on Economic Decisions under Uncertainty*. Cambridge University Press, Cambridge.
- Ehrlich, C. (1985). *The Music Profession in Britain since the Eighteenth Century*. Clarendon Press, Oxford.
- Elster, J. (1985). *Making Sense of Marx*. Cambridge University Press, Cambridge.
- Faulkner, R. (1973). "Career concerns and mobility motivations of orchestra musicians". *Sociological Quarterly* 14, 334–349.
- Faulkner, R.R. (1983). *Music on Demand*. Transaction Books, New Brunswick.

- Faulkner, R.R., Anderson, A.B. (1987). "Short-term projects and emergent careers: Evidence from Hollywood". *American Journal of Sociology* 92, 879–909.
- Federico, R. (1983). "The decision to end a performing career in ballet". In: Kamerman, J., Martorella, R. (Eds.), *Performers and Performances*. Praeger, New York.
- Feist, A. (1998). "Great Britain". In: Bodo, C., Fisher, R. (Eds.), *New Frontiers for Employment in Europe*. Circle Publications No. 9, Rome.
- Filer, R.K. (1986). "The 'starving artist' – Myth or reality? Earnings of artists in the United States". *Journal of Political Economy* 96, 56–75.
- Filer, R.K. (1990). "Arts and academe: The effects of education on earnings of artists". *Journal of Cultural Economics* 2, 15–38.
- Florida, R. (2002). *The Rise of the Creative Class*. Basic Books, New York.
- Fohrbeck, K., Wiesand, A. (1975). *Der Künstler-Report*. Hanser Verlag, München.
- Frank, R.H., Cook, P.J. (1995). *The Winner-Take-All Society*. The Free Press, New York.
- Freidson, E. (1986a). *Professional Powers*. The University of Chicago Press, Chicago.
- Freidson, E. (1986b). "L'analyse sociologique des professions artistiques". *Revue Française de Sociologie* 27, 431–443.
- Freidson, E. (1990). "Labors of love: A prospectus". In: Erikson, K., Vallas, S.P. (Eds.), *The Nature of Work: Sociological Perspectives*. Yale University Press, New Haven.
- Frey, B., Pommerehne, W. (1989). *Muses and Markets*. Basil Blackwell, Oxford.
- Ginsburgh, V. (2003). "Awards, success and aesthetic quality in the arts". *Journal of Economic Perspectives* 17, 99–111.
- Ginsburgh, V., van Ours, J. (2003). "Expert opinion and compensation: Evidence from a musical competition". *American Economic Review* 93, 289–296.
- Glaeser, E.L. (1998). "Are cities dying?" *Journal of Economic Perspectives* 2, 139–160.
- Graña, C. (1964). *Bohemian versus Bourgeois*. Basic Books, New York.
- Gray, L.S., Seeber, R.L. (Eds.) (1996). *Under the Stars Essays on Labor Relations in Arts and Entertainment*. Cornell University Press, Ithaca.
- Habermas, J. (1988). *Le discours philosophique de la modernité*. Gallimard, Paris.
- Hall, R.E. (1970). "Why is the unemployment rate so high at full employment?" *Brookings Papers on Economic Activity* 3, 369–403.
- Hamlen, W. Jr. (1991). "Superstardom in popular music: Empirical evidence". *Review of Economics and Statistics* 4, 729–733.
- Hamlen, W. Jr. (1994). "Variety and superstardom in popular music". *Economic Inquiry* 3, 395–406.
- Handy, C. (1989). *The Age of Unreason*. Harvard Business School Press, Boston.
- Hesmondhalgh, D. (1996). "Flexibility, post-Fordism and the music industry". *Media, Culture and Society* 18, 469–488.
- Hesmondhalgh, D. (1998). "The British dance music industry". *British Journal of Sociology* 2, 234–251.
- Hirsch, P. (1972). "Processing fads and fashions". *American Journal of Sociology* 77, 639–659.
- Hirschman, A.O. (1986). *Rival Views of Market Society*. Viking, New York.
- Jackson, C., Honey, S., Hillage, J., Stock, J. (1994). *Careers and Training in Dance and Drama*. Institute of Manpower Studies, Brighton.
- Jeffri J. (1991). "The artists training and career project: Painters". Mimeo, Columbia University, New York.
- Johnson, W. (1978). "A theory of job shopping". *Quarterly Journal of Economics* 92, 261–278.
- Jovanovic, B. (1979). "Job matching and the theory of turnover". *Journal of Political Economy* 87, 972–990.
- Kalleberg, A.L. (2000). "Nonstandard employment relations: Part-time, temporary and contract work". *Annual Review of Sociology* 26, 341–365.
- Kanter, R.M. (1989). "Careers and the wealth of nations: A macro-perspective on the structure and implications of career forms". In: Arthur, M., Hall, D., Lawrence, B. (Eds.), *Handbook of Career Theory*. Cambridge University Press, Cambridge.
- Kanter, R.M. (1995). *World Class: Thriving Locally in the Global Economy*. Simon and Schuster, New York.
- Kealy, E. (1979). "The case of sound mixers and popular music". *Sociology of Work and Occupations* 1, 3–29.

- Killingsworth, M. (1983). *Labor Supply*. Cambridge University Press, Cambridge.
- Knott, C. (1994). *Crafts in the 1990s*. The Crafts Council, London.
- Kremer, M. (1993). "The O-rings theory of economic development". *Quarterly Journal of Economics* 108, 551–575.
- Kris, E., Kurz, O. (1987). *L'image de l'artiste*. French translation. Editions Rivages, Marseille.
- Krueger, A.B. (2005). "The economics of real superstars: The market for rock concerts in the material world". *Journal of Labor Economics* 23, 1–30.
- Lancaster, K. (1979). *Variety, Equity and Efficiency*. Columbia University Press, New York.
- Laplante B. (1990). "La vocation de l'acteur". Unpublished PhD, University of Montreal.
- Lazear, E.P., Rosen, S. (1981). "Rank-order tournaments as optimum labor contracts". *Journal of Political Economy* 89, 841–864.
- Lenman, R. (1989). "Painters, patronage and the art market in Germany 1850–1914". *Past and Present* 123, 109–140.
- MacDonald, G.M. (1982). "A market equilibrium theory of job assignment and sequential accumulation of information". *American Economic Review* 72, 1038–1055.
- MacDonald, G.M. (1988). "The economics of rising stars". *American Economic Review* 78, 155–166.
- Marshall, A. (1947). *Principles of Economics*, eighth ed. MacMillan, New York.
- Menger, P.-M. (1983). *Le paradoxe du musicien*. Flammarion, Paris.
- Menger, P.-M. (1989). "Rationalité et incertitude de la vie d'artiste". *L'Année Sociologique* 39, 111–151.
- Menger, P.-M. (1993). "L'hégémonie parisienne. Economie et politique de la gravitation artistique". *Annales ESC* 6, 1565–1600.
- Menger, P.-M. (1997). *La profession de comédien. Formations, activités et carrières dans la démultiplication de soi*. La Documentation Française, Paris.
- Menger, P.-M. (2003). *Portrait de l'artiste en travailleur. Métamorphoses du capitalisme*. Seuil/La République des Idées, Paris.
- Menger, P.-M. (2005). *Les intermittents du spectacle. Sociologie d'une exception culturelle et sociale*. Editions de l'Ecole des Hautes Etudes, Paris.
- Menger, P.-M., Gurgand, M. (1996). "Work and compensated unemployment in the performing arts. Exogenous and endogenous uncertainty in artistic labour markets". In: Ginsburgh, V., Menger, M. (Eds.), *Economics of the Arts*. North-Holland, Amsterdam.
- Miller, R.A. (1984). "Job matching and occupational choice". *Journal of Political Economy* 92, 1086–1120.
- Mitchell, R. (1992). "Artists policies". In: Irjala, A. (Ed.), *European Symposium on the Status of the Artist*. UNESCO Publications, Helsinki.
- Montias, J.M. (1982). *Artists and Artisans in Delft*. Princeton University Press, Princeton.
- Montias, J.M. (1996). *Le marché de l'art aux Pays-Bas, XVe-XVIIe siècles*. Flammarion, Paris.
- Moulin, R. (1987). "The French Art Market. A Sociological View". Rutgers University Press, New Brunswick.
- Moulin, R. (1992). *L'artiste, l'institution et le marché*. Flammarion, Paris.
- Musgrave, R.A. (1959). *The Theory of Public Finance*. McGraw-Hill, New York.
- Nash, D. (1955). "The American composer's career". Reprinted in: Albrecht, M., Barnett, J., Griff, M. (Eds.), (1970). *The Sociology of Art and Literature*. Duckworth, London.
- Netzer, D. (1978). *The Subsidized Muse: Public Support for the Arts in the United States*. Cambridge University Press, Cambridge.
- Nisbett, R., Ross, L. (1980). *Human Inference*. Prentice Hall, Englewood Cliff, NJ.
- O'Brien, J., Feist, A. (1995). "Employment in the arts and cultural industries: An analysis of the 1991 Census". Mimeo, Arts Council of England, London.
- Observatoire de l'emploi culturel (2004a). "L'emploi dans le secteur culturel en 2002". Note No. 35, Ministère de la Culture, Paris.
- Observatoire de l'emploi culturel (2004b). "L'emploi dans les professions culturelles". Note No. 30, Ministère de la Culture, Paris.
- Okun, A.M. (1981). *Prices and Quantities*. Basil Blackwell, Oxford.

- Paradeise, C. (1998). *Les comédiens*. Presses Universitaires de France, Paris.
- Paul, A., Kleingartner, A. (1994). "Flexible production and the transformation of industrial relations in the motion picture and television industry". *Industrial and Labor Relations Review* 4, 662–678.
- Peacock, A. (1970). *A Report on Orchestral Resources*. The Arts Council of Great Britain, London.
- Peacock, A., Weir, R. (1975). *The Composer in the Market Place*. Faber, London.
- Powell, W.W. (1985). *Getting into Print*. The University of Chicago Press, Chicago.
- Powell, W.W. (1995). "Neither market nor hierarchy: Network forms of organization". *Research in Organizational Behavior* 12, 295–336.
- Quingley, J. (1998). "Urban diversity and economic growth". *Journal of Economic Perspectives* 2, 127–138.
- Rannou, J., Roharik, I. (2006). *Les danseurs, un métier d'engagement*. La Documentation Française, Paris.
- Romer, P.M. (1990). "Endogenous technical change". *Journal of Political Economy* 98, 71–102.
- Rosen, S. (1981). "The economics of superstars". *American Economic Review* 75, 845–858.
- Rosen, S. (1986). "The theory of equalizing differences". In: Ashenfelter, O., Layard, R. (Eds.), *Handbook of Labour Economics*. North-Holland, Amsterdam.
- Sartre, J.P. (1971). *L'idiote de la famille*. Gallimard, Paris.
- Scott, A.J. (2000). *The Cultural Economy of Cities*. Sage, London.
- Seaman, B.A. (2003). "Cultural and sport economics: Conceptual twins?" *Journal of Cultural Economics* 27, 81–126.
- Sen, A. "Rational fools". Reprinted in: Sen, A. (1982). *Choice, Welfare and Measurement*. Basil Blackwell, Oxford.
- Simpson, C. (1981). *SoHo: The Artist in the City*. The University of Chicago Press, Chicago.
- Smith, V. (1997). "New forms of work organization". *Annual Review of Sociology* 23, 315–339.
- Spilerman, S. (1977). "Careers, labor market structure, and socioeconomic achievement". *American Journal of Sociology* 3, 551–593.
- Stephan, P. (1996). "The economics of science". *Journal of Economic Literature* 34, 1199–1235.
- Stigler, G.S. (1962). "Information in the labor market". *Journal of Political Economy* 70, 94–105.
- Stinchcombe, A.L. (1963). "Some empirical consequences of the Davis–Moore theory of stratification". Reprinted in: Stinchcombe, A.L. (1986). *Stratification and Organization*. Cambridge University Press, Cambridge.
- Stinchcombe, A.L. (1968). *Constructing Social Theories*. The University of Chicago Press, Chicago.
- Storper, M. (1989). "The transition to flexible specialisation in the film industry". *Cambridge Journal of Economics* 13, 273–305.
- Storper, M. (1993). "Flexible specialisation in Hollywood: A response to Aksoy and Robins". *Cambridge Journal of Economics* 17, 479–484.
- Storper, M., Walker, R. (1989). *The Capitalist Imperative. Territory, Technology and Industrial Growth*. Basil Blackwell, Oxford.
- Sussmann, L. (1984). "Anatomy of the dance company boom". *Dance Research Journal* 2, 23–28.
- Taylor, B. (1987). "Artists in the marketplace: A framework for analysis". In: Shaw, D., Hendon, W., Waits, C.R. (Eds.), *Artists and Cultural Consumers*. Association for Cultural Economics, Akron.
- Throsby, D. (1992). "Artists as workers". Reprinted in: Towse, R. (Ed.), (1997). *Cultural Economics: The Arts, the Heritage and the Media Industries*. Edward Elgar, Cheltenham.
- Throsby, D. (1994a). "A work-preference model of artist behaviour". In: Peacock, A., Rizzo, I. (Eds.), *Cultural Economics and Cultural Policies*. Kluwer Academic, Dordrecht.
- Throsby, D. (1994b). "The production and consumption of the arts". *Journal of Economic Literature* 32, 1–29.
- Throsby, D. (1996). "Disaggregated earnings functions for artists". In: Ginsburgh, V., Menger, M. (Eds.), *Economics of the Arts*. North-Holland, Amsterdam.
- Throsby, D., Hollister, V. (2003). *Don't Give Up Your Day Job: An Economic Study of Professional Artists in Australia*. Australia Council for the Arts, Sydney.
- Topel, R., Welch, F. (1980). "Unemployment insurance: Survey and extensions". *Economica* 47, 351–379.
- Towse, R. (1993). *Singers in the Marketplace: The Economics of the Singing Profession*. Clarendon Press, Oxford.

- Towse, R. (1996). *The Economics of Artists' Labour Markets*. The Arts Council of England, London.
- Warnke, M. (1985). *Hofkünstler. Zur Vorgeschichte des modernen Künstlers*. DuMont Buchverlag, Köln.
- Wassall, G., Alper, N., Davison, R. (1983). *Art Work: Artists in the New England Labor Market*. New England Foundation for the Arts, Cambridge, MA.
- Wassall, G., Alper, N. (1990). "When is an artist an artist: An analysis of factors related to claiming membership in this profession". *Journal of Arts Management and Law* 4, 37–50.
- Weick, K. (1979). *The Social Psychology of Organizing*. Random House, New York.
- Weick, K., Berlinger, L. (1989). "Career improvisations in self-designing organizations". In: Arthur, M., Hall, D., Lawrence, B. (Eds.), *Handbook of Career Theory*. Cambridge University Press, Cambridge.
- Westby, D. (1960). "The career experience of the symphony musician". *Social Forces* 2, 224–230.
- White, H.C. (1993). *Careers and Creativity*. Westview Press, Boulder.
- White, H.C., White, C.A. (1965). *Canvases and Careers*. Wiley, New York.

## ARTISTS' CAREERS AND THEIR LABOR MARKETS\*

NEIL O. ALPER and GREGORY H. WASSALL

*Northeastern University, Boston, MA, USA*

### Contents

Abstract	814
Keywords	814
1. Introduction	815
2. A classification scheme for studies of artists' careers	815
3. Empirical studies of artists' careers	818
3.1. Careers but not panel	818
3.2. Surveys via retrospection	819
3.3. Quasi-panels	822
3.4. True panel data	825
4. An empirical investigation of artists in the US: 1940–2000	827
4.1. Using census data	828
4.2. Size and growth of the artist profession: 1940–2000	831
4.3. Labor supply and labor market characteristics	832
4.4. Earnings, calculated wage rates, and earnings variability	836
4.5. Earnings functions	841
5. US artists' careers: 1979–1998	847
5.1. Flow and tenure	848
5.2. The start	849
5.3. The transition	850
5.4. The leavers	852
5.5. The post-artist period	854
6. Summary, conclusions and future research	856
Appendix A	858
References	858
Further reading	861

\* We would like to thank Rochelle (Shelly) Tapper and David Hyde for their able assistance in preparing the data and tables.

## Abstract

This chapter is a continuation of ongoing work by economists and others on artists' labor markets and careers. It highlights the use of quasi-panel data obtained from census data to examine the employment and earnings of artists while comparing them to all the other professional and technical workers. It also provides a glimpse into what can be learned about artists' careers from true panel data.

Quasi-panels from the seven most recent US censuses (1940–2000) provide a reasonably consistent set of findings in each census year. Artists are found to work fewer hours, suffer higher unemployment and earn less than members of the reference group. Over the sixty year period, disparities in unemployment and annual hours worked are found to shrink somewhat, but disparities in earnings do not. Artists earned less across all years even when only members working full-time year-round of each group are compared. The earnings of artists are found to display greater variability than those of other professional and technical workers.

The National Longitudinal Survey of Youth 1979 is used to examine almost twenty years in the artists' lives and provides some insights into their careers. It suggests that many people participate in the artistic labor market, but that few succeed to the point that enables them to develop a career in the arts. In part due to their relatively high educational levels, artists are found to be able to transition from forays into arts occupations to jobs in professional and managerial occupations, not into service occupations as artist 'mythology' might suggest. We find that when the artists are young and struggling to make it they do work in various service occupations that tend to provide greater work schedule flexibility.

## Keywords

artists, artists' earnings, artists' labor markets, artists' careers, National Longitudinal Survey of Youth 1979

*JEL classification:* Z11, J44



"It was time. It got to the point where you're just tired of being poor." Bassoonist Chad Alexander ten years after graduating from Juilliard who recently sold his bassoon to cover credit card bills and now works as an assistant insurance underwriter.

(Wakin, *New York Times*, December 12, 2004)

## 1. Introduction

Economists' research into the economics of art and culture has a relatively short history. Baumol and Bowen's (1966) book, written forty years ago, is seen by many as the starting point of the economists' foray into this sub-discipline. Research on the artists' labor markets and their careers has even a shorter history. Its empirical roots lie in studies of single artistic occupations that are, for the most part, barely 30 years old.

As might be expected, these early studies were cross-sectional and provided important, but limited information on careers and labor market behavior. Economists learned a great deal from these studies, but they also raised a number of unanswered questions. Table 1 provides a list of some of the landmarks among the empirical studies of artists many of which are discussed in more detail below. The passage of time has allowed economists to enhance their knowledge through the use of increasingly available panel data of various types.

This chapter starts with a description of a framework for categorizing and reviewing existing studies (Section 2). This is followed by a review of some of the major studies in these categories (Section 3) and concludes with two studies of our own. One will provide an overview of many major themes learned from the cross-sectional analysis by examining a quasi-panel for the US (Section 4) based on the US Census Bureau's Public Use Microdata Samples (PUMS) drawn from the US decennial censuses for 1940 through 2000. The other (Section 5) addresses some themes regarding the artists' careers and labor market experiences that cannot be answered by cross-sectional data through the use of the 1979 National Longitudinal Survey of Youth (NLSY79). The last part (Section 6) provides a summary, conclusions and suggestions for further research.

## 2. A classification scheme for studies of artists' careers

For the most part current research examining the careers' of artists can be placed into four categories. The one to receive the least attention in this chapter develops theoretical models of artistic career processes. MacDonald (1988), for example, developed a two-period model of performers' behavior that predicts young performers will earn incomes that are less than what they could earn outside the arts. The three additional categories of current research relate to empirical research.

The second category utilizes existing information on a group of artists obtained from a variety of sources to develop an understanding of an aspect of artists' careers. Sometimes the data for this research is anecdotal, as in Richardson's (1980) study of opera

Table 1  
Some landmarks in empirical research into artist labor markets

Empirical Finding	Authors	Artist Group/Database
Artists as risk-takers	Santos (1976)	US Census
Determinants of artists' earnings using statistical earnings function	Filer (1986)	US Census
Artists' dependence on non-arts jobs for income	Throsby (1986) Jeffri (1988) Wassall, Alper and Davison (1983)	Australian artists Artists in several US cities New England artist survey
Issues in identifying and defining artists	Filer (1986) Wassall and Alper (1985) Karttunen (1993, 1998a)	US Census New England artist survey Finnish artists
Artist moonlighting patterns	Alper and Wassall (2000)	Current Population Survey
Longitudinal analyses of artists' careers	Rengers (2002)	Dutch artists
Career transitions of artists	Alper and Wassall (1998) Smith (2000) Montgomery and Robinson (2003)	Nat'l. Survey of College Graduates US Census College graduates
Examination of artists' entire careers	Galenson (2000a, 2000b)	Painters

singers. Other times it is much more quantitative as in Galenson's (2000a, 2000b, 2001, 2002) studies of painters' careers. Galenson uses auction transactions in order to identify at what point over their lifetimes artists produce their best work.

The third category is based on surveying artists and asking them to recreate their careers by responding to written questionnaires or personal interviews. There are two types of retrospective studies, one utilizing special surveys of artists and the other using surveys of a broader group, including the general population. Examples of the former include Montgomery and Robinson's (2003) study of dance majors and Stohs' (1989, 1990, 1991a, 1991b, 1992a, 1992b) study of graduates of the School of the Art Institute of Chicago. The research by Smith (2000) using the 1970 census for the US and Alper and Wassall (1998) are examples of the latter.

The last category is based on panel data. There are very few examples of research on artists' careers based on true panel data. There are a number of studies that are what might be called "quasi-panels". They follow groups of artists, many of whom are likely to be the same from survey to survey over time rather than following the same group of people/artists from the start to the end of a multiyear period. Two examples of quasi-panel research are the studies of Bielby and Bielby (1987, 1989, 1993b, 1998) based on the administrative records of the Writers Guild of America, west, the union for writers

in the US television and film industries, and the work of Throsby and his colleagues (1986, 1989, 1994, 1995, 2003) utilizing the four surveys of Australian artists for 1983, 1988, 1993 and 2002. The research on US artists from 1940 to 2000 to be discussed later in this paper would also fit into this category.<sup>1</sup>

There appear to be only a few true panel studies. One, by Rengers (1998a, 2002), specifically targeted artists by following a group of 540 artists who graduated from several art schools in the Netherlands. The artists were surveyed twice over a six-year post-graduation period. The first interview was approximately one and one-half years post-graduation and the other was six years post-graduation. The other by Alper and Wassall (2002), which will be presented in detail below, is a study of artists who were part of a panel survey of the US population. It has followed essentially the same group of people for more than twenty years. The former study is limited by the small number of years of information making it difficult to observe and examine the artists' career paths. The latter, while covering a significantly longer time period and thus potentially providing greater insights into artists' careers, was not specifically designed to examine issues that are unique to artists. It also suffers from being a survey designed to represent the general population and therefore includes relatively few artists.

There are a number of advantages to using panel data rather than cross-sectional data [Federal Committee (1985)] when studying artists. Perhaps most important is the reduced variability in the estimates of change when compared to the results from unrelated or pooled cross-sections taken over time. There is no need to control for as many factors related to the change that might impact the reasons for the observed patterns, because in each time period it is the same people being examined. For example there is generally no change in the composition of the sample related to gender, race or ethnicity, all of which can influence the phenomena being measured.

The reliability and accuracy of the information obtained is also another advantage of a panel survey. A panel survey is repeated on a regular basis, so the reference period is generally shorter resulting in less recall bias than a single retrospective survey. Trying to remember what one did over the last twelve months when asked about work activities is much easier than trying to recreate an entire working career that may extend 10, 15 or 20 or more years. Information on income, expenses, and the like, can be recalled for the last year or so, but certainly most people would find it difficult to recall much beyond that. In retrospective surveys, participants tend to selectively telescope, suppress or embellish events that happened to them in the more distant past.

A panel enables data to be collected in a clearly identified time sequence so the direction and causation of change can be more easily and reliably identified. For example, when examining the determinants of who becomes a successful artist it might be interesting to study how a young person's goals and attitudes towards work and his/her

<sup>1</sup> Except for the works discussed in this paper or listed among the references, the authors were able to find little evidence of empirical research on artists in other countries. Most of the research not directly discussed in the paper is either simply cross-sectional or quasi-panel.

chosen profession influences his/her future success. Retrospective information on these issues may be influenced by events that have occurred while information collected at the time these goals and attitudes are being developed is unaffected by subsequent outcomes. Also, it might be useful to examine how attitudes toward the profession are developed over time and the relationship individual experiences have in this process.

The repeated nature of a panel, with a short time period between surveys, tends to lessen the burden on the respondents and therefore is likely to increase their willingness to participate and to provide accurate information. The amount of information that needs to be collected at each interview is less than for a single retrospective survey so the time commitment for each interviewee is reduced. Many questions would not need to be repeated and may simply be updated when necessary, so that new information can be obtained without increasing the respondents' burden. The total amount of information collected through a panel is generally greater than what could be collected in a single retrospective survey.

This is not to say that panel studies do not have disadvantages and problems as well. Perhaps the most problematic is that the full set of benefits to be gained from panel surveys generally cannot be obtained for a number of years. While information from each cross-section can be very useful, the benefits from being able to measure change and long-term outcomes will take time. Participant attrition can lead to serious biases above the typical refusals of any cross-section survey. The cost of a panel survey is generally higher than a retrospective cross-section. For there to be significant value from a panel survey the original budget commitment must be for more than one survey.

Using a panel to study artists has an additional problem if it is not designed specifically for a sample of artists. In the US artists comprise less than two percent of the US labor force so that unless it is a large survey, the number of artists who are likely to be part of the sample is small making it difficult, if not impossible, to obtain reliable estimates. This is an especially difficult problem if there is an interest in exploring the differences among the various genres of artists.

### **3. Empirical studies of artists' careers**

#### *3.1. Careers but not panel*

First is an examination of research on artists' careers based on data that are neither panel in nature nor based on surveys of artists. Examples that fit into this category are studies done by Galenson (2000a, 2000b, 2001, 2002), Galenson and Weinberg (2001) and Richardson (1980).

Galenson is interested in determining at what point in the artist's career he/she is most productive. In particular he attempts to identify when artists, i.e., painters, produce their most valuable work and their best work. He is also interested in seeing whether this point in an artist's career has remained constant over time. He uses two sources of information. To determine value he uses data from art auctions. To determine quality he

uses surveys of art history, texts and published monographs, and the paintings that are selected for retrospective exhibitions of the artists' works. He characterizes the authors of the surveys and the curators of the exhibitions as art experts who can provide quality assessments of the artists' works.

Galenson examined the careers of French artists born between 1796 and 1900 and American artists born 1870 to 1940. He found that both the artists' most valued works and their best works tended to occur at particular points in their careers and that these changed over time. For both the French and American artists he determined that the artists born at the beginning of each period produced their most valued and their best work at significantly later points in their careers than those born at the end of each period. His explanation for this result is that the two artist groups introduced innovations into their art in very different ways. The younger artists were "conceptual innovators" who were quick to introduce new ideas into their paintings. As in many other fields requiring considerable amounts of intellect, e.g., physics and mathematics, the best years for what Galenson calls "radical conceptual innovation" tend to be early in a person's career and it is this innovation that makes the artist's product the most valued and his/her best work. The older artists, those born earlier in the periods, were "experimental innovators" who spent many years working to solve a single problem and thus produced a body of work that illustrated the evolution of the problem's solution they wanted to solve rather than a single, revolutionary work. Thus their most valued and best paintings tended to occur much later in their careers.

Richardson's (1980) work was primarily case studies of various American opera singers' careers. He noted that unlike other performers, opera singers cannot really start to train their voices until their late teens or early twenties and that it takes at least five years to develop the voice; the stamina needed to sing an entire opera; to learn to sing the repertoire and how to act it on stage. Opera singers continue to train throughout their careers. Even with constant training few opera singers continue to perform past their early fifties. For many years American opera singers, especially black Americans, had to go to Europe for experience and to develop a reputation and then they would be able to return to a successful career in the US.

### 3.2. *Surveys via retrospection*

Montgomery and Robinson (2003) studied the graduates of the Five College Dance Program.<sup>2</sup> The graduates were asked to recreate, through a retrospective survey, their post-graduation behavior. Those surveyed completed their undergraduate degrees over the 1970 to 1998 period. The vast majority (84 percent) worked in dance at some time during the post-graduation period but that proportion decreases quite rapidly over time. Of those who graduated post-1990, 78 percent were still involved in dance in 1998

<sup>2</sup> The five colleges involved are: University of Massachusetts Amherst, Amherst College, Mt. Holyoke College, Hampshire College and Smith College.

with 41 percent dancing and an equal number teaching dance. Of those who graduated pre-1990, only 40 percent were still in dance with fewer than 20 percent indicating that they were dancers and about one-third identified themselves as dance teachers. The average number of years spent dancing, teaching dance or as a choreographer ranged from three to six years. One ramification of the career transition out of dance into other occupations Montgomery and Robinson found was that the program's graduates became more geographically mobile as they moved out of the dance occupations.

Stohs' (1989, 1990, 1991a, 1991b, 1992a, 1992b) studies of graduates of the School of the Art Institute of Chicago were more sociological and psychological than economic, but did identify some aspects of their careers that are of interest to economists. In 1981 she surveyed a group of students who had graduated in 1963. She categorized these artists as either fine artists, those who produce art products, or applied artists, those who produce art for commercial purposes. Stohs found that most of the fine arts majors left the fine arts for advertising, teaching or a non-arts occupation. Only six percent of the fine arts majors were still supporting themselves in the fine arts 18 years after graduation, what she identifies as 'midlife', while almost half were doing so as young adults shortly after graduation.

Stohs characterized artistic careers based on the number of related job changes the artists made over their careers. Her characterization of a continuous career was one with three or fewer related job changes. An interrupted career was one with four or more job changes. Perhaps not surprisingly, she found that men were much more likely to experience careers that were steady with steady promotions than women. Two-thirds of the men and only one-quarter of the women had steady careers with promotions from young adulthood to midlife. Reflective of this difference, Stohs found that for almost three-quarters of the men and just over half the women their artistic work was proving to be their primary source of support. At midlife the female continuous career artists were found to have greater personal income, a higher occupational prestige and significantly fewer children than those with interrupted careers. In a study of a subset of male artists, Stohs found that fine artists were more likely to identify intrinsic motives to explain why they were artists than were the applied artists.

Smith (2000) used the question in the 1970 US Census on the occupation held five years earlier and the respondent's occupation in 1970 to examine the movement of people into and out of artistic occupations. He found that slightly more than 70 percent of the artists in 1965 persisted as artists in 1970. Approximately one-third of those who were no longer artists were working in other professional occupations<sup>3</sup> in 1970 with the others distributed among the non-professional occupations. There were also new entrants into the arts over the five year period. They made up approximately 0.3 percent of those in non-arts occupations in 1965. Almost 55 percent of those who were artists in 1970 had also been artists five years earlier. Almost one-quarter of the artists in 1970 had not been working at any job five years earlier.

<sup>3</sup> The US Census Bureau classifies the artist occupations among the professional occupation category. This category includes other professional workers such as doctors, lawyers and professors of economics.

Smith's primary focus was the determinants of artists' earnings over their careers. He found that having been an artist in 1965 leads to significantly lower earnings (approximately 15 percent less) for people who were no longer artists in 1970. Overall work experience has the inverted 'U' shape effect on earnings for those former artists working in non-arts jobs suggesting that the skills learned while working as an artist may not be readily transferable to non-arts jobs. He found that experience in the arts and the apparent development of arts specific human capital had a positive and significant effect on arts earnings.

In a similar study, Alper and Wassall (1998) utilized the first round of the National Science Foundation's National Survey of College Graduates (NSCG) to examine the determinants of persistence in arts occupations. Like the census data used by Smith, the information in the NSCG was collected in the first interview round as part of a retrospective survey. Information was obtained for 1988 and 1993. Data from the 1990 Census was also part of the NSCG data so that there were three observations on these individuals over this five year period. Those who had worked as artists in any of the three years were included in Alper and Wassall's study.

Over the five year period artists' careers were found to be not that much less stable, at least in terms of occupational changes, than many other occupations. Approximately three-quarters of the artists in 1988 were still artists five years later. This is slightly larger than what Smith found and may reflect that these were artists with at least bachelor's degrees. In the majority of the non-arts occupations examined by Alper and Wassall, between 70 and 85 percent persisted in their occupations over the five year period.

When these artists did change occupations more than one-quarter entered managerial and executive occupations and fewer than two percent were working in food service occupations. An equal amount, approximately ten percent, worked in sales and clerical occupations. They also found a larger percentage of artists left for better pay/promotion or for a career change than did other professionals who left their jobs over the same time period.

There was a considerable inflow of people into the arts occupations over the five year period as well. Fully one-third of the artists in 1993 were not artists in 1988. Approximately one-third of the new entrants came from the stereotypical artists' non-arts occupations, e.g., sales, clerical and food service. People were considerably more likely to indicate that they entered an arts occupation to change their career than were those who entered other professional occupations. There was a smaller, but still considerable, difference in the proportion of people entering arts occupations who indicated they did so to change their working conditions than among those who entered other professional occupations.

Alper and Wassall, using a probit model, examined the determinants of occupational persistence for artists over the 1988 to 1993 period. Several demographic characteristics, the artist's age and gender, were found to significantly impact persistence as an artist over the five year period. Male artists were more likely to persist than female artists. The likelihood of persisting tends to increase with age, but at a decreasing rate with artists older than 50 having an increasing likelihood of leaving the profession, perhaps into re-

tirement. Neither the artist's race, ethnicity nor change in marital status had any impact on persisting as an artist. The only measure of human capital investments that significantly impacted persistence was experience. Whether the experience was full-time or part-time did not matter, the more experience the greater the likelihood of persisting as an artist. There was no significant impact due to differences in the highest degree received or in being an art major. Even with only a limited number of identifiable artistic occupations in the NSCG data available for analysis, Alper and Wassall did find that being a post-secondary school teacher of art significantly decreased the likelihood of remaining an artist over this five year period. Other factors that had significant effects were having a spouse who works part-time, which has a positive impact on persistence, and having a disability, which has a negative impact.

### 3.3. *Quasi-panels*

Bielby and Bielby's (1987, 1989, 1993b, 1998) analysis of the Writers Guild of America, west (WGA) membership initially focused on issues related to gender, age and minority status. They started with data from 1982<sup>4</sup> and continue to study these authors today. Over time their focus shifted from analyzing earnings and employment conditions at a point in time to changes over time and the writers' careers.

The membership of the WGA, west, is inclusive of most authors who write for television and film.<sup>5</sup> The WGA's data on these writers is limited to a few basic demographic characteristics (age, gender and ethnicity) and to the earnings received in activities covered by the union's agreements with television and film producers. This limits the conclusions. Without a broader set of information, especially on their other income generating activities, the Bielbys only have a partial picture of what these writers are doing at any point in time and, therefore, over time as well. Since less than half the union's membership works in the industry in any given year, the absence of information on what they are doing when not employed in television or film is difficult to overcome. Therefore, the Bielbys pool the information on the annual panels of WGA members rather than follow the behavior of individual members over time.

In their earliest reports the Bielbys found that women and minorities were disadvantaged relative to the white male members of the WGA. This was true from both the perspective of earnings and employment. While this has changed over time, differences still exist. Throughout the 1980s they estimated that the earnings gap between all female and white male writers ranged between 25 and 40 percent. In the 1990s, through 1997, the gap narrowed to between 10 and 15 percent. The gaps were larger in film than in television. Minority earnings in the 1980s were approximately one-half the earnings of white males, but in the 1990s they were approximately the same. In recent work they

<sup>4</sup> The first year the WGA, west computerized its data.

<sup>5</sup> Television writers do not include those who write for news, sports, other non-fiction shows, game shows and daytime dramas. Film writers do not include those who own the copyrights to their material.



found a much smaller gender gap in earnings between white male and female writers who are recent entrants into the occupation. The employment situation for female writers, as a percentage of total employment, has not changed much over the time period, especially in film and in television since the early 1990s. Minority employment in the industry increased by 66 percent during the 1990s, bringing it to seven percent of total employment.

Recently the Bielbys identified significant changes in the relationship between the writer's age and his/her likelihood of being employed as well as in the writers' age-earnings profiles. The biggest changes in the likelihood of employment occurred from the late 1980s through 1997. The biggest change was the decreased likelihood of older writers finding employment. Between 1987 and 1997, the proportion of writers 51 to 60 years old employed decreased by one-third. For writers in the two older cohorts (61 to 70 and 71 to 80 years old) the proportion employed decreased by 50 percent. The writers in the 30 years old and under cohort were the only ones to see the likelihood of being employed increase.

The change in age-earnings profiles over the 1982 to 1997 period was most profound among writers employed in television and much more modest for those employed in film. For the television writers the transition was from the traditional inverted 'U' shape in 1982 with a peak for writers 51 to 60 years old, to an 'M' shape in 1987 with dual peaks for writers 31 to 40 and 51 to 60 years old, to a relatively flat inverted 'U' in 1991 with little difference in median earnings for writers in the three cohorts between 31 and 60 years old, and ended in 1997 with another highly peaked inverted 'U' shape, as in 1982, but with the peak earnings for writers who were 31 to 40. For writers in film the pattern of change was similar but did not reflect as much change in the age of the peak earners. In 1982 the inverted 'U' shape identified the film writers 41 to 50 years old as having the peak earnings while in 1997 it was those 31 to 40, with the writers in the 41 to 60 cohorts earning only slightly less.

The Bielbys examined various career related issues through the estimation of pooled earnings functions that followed the basic form:

$$\ln Y_{ict} = \alpha + \beta_1 X_i + \beta_2 W_{it} + \gamma_c + \delta_t + \varepsilon_{ict},$$

where  $Y_{ict}$  is the WGA earnings for the  $i$ th artist who in the  $c$ th cohort<sup>6</sup> in year  $t$ ,  $X_i$  are time invariant characteristics of the artist (e.g., gender),  $W_{it}$  are traits that vary for the  $i$ th artist over time (e.g., experience),  $\gamma_c$  captures the effects that are unique to a specific cohort,  $\delta_t$  captures year specific effects (i.e., a vector of dummy variables for the year) and  $\varepsilon_{ict}$  is a vector of random errors [Bielby and Bielby (1993a, p. 14)]. They found that, regardless of the writer's age, experience in the industry has a positive impact on earnings, but that having been recently employed in the industry (i.e., within the past three years) overwhelms the importance of overall experience in determining writers' earnings.

<sup>6</sup> Was admitted as a member of the WGA in year  $c$ .

Third parties represent many writers, and other artists.<sup>7</sup> The Bielbys found that representation had a significant impact on both earnings and employment (through the estimation of a logit model for the probability of employment) and that the type of representation was also important. They separated “core” agencies<sup>8</sup> from all other agencies that represent writers and found that writers with “core” agency representation had an increased likelihood of being employed and significantly higher earnings than all other writers. When examining earnings data from 1982 to 1990 they also determined that the disadvantage faced by female writers was continuous over their careers at a constant differential and not cumulative.

Throsby and his colleagues (1986, 1989, 1994, 1995, 2003) have been studying Australia’s artists through direct surveys since the early 1980s. The first survey was undertaken in 1983, was repeated in 1987, 1993 and 2002, but does not represent a true panel as each survey involved a different group of artists. Its advantage over the Bielbys’ research is that Throsby et al. utilized a data collection instrument designed for artists. A considerably wider range of information tailored to the artists’ activities, their earnings and work experiences was therefore obtained.

The findings from the surveys suggest some significant changes in Australia’s artists and the career paths they have apparently taken. First and foremost there has been a considerable change in gender composition. In 1983 a bit more than one-third were female while in 2002 they were evenly split. Over the same twenty year period, the artists were getting older (mean age of about 46 years compared to 41.5) while the Australian labor force was staying about the same (37.5 years old in 2002 compared to 38.2 in 1983). They were also more likely to be native born Australians in 2002 than in 1983 (74 percent compared to 67 percent). In examining artists’ training, an important aspect of the development of artists’ careers, they found that the average time spent training went from about five years to slightly more than four (4.3) in 2002. While the time spent training showed a slight decline, the average age at which the artists start working as professional artists increased from 23 in 1983 and 27 in 1993 to 30 years old in 2002.

It appears that Australia’s artists spent considerably more time working at their art in 1993 than they did in 1983. This includes working in their primary artistic occupation, in another arts field and in arts related work, a circumstance that is not unusual for many artists. The increase in hours was approximately 25 percent over the ten year period. This increase in hours working in the arts led to a change in the proportion of time working in the arts, as well, from just over 71 percent to slightly more than 82 percent in 1993. Over the 1993 to 2002 period the time spent in the arts declined by almost eight percent and the proportion of their total time spent working in the arts decreased slightly to 81 percent.

<sup>7</sup> In the US unionization is relatively common among the performing artists, such as actors and musicians, and very uncommon among most other artists.

<sup>8</sup> “They negotiate unique arrangements with the talent guilds and cultivate long-term relationships with those who finance, produce, and distribute new projects.”

The artists' increased time and effort working apparently did not pay off in terms of earnings. Throsby (1986) reported that in 1983 artists' earnings from their art was approximately 60 percent of the earnings for all workers in Australia (full-time employees in their main jobs). In both 1993 and 2002<sup>9</sup> the artists' artistic earnings were only 40 percent of the earnings of all workers. If arts related earnings are included the difference was not quite as large, 70 percent in 1983, 60 percent in 1993 and 57 percent of the average earnings for all workers in 2002, but it still shows a decline in the relative earnings of Australia's artists. A comparison of total earnings from all sources shows that in 1983 artists earned approximately six percent more than the average Australian worker; in 1993 they earned only 80 percent of the average Australian's mean earnings. The period from 1993 through 2002 saw a slight improvement in overall earnings to the point where artists' mean earnings were 87 percent of the average full-time Australian employees' earnings.

In Australia, as in much of the rest of the world, artists tend to hold more than one job during the year in order to earn the income needed to survive. The proportion of artists who were multiple jobholders, decreased from almost three-quarters (72 percent) of the artists in 1988 to under two-thirds (63 percent) in 2002.

As far as career development, Throsby et al. find that artists are starting their professional careers increasingly later in life. In 1983 the mean age for the start of their professional careers was 23. A decade later it was 27. In 2002 it is 30. The increase was not uniform across all arts occupations. The 'community artists/community cultural development workers' had the largest relative increase in the average age at which they started their careers. The 'crafts practitioners' had the smallest increase. Perhaps this reflects differences in the relative importance of the formal schooling and training needed to enter these professions. Throughout this 20 year period it was the dancers who started their professional careers at the youngest age (19 in 1983 and 24 in 2002) and the writers who started at the oldest age (28 in 1983 and 37 in 2002).

### 3.4. True panel data

Rengers (2002) was interested in determining the best model to explain artistic careers. He compared the traditional human capital model, in which the artists' careers are affected by their intrinsic and learned qualities, to the winner-take-all model, in which the relative differences among artists are the most significant factor in explaining the differences in careers. He utilized data from several panels. One panel was comprised of 540 art college graduates in the Netherlands who entered the labor market between 1993 and 1995. Rengers had two observations on their activities. The first observation was approximately one and a half years after graduation and the second was six years

<sup>9</sup> In general, financial information from surveys and censuses are for the year prior to the survey year. For Throsby's 2002 survey the financial year extended from July 1, 2000 to June 30, 2001. In general the survey or census year will be used to identify all the data collected at that time.

later. Another panel represented all the visual artists in the Netherlands who had been surveyed annually starting in 1993. He used the results from four surveys of approximately 500 visual artists for the years 1993 through 1996. A third panel was comprised of 575 visual artists and covered the period 1980 to 1991. While Rengers' analysis was based on a panel, his ability to generalize and fully explore artists' careers was limited because some panels only covered a short period of time while others were limited to one group of artists.

The models estimated were for labor supply, as measured by hours of work, wages and gross earnings as well as for two non-monetary measures of artistic achievement. Rengers' results provided mixed support for both models. He found that inequality in hours, wages and earnings diminished over time, which would be consistent with the human capital model not the winner-take-all model. He did not find any evidence that the art school graduates who left the arts for non-artistic occupations were penalized in terms of their earnings, as one might have suspected in the winner-take-all model. He did find that differences in school location and field of study, measures of signals apparently used by the market, were important in explaining the variation in wages and he believes this provides support for the winner-take-all model. Additionally, Rengers finds that the winner-take-all model better explains the variation in the non-monetary measures he explores, one being the artist's assessment of his/her own reputation and the other a measure of whether the artist received the attention of the media.

Rengers also provides evidence from the Dutch artist samples for some overarching themes found in the literature on artists' careers. Dutch artists started their education and training at young ages, often outside formal schooling, and the vast majority received arts degrees (90 percent), but begin their arts careers relatively late in life because of the amount of education they receive. Their earnings early in their careers tend to be less than that of non-artists with the same amount of education because artists work fewer hours, not because they receive a lower wage. Characteristics of the artist's education were found to have little or no impact on the artist's career. Self-educated artists have the same earnings and the same supply behavior as those with formal arts education and the prestige of the arts college attended does not have long-lasting effects. Female artists' earnings are about one-third less than their male colleagues, a difference that is comparable to the non-arts labor market in the Netherlands. Six years after graduation approximately two-thirds of the artists work exclusively as artists and about an equal proportion work both in and outside the arts (approximately 13 percent). Both age and experience have positive impacts on a variety of measures of success including participation, wages, sales of art works, prices of art works and the frequency of exhibiting in art galleries. Receiving government grants has a positive impact on artists' earnings late in their careers.

Perhaps unique to the Dutch experience, Rengers identifies two distinct career paths for artists. One is the government market and the other is the private market. He finds that the majority of visual artists (approximately 60 percent) participate in both markets with the smallest proportion (approximately 15 percent) with earnings only from the public market and that getting established in a career in the private market takes longer

than getting established in one in the public market but that in both markets success breeds greater success.

#### 4. An empirical investigation of artists in the US: 1940–2000

This section analyzes the employment and earnings of American artists using decennial US Census data from 1940 to 2000. Since it is based on seven unique cross-sections of US artists it is most comparable to Throsby's work and can be categorized among those studies that are quasi-panels via separate cross-sections.

The information from the Census public use samples<sup>10</sup> will be used to examine various aspects of artists' employment patterns, earnings and earnings variability including the estimation of earnings functions for artists for census years between 1950 and 2000. The use of Census data over a 60 year time span provides a clearer perspective of three related issues that have been debated in the economics literature on artists: (1) As working professionals, to what extent have artists fared less well than comparably educated persons in other disciplines? (2) Has the oft-reported disparity in earnings between artists and comparably educated groups grown or shrunk over time? (3) Is this apparent "earnings penalty" due to the characteristics inherent in the nature of the artistic labor market? For example, are artists unusual risk-takers? Or are they unable to properly assess their likely outcomes in a job market increasingly dominated by "winner-take-all" characteristics? This section will not address a fourth important issue: the decision processes of those artists who routinely moonlight, both inside and outside the artistic profession.<sup>11</sup>

This section starts with (A) a discussion of the use of US census data to study artists. It is followed by (B) an analysis of growth in the artist profession between 1940 and 2000, and (C) a summary of the labor market for artists over the same period. The last two sections examine (D) the level variability in artists' earnings, compared to other professionals, and (E) the determinants of their earnings, and the existence of an artistic earnings penalty.

<sup>10</sup> The US Census Bureau as part of its decennial census collects detailed socioeconomic information from a sample of the population using the 'long-form questionnaire'. This information is made available to researchers through data sets that represent either 1 percent or 5 percent of the US population. They are known as the Public Use Microdata Sample or PUMS. The Census Bureau first created PUMS data from the 1960 Census, a single one percent sample was generated. In 1970, six independent one percent samples were generated; three were based on a long-form that went to five percent of the population, and three were based on a long-form that went to fifteen percent. This study uses data from four of these one percent samples to calculate descriptive statistics. In the regression analysis, all three one percent samples from the five percent survey are used. In 1980, three samples were generated: a five percent sample, used in this study, and two one percent samples. In 1990 and 2000, two samples were generated: a five percent sample, used in this study, and a one percent sample. Researchers at the Center for Demography and Ecology at the University of Wisconsin, in collaboration with the Census Bureau, reconstructed the one percent PUMS for 1940 and 1950.

<sup>11</sup> A detailed discussion of the multiple jobholding, moonlighting behavior of artists can be found in [Alper and Wassall \(2000\)](#).

#### 4.1. Using census data

A major constraint facing social scientists who study artists is the paucity of labor market data bases containing sufficiently large numbers of them. The artist labor force in the US is small having grown from 0.7 percent of the entire civilian labor force in 1940 to over 1.4 percent in 2000 (Table 2). Of the cross-sectional and quasi-panel data bases available to researchers in the United States, only the decennial Census Public Use Microdata Samples (PUMS) are large enough to permit meaningful statistical analyses. As a consequence, the information on artists presented herein utilizes public use samples from the 1940 through 2000 censuses. The actual number of artists contained in these data ranges from 3863 in 1940 to 109,469 in 2000, but they represent the entire population of artists in each census year.

Extensive information on housing and personal characteristics of household members is contained in the samples. The information collected in each decade has changed somewhat; a few questions have been dropped or added, and the definitions of terms used in some continuing questions have been altered. Nevertheless, there remains a substantial core information base across all six decades that can be utilized for comparative analyses.

Defining who is an artist can be problematic. Until recently, the National Endowment for the Arts (NEA) used the following eleven Census occupations: (1) actors and directors, (2) announcers, (3) architects, (4) post-secondary art, drama and music teachers, (5) authors, (6) dancers, (7) designers, (8) musicians and composers, (9) painters, sculptors, craft artists and printmakers, (10) photographers, and (11) artists not elsewhere classified (*nec*). They were originally located within the category called “professional and technical workers”. Since the 2000 Census this category has been narrowed to “professional workers”. Also, after the 2000 Census the NEA changed its artist definition.<sup>12</sup> The eleven occupations now defined as artistic are: (1) actors, (2) announcers, (3) architects, (4) artists and related workers, (5) authors, (6) dancers and choreographers, (7) designers, (8) entertainers and performers, sports and related workers, (9) musicians, singers and related workers, (10) photographers and (11) producers and directors [US Census Bureau (2003)]. Table A.1 in Appendix A provides a brief evolution of these categories, and reports on the sizes of the samples that were extracted for each artist category.<sup>13</sup>

<sup>12</sup> The Census no longer separately identified post-secondary art, drama and music teachers, and the actors and directors occupation was separated into two groups, one for the actors and the other for the producers and directors. The painters, sculptors, crafts artists and printmakers occupation was subsumed into the new artists and related workers occupation which also includes some artists who were previously in the artists not elsewhere classified occupation. The artists *nec* has been replaced by the narrower entertainers and performers, sports and related workers *nec* occupation.

<sup>13</sup> The definitions of many of the artist occupational categories used by the Census underwent changes. The major changes were as follows. The category “showmen” appeared in the 1940 Census only. Directors were not lumped with actors until 1980 (and were mainly classified as managers, and not artists, prior to 1980) and

The availability of Census public use samples to study the labor market behavior of artists is, however, a mixed blessing. Three principal issues arise from the Census data gathering methodology that affect the information published and its interpretation.<sup>14</sup>

One is that the Census artistic occupational categories used by the NEA and most researchers appear to be very inclusive. The Census dancer category incorporates ballet dancers, Las Vegas showgirls and barroom strippers, and the author category includes magazine feature writers as well as poets. The NEA definition of artists includes entire occupational categories that some would argue lie outside the fine arts, such as architects, designers, and radio and television announcers. However, in one respect, it might not be as inclusive as perhaps it should be. Prior to 2000, college and university teachers of artistic subjects were included among the artists, but not their counterparts at the elementary and secondary school levels; currently both groups are excluded from the definition. The eleven Census categories used by the National Endowment for the Arts define the artist profession in what follows.

The second and more critical issue is that the Census Bureau's methodology obscures important aspects of artists' occupational choices. The artistic occupation is unusual in that many of those in it also hold non-artistic jobs, i.e., they are multiple job holders. It has been frequently documented, using direct economic surveys of artists, that most artists report working in non-artistic jobs in a given year [e.g., Alper and Wassall (2000); Kingston et al. (1981, 1986); Throsby and Hollister (2003); Wassall and Alper (1984, 1985, 1990, 1992); Wassall, Alper and Davison (1983)]. Those who report working in non-artistic jobs also report that a significant percentage of their time is devoted to working in these other occupations, and that a significant percentage of their total earnings, typically well over half, derives from this non-artistic work.<sup>15</sup> However, the Census requires that a person filling out its long form choose a single occupation. This choice

in 2000 they were removed and placed in a separate category with producers, but the combined category was retained among the artist categories. Announcers was not a separate category until 1970; prior to that, they were not enumerated as artists at all. The category of art, drama and music teachers did not exist prior to 1970; before 1970, these artist/teachers were classified with their artist counterparts (e.g., "musicians and music teachers"). In 2000 the post-secondary school teachers of art category was no longer separately identified and therefore could not be included in the artist definition. Prior to 1980 window dressers were a category separate from designers; in 1980 both were lumped into the designer category. Other artist occupations experienced small changes over this period. The sole exception is authors, whose definition remained unchanged over the period. Citro and Gaquin (1987) have estimated the total artist population for the census years of 1950 through 1980 using reconstructed artist occupational categories based on the 1980 definitions cited above. Their estimates of total artists in these years are slightly different from ours, which are directly projected from Census Public Use Microdata Samples.

<sup>14</sup> For a more exhaustive discussion of the strengths and weaknesses of census data on artists see McNertney and Waits (1989) and Wassall and Alper (1992).

<sup>15</sup> Kingston, Cole and Merton (1981), in their survey of 2241 American authors, report that while the median 1979 income from the writings of these authors was \$4800, their median personal income from all sources was \$27,000. In a survey of 494 American composers, Felton (1978) reports a median family income of \$20,000 in 1974, but a median income of only \$168 from composing. Wassall, Alper and Davison (1983) asked 3027 New England artists to classify their 1981 work time and earnings into artistic, arts-related, and non-arts-related. With a definition of artistic occupations that was more narrowly defined than in the Census it was found that

is based on time spent at work during a single reference week. Once this occupational choice is made, all time spent working and all earnings in the year prior to the Census become attributed to that occupation.<sup>16</sup>

Nevertheless, the Census method of defining an occupation based on one's principal activity in a reference week is appealing to economists. As Filer (1988) notes, "a worker who spent 20 hours every week producing paintings and 30 hours a week teaching would be classified as a teacher. Conversely, a worker who spent 20 hours every week in the classroom and 30 at her easel would always be a painter . . . Census definitions result in a bias towards including only those who achieve the most success in their art form as artists."

The use of a market test to define who is an artist troubles some observers. Regardless of one's position on this issue, one must be aware that, when working with aggregated Census data on artists, unlike most other professions, an unknown but significant percentage of the reported work effort and earnings will be from other, non-artistic occupations. The dynamic choices members of this unique profession make in dividing their work time among diverse occupations cannot be observed using Census data.

The third issue is the Census policy of top-coding certain variables for reasons of confidentiality, including the individual's age and the individual's income and earnings. Most importantly, in the 2000 Census an individual's reported wage and salary earnings is capped (top-coded) at a maximum level of \$175,000.<sup>17</sup> All those individuals earning above this level are assigned the state's mean earnings for those earning above the top-code. For example, all Massachusetts residents who had wage and salary earnings in 1999<sup>18</sup> above \$175,000 would have their earnings reported as \$322,000. Thus the

only 24 percent of the survey respondents held artistic jobs alone during the year. Of the artists who held other types of jobs as well, 52 percent held artistic and arts related jobs, 29 percent held artistic and non-arts related jobs and 19 percent held all three types of jobs. Of the artists' average labor earnings of \$14,079 in 1981, 46 percent was derived from artistic work, 34 percent from arts-related work and 20 percent from work which was neither artistic nor arts-related in nature. In this study teaching and coaching were defined as arts-related work. Direct surveys of artists typically employ a more narrow definition of "artist" than the Census. While they contain a higher percentage of "traditional" artists, they incorporate some persons who spend little time and earn little money at their art, but who define themselves as artists. By defining artists' occupations more narrowly, these studies may also attribute a smaller percentage of their total earnings to their art work. Also, the low percentages of earnings from artistic work cited in the studies above are due in part to their not counting earnings from teaching in one's profession as part of artistic income.

<sup>16</sup> An intriguing question is: What choice do artists make when confronted with the reference week work question? In direct surveys of artists, it is common that they signify being an artist is their principal profession despite little evidence of work time or of financial success. Their artistic occupation often is more prestigious than any other stated occupation(s). Do artists behave similarly when filling out the Census form, or do they answer the reference week work question literally?

<sup>17</sup> Earnings in each category are top-coded in every year: at \$5000 for 1939, at \$10,000 for 1949, at \$25,000 for 1959, at \$50,000 for 1969, at \$75,000 for 1979, and at \$90,000 in 1989. In 1999 there were top-codes for each type of income with wage or salary income top-coded at \$175,000, self-employment income top-coded at \$126,000 and total earnings top-coded at \$310,000. Summing the categories to determine total earnings reduces the compression somewhat, but it is still true that actual means and standard deviations will be underestimated.

<sup>18</sup> US Census data on earnings and time spent working are for the calendar year prior to the census year.



earnings of the highest paid members of any occupation may be underestimated. These earnings caps have risen from decade to decade. To a small but unknown extent, the stated mean earnings and degree of earnings inequality in the artistic (and any other) profession will be underestimated using Census data.

#### 4.2. *Size and growth of the artist profession: 1940–2000*

The growth in the artist labor force during the period 1940–2000 has been dramatic. In 1940, there were 386,000 artists in the US labor force, or 0.7 percent of all its members. By 2000, the number of artists had increased fivefold to 1,931,000, or 1.4 percent of the labor force.<sup>19</sup>

During the post-World War II period, the growth of the civilian labor force was fueled by rapid growth of the service sector. Most professional workers, and most artists, work in service industries. A second trend abetting the post-war growth in the civilian labor force has been increasing labor force participation rates of women. The growth in both male and female professional and technical workers (and artists as well) substantially exceeded the growth in the labor force as a whole; part of this rapid growth reflects a higher proportion of women who work in these professions. Even though women have been disproportionately represented in the professions, professions other than artists have typically contained a higher proportion of women.

These trends are reflected in the relative growth rates of the three labor force categories in [Table 2](#). Between 1940 and 2000, the civilian labor force grew by 166 percent. The professional labor force, however, grew at a rate of 542 percent over this period. The artist labor force increased by 400 percent, somewhat less than the professionals overall but considerably greater than the labor force. In recent years, the growth in artists has been more pronounced. Between 1980 and 2000, for example, the professional labor force increased at 103 percent while the rate of increase of the artistic labor force was 77 percent.

The growth in artists is drawn from Census information published after each Census; thus they incorporate the differences in the definition of artists noted above and in [Table 2](#). In the sections which follow, the implications of changing occupational definitions of artists over time are minimized primarily by analyzing all artists as one broad occupation. There still exist less important definitional changes which over time have added to or subtracted from the artistic labor force. These, noted above, primarily affect the number of directors and announcers enumerated as artists from 1960 to 1980. The exclusion of the post-secondary school teachers of art from the 2000 Census is likely to have an impact, but the extent is not clear.

<sup>19</sup> Authors' calculations and Ellis and Beresford (1994, p. 6).

Table 2  
Growth in artists vs. growth in the civilian labor force and all professional workers: 1940–2000 (numbers in thousands)

Category	1940	1950	1960	1970	1980	1990	2000	Change (%) 1940–2000
Civilian labor force	51,742	58,999	67,378	79,802	104,058	122,473	137,669	166.1
Professionals	3879	5081	5543	8800	12,275	16,648	24,905	542.0
Artists	386	441	492	737	1086	1671	1931	400.3

Source: US Bureau of the Census (1975, Series D 182–232, p. 139 and Series D 232–682, p. 140) for 1940–1950; Citro and Gaquin (1987, Table II.1), for 1960. Ellis and Beresford (1994) for 1970–1990. 1997 Statistical Abstract, Table 645, for 1996; authors' calculations.

#### 4.3. Labor supply and labor market characteristics

In examining the labor market outcomes of artists, some perspective can be gained by comparing them to an appropriate reference group of workers. Most Census-based studies have compared artists' labor market outcomes to a reference population. The choice of reference population has not been consistent, ranging from specific occupations with comparable educational attainment [Santos (1976)] to specific professional occupations [Waits and McNertney (1980)], to all workers [Filer (1986)], and to all managerial, professional, and technical workers [Filer (1988, 1989)].

This research lies in a similar vein. The conclusions drawn about artists are compared to those for all professional workers, excluding artists. All other professional workers were chosen as a reference group because (a) they possess similar demographic and socioeconomic characteristics, (b) they comprise the category in which artists were initially classified by the Census and (c) by choosing all other professional workers, one is unlikely to bias comparisons to artists as might occur by selecting specific occupations.

Some basic comparisons are shown in Table 3. In this table, labor force and socioeconomic characteristics of both artists and professional workers, excluding artists, are reported and compared.

In many respects artists closely resemble members of other professions. Some distinctions, however, are apparent. Perhaps most important is the additional years of schooling that other professional workers have achieved relative to artists. This difference has narrowed from 2.5 years in 1940 to less than one year in 2000.<sup>20</sup> Another less obvious (and perhaps less expected) difference is that the artist labor force has, for

<sup>20</sup> The comparison of educational levels for 1990 and 2000 to prior census years is weakened because since 1990 the Census has reported on highest degree or level of schooling completed and not years of schooling. Using data on both years of schooling completed and highest grade/degree completed for 1990 collected by the US Census Bureau [Kominiski and Siegel (1993)] the authors developed a method to convert highest grade/degree completed to years of schooling based on the average number of years it took to complete the grade/degree.

Table 3  
 Labor market characteristics of artists and other professionals in the experienced civilian labor force: 1940–2000 (median in parentheses)

	Age	Years of school	Women (%)	White (%)	Black (%)	Unemp. (%)	Worked in 1999 (%)	Self-emp. (%)	Full-time*	Hours worked*	Weeks worked*
1940											
Artists	37.9	11.7	33.0	95.4	4.2	11.0	92.1	33.4	39.2	30.1 (40.0)	38.3 (50.0)
Prof.	38.3	14.2	43.7	95.9	3.9	3.0	96.0	16.3	49.4	39.4 (40.0)	42.6 (52.0)
1950											
Artists	37.8	12.8	34.7	95.6	4.0	4.9	91.8	26.5	25.5	33.9 (40.0)	40.1 (51.0)
Prof.	38.9	14.4	35.2	95.5	4.2	1.4	94.5	12.8	41.6	40.9 (40.0)	42.7 (52.0)
1960											
Artists	40.2	13.7	38.6	96.1	3.0	4.0	98.4	28.2	42.9	33.2 (40.0)	41.8 (51.0)
Prof.	40.0	15.0	38.2	95.0	4.1	1.2	98.5	10.4	57.6	39.8 (40.0)	44.1 (51.0)
1970											
Artists	38.6	13.8	29.7	94.7	3.5	4.4	98.1	22.4	50.1	33.9 (40.0)	42.9 (51.0)
Prof.	39.0	15.2	41.1	93.0	5.5	1.7	98.2	8.1	52.9	37.3 (40.0)	43.9 (51.0)
1980											
Artists	36.8	14.5	38.2	92.1	4.3	5.5	95.8	30.1	49.6	35.5 (40.0)	42.1 (52.0)
Prof.	38.0	15.7	47.4	88.7	7.3	2.0	98.1	7.6	58.8	38.5 (40.0)	44.9 (52.0)
1990											
Artists	38.3	14.4	44.3	89.9	4.5	4.8	96.6	31.3	53.1	37.0 (40.0)	43.3 (52.0)
Prof.	39.2	15.3	52.7	86.2	8.0	2.1	98.3	7.5	61.9	39.6 (40.0)	45.9 (52.0)
2000											
Artists	40.2	14.7	45.7	86.2	4.7	4.5	96.8	31.5	56.0	37.9 (40.0)	44.2 (52.0)
Prof.	40.9	15.5	56.6	81.5	8.5	2.2	98.3	6.6	59.8	39.4 (40.0)	45.7 (52.0)

Source: Authors' tabulations from the 1940–2000 Census PUMS.

\*In year prior to census year.

almost all the period, been composed of a higher percentage of men and whites than the other professions.

However, the most striking findings are related to the consistently poorer labor market outcomes of artists. Although the disparities between artists and other professional workers narrowed over the 1940–2000 period, they by no means disappeared. Across all seven census years, artists were more likely to be unemployed. Up to 1960, artist unemployment rates were at least three times that of other professional and technical workers; after 1960 artist unemployment rates were between two and three times greater. Artists also have consistently worked fewer weeks per year and fewer hours per week over this period. In each census year the proportion of artists working full-time, year-round (defined as working at least 50 weeks per year and 35 hours per week) was less than the proportion of professional workers. However, the gap between annual hours worked by artists and professional–technical workers fell over this period, with a difference of approximately 125 hours in 1999 down from 520 hours in 1939 and less than 200 hours in 1969. Similarly, the gap in percentage working full-time was lower between 1969 and 1999 than in earlier Census years.

Given that these labor market disparities persist across seven census years, any type of disequilibrium is unlikely to be the cause. Over a sixty year period, one would certainly expect any disequilibrium to have adjusted. This leads to a search for alternative explanations.

Several hypotheses have been advanced as to what characteristics make the artist labor market unique. First, it seems difficult to accept any explanation based on informational asymmetries; surely over sixty years artists and those training to become artists would recognize these disparities and alter their career choices.<sup>21</sup> It is possible that artists value leisure more highly than do members of other professional occupations. However, it seems unlikely that any single occupation would attract only those with a higher value for leisure. This “leisure theory” can be contrasted to Throsby’s “work preference” model of artist behavior [Throsby (1994a)], which postulates that the artist is driven to create, and will maximize time spent working as an artist subject to constraints of earning sufficient income, from either inside or outside the arts, to finance an acceptable level of consumption. On the one hand, this theory would imply that artists spend more time working than other workers who receive (greater) disutility from their work. This does not show up in the annual hours worked data. On the other hand, the theory also predicts that, for artists who supplement their arts earnings with jobs outside the arts, as wage rates for non-arts jobs rise, more time will be freed to create art. As the premium to a college or higher level education (which most artists possess) has risen over the last two to three decades, so has the gap in hours worked narrowed between artists and other professionals. This is consistent with artists being able to finance more creative time from higher non-artistic wage rates. Unfortunately,

<sup>21</sup> Compare this perspective to that of Towse (1992a), who argues that artists tend to enter the artistic labor market too frequently because they overestimate the likelihood of future success.

with all income attributed to the reported profession, it is impossible to test these relationships using Census data.

Another set of theories lies in the roles of risk-taking among artists and the rewards to those who rise to the top of their profession. Although these theories were mainly used to explain income distributions among artists, they also have implications for labor supply. This line of reasoning was initially raised by Santos (1976), who asserted that performing artists belong to a class of risk-taking workers who, unlike other workers, are willing to trade off a small chance at substantial financial rewards for a much larger chance of low earnings. Finding that artists' investments in training do not yield financial returns consistent with those in the labor market as a whole, he noted that "risk-preference and psychic income apparently prevail over financial considerations when considering the pursuit of a career in the performing arts" (p. 257). In a nutshell, too many persons pursue careers in the arts because of these two factors. Santos does not explain why artistic careers would disproportionately attract occupational risk-takers, however.

A related theory is found in the literature on the earnings of superstars [Rosen (1981); Adler (1985)]. The predicted effects of this theory on labor supply are more prominent in Frank and Cook (1995); they refer to superstar labor markets as "winner-take-all markets". They note that "market incentives lure too many contestants into winner-take-all markets, and too few into other careers" (p. 103). They attribute this outcome in part to contestants' overestimates of their talent. However, they add that as long as market entrants base their decision on expected earnings rather than marginal earnings, the number of entrants will be greater than that what is socially optimal.

In somewhat different ways, both approaches imply that the nature of artistic labor markets will induce too many entrants. These arguments are consistent with the observed data on employment found in the Census.

Benhamou (2000) has an additional set of hypotheses that do not focus on the artist. One focuses on the behavior of the arts institutions and organizations that employ artists rather than the artists' choices. The other focuses on the public policies that support artists, especially those who choose to be self-employed.

Finally, it is worth noting that ease of entry into artistic occupations combined with greater loss of work time due to more frequent transitions from one job to another (both inside and outside their occupation) have also contributed to the consistently lower labor market participation experiences of artists. Thus artists would likely experience greater friction in moving among jobs and, even in a world of perfect information, would spend more time unemployed than members of other occupations. Given the unique extent of multiple job-holding among artists, these phenomena can be viewed as occupational hazards of working in the arts, which must be tolerated even if known in advance. As noted, this conclusion could not be drawn from an examination of Census data alone.<sup>22</sup>

<sup>22</sup> Corroborating evidence may be found in Wassall, Alper and Davison (1983), who report that New England artists were unemployed more times in a year but for shorter durations than members of the labor force in gen-

#### 4.4. Earnings, calculated wage rates, and earnings variability

As noted above, two principal issues have been raised in the literature on artists' earnings. One is the existence and extent of an earnings penalty facing artists, compared to other workers of comparable education and skills. The other is the existence of unusual earnings patterns in the artistic labor market, such as greater earnings uncertainty and variability, relative to other occupations. The availability of Census earnings data over a 60-year period enables an extended comparison of earnings, and thus of these two issues. In this section, we examine earnings variability. In the following section, the existence of an earnings penalty is investigated.

In [Table 4](#) the current and real earnings (in 2004 dollars) and estimated hourly wages of artists and other professional workers are contrasted. Means as well as medians are reported since several categories have medians of zero.<sup>23</sup> We report wage and salary, self-employment, and total earnings separately.<sup>24</sup> The definition of labor earnings has remained consistent since the 1950 Census. In 1940, only wage and salary earnings were reported.

The information in [Table 4](#) shows that artists have earned less than other professional workers throughout this period. Both mean and median annual earnings of other professional workers have been higher in every Census year. Unlike disparities in time worked, differences in earnings have not narrowed over time. The difference in median earnings was 33 percent in 1949, and 30 percent in 1999. It has ranged from a low of 14 percent in 1969 to a high of 45 percent in 1979. Although there is no clear cut trend, the largest percentage earnings differences occurred in 1979 and 1989. The differences in mean earnings over the same period are similar, though slightly less. They range from 12 percent in 1969 to 30 percent in both 1979 and 1989.

A major part of the earnings differentials is attributable to artists working fewer hours per year. This is reflected in the generally smaller differentials in median wages between the two groups, and in the virtually equivalent mean wages. These figures suggest that a comparison of only full-time workers in the two groups would reduce earnings differences; in fact, it does.<sup>25</sup> This approach of limiting comparisons only to full-time workers

eral. It can also be found in [Menger and Gurgand's \(1996\)](#) study of performing arts in France. They document the role played by the French unemployment insurance system in the pattern of performers' employment over their careers.

<sup>23</sup> A median of zero exists for self-employment earnings because the majority of artists and professional and technical workers are employees and not self-employed.

<sup>24</sup> In the Public Use Microdata Samples, labor earnings are composed of (1) wage or salary income, (2) non-farm self-employment income, and (3) farm self-employment income. From 1970 to 1990 earnings are reported separately in each of the above categories. From 1950 to 1960 and again in 2000 earnings from (2) and (3) were collapsed into one category. In 1940 only wage and salary earnings were reported.

<sup>25</sup> For example, the median earnings of full-time year-round professional and technical workers were the same as those of artists in 1969, but exceeded artists' earnings by 15.2 percent in 1979, by 16.5 percent in 1989 and by 15.4 percent in 1999. The pattern of differences in mean earnings between the two groups was similar over the same time period.

Table 4  
 Mean earnings of artists vs. other professional and technical workers in the experienced civilian labor force: 1940–2000\* (in current and real (2004) dollars; medians in parentheses)

	Artists					Professional and technical				
	Wage & salary	Self-emp.	Total earnings	Wage	Below poverty (%)	Wage & salary	Self-emp.	Total earnings	Wage	Below poverty (%)
1940										
Current	905 (588)	NA	NA	0.78 (0.26)	NA	1271 (1000)	NA	NA	0.77 (0.56)	NA
Real	12,436 (8080)			10.72 (3.57)		17,465 (13,741)			10.58 (7.69)	
1950										
Current	1913 (1200)	746 (0)	2658 (2100)	1.57 (0.96)	NA	2510 (2400)	659 (0)	3170 (2800)	1.52 (1.35)	NA
Real	15,352 (9630)	5987 (0)	21,331 (16,853)	12.60 (7.70)		20,143 (19,261)	5289 (0)	25,440 (22,471)	12.20 (10.83)	
1960										
Current	3619 (2800)	1238 (0)	4857 (4000)	3.63 (2.49)	NA	4759 (4500)	1158 (0)	5917 (5000)	3.25 (2.70)	NA
Real	23,754 (18,378)	8126 (0)	31,879 (26,254)	23.83 (16.34)		31,236 (29,536)	7601 (0)	38,837 (32,818)	21.33 (17.72)	
1970										
Current	6612 (5400)	1688 (0)	8305 (7000)	5.65 (3.98)	4.9	8019 (7500)	1309 (0)	9342 (8000)	5.34 (4.36)	3.3
Real	34,411 (28,104)	8785 (0)	43,222 (36,431)	29.40 (20.71)		41,734 (39,033)	6813 (0)	48,619 (41,635)	27.79 (22.69)	
1980										
Current	9942 (6505)	2696 (0)	12,657 (9605)	9.13 (5.97)	7.6	14,882 (13,205)	1562 (0)	16,476 (14,005)	9.48 (7.60)	3.5
Real	26,156 (17,114)	7093 (0)	33,299 (25,269)	24.02 (15.71)		39,152 (34,740)	4109 (0)	43,346 (36,845)	24.94 (19.99)	
1990										
Current	18,985 (12,000)	4966 (0)	23,992 (18,000)	16.76 (10.38)	7.2	28,465 (25,000)	2616 (0)	31,117 (26,000)	17.31 (13.46)	3.3
Real	29,243 (18,484)	7649 (0)	36,955 (27,726)	25.82 (15.99)		43,845 (38,501)	4029 (0)	47,930 (40,048)	26.66 (20.73)	
2000										
Current	29,353 (20,000)	7238 (0)	36,590 (27,000)	23.81 (15.38)	6.9	41,235 (34,000)	3385 (0)	44,620 (35,000)	25.10 (18.75)	4.2
Real	33,652 (22,929)	8298 (0)	41,949 (30,954)	27.30 (17.63)		47,274 (38,980)	3881 (0)	51,155 (40,126)	28.78 (21.50)	

Source: Authors' tabulations from the 1940–2000 Census PUMS.

\*Year of data is for calendar year prior to census year.

is not pursued here, since a defining characteristic of artists' careers is that they work fewer hours per year than members of virtually all other professions.

Real earnings of artists and other professional workers increased from 1939 through 1969. The decade of the 1970s was a period of high inflation so by 1979 the real earnings of the artists and other professionals had decreased considerably. Unlike the other professions whose real 1999 wage and salary earnings were thirteen percent higher and whose real total earnings were five percent higher, the real earnings of artists, regardless of source, had still not reached their 1969 levels thirty years later. In fact their 1999 wage and salary earnings were two percent lower than in 1969 and their total earnings were three percent lower.

The other issue raised above addresses the question of a possible relationship between the poorer labor market outcomes that artists experience and greater earnings variability among artists than other professions. In the previous section, we reviewed two explanations of such a relationship. Santos' (1976) theory of artists as risk-takers suggests that artistic careers should be characterized by earnings with lower means and greater variability than those in comparable occupations inhabited by risk-averse workers. In fact, he presents evidence from the 1960 Census that the earnings of dancers and singers had higher coefficients of variation than other workers with the same amount of education.

Rosen's (1981) superstar theory also implies greater variability in earnings. However, it further implies that the distribution of income will become more skewed as (1) improvements in technology make it increasingly possible for the top superstars in each field to reap relatively greater rewards, due to the joint consumption characteristics of their product (book publishing and recordings, e.g.), and (2) the number of consumers or the intensity of their demands increases. Both these conditions appear to hold across artistic professions.<sup>26</sup> The superstar model does not predict that mean earnings will necessarily be lower than in comparable occupations without superstar characteristics, however.

Surprisingly few attempts have been made to test these hypotheses. Filer (1989) tested extensively for the presence of greater earnings inequality among artists in the 1980 Census. Using several measures of inequality, he found that "there is greater dispersion of incomes among artists than among the entire work force, or among managers, professionals, and technicians", but that "this difference does not appear to be large" [Filer (1989, p. 72)]. This difference in inequality seems attributable to two factors. One is the greater variability in annual hours worked by artists compared to the above reference groups. Differences in inequality were less when only full-time, year-round workers were compared. They were also less when artists were compared to selected, narrowly-defined occupations. Looking at three-digit occupations, he found that measures of earnings inequality "for occupations where individual talent and performance are important determinants of earnings tend to be similar to those for artists" (p. 74).

<sup>26</sup> Frank and Cook put it like this (p. 121): "The growing importance of winner-take-all markets thus implies a change in the pattern of incomes observed in the economy. More specifically, it implies that even if we control for age, education, experience, ability, and other individual characteristics thought to influence productivity and hence income, we should see greater income variability now than in the past."



To address this issue more comprehensively, in [Table 5](#) we compare various measures of earnings variability by comparing artists to other professionals from 1940 to 2000. In [Table 6](#), we make similar comparisons among selected occupations at the three-digit

Table 5  
Measures of low and high incomes and income variability, artists and professionals in the experienced civilian labor force: 1940–2000\*

		Percent with zero or less income from				Percent with income at maximum from				Variability in total earnings	
		Wage & salary	Self <sup>1</sup>	Total <sup>2</sup>	All <sup>3</sup>	Wage & salary	Self <sup>1</sup>	Total <sup>2</sup>	All <sup>3</sup>	CV <sup>4</sup>	Mean/median
1940	Artists	32.7	NA	NA	NA	2.0	NA	NA	NA	1.231	1.540
	Prof.	22.6	NA	NA	NA	2.6	NA	NA	NA	0.944	1.271
1950	Artists	33.7	73.8	12.1	9.9	1.4	2.1	3.5	3.8	0.940	1.266
	Prof.	20.5	84.5	8.5	6.9	1.3	2.7	4.1	4.7	0.766	1.132
1960	Artists	28.0	67.5	7.2	4.1	0.5	0.7	1.4	1.6	0.988	1.214
	Prof.	13.4	84.8	4.4	3.6	0.4	1.0	1.7	2.0	0.849	1.198
1970	Artists	19.4	77.9	3.9	2.9	0.3	0.3	0.7	0.9	0.951	1.186
	Prof.	8.4	90.3	2.4	1.8	0.2	0.5	0.6	1.0	0.845	1.168
1980	Artists	24.6	73.9	5.4	3.4	0.7	0.5	1.3	1.5	1.036	1.318
	Prof.	6.7	91.8	2.2	1.3	0.8	0.4	0.9	1.5	0.827	1.176
1990	Artists	23.5	69.9	5.0	3.3	1.7	0.6	2.6	3.1	1.132	1.333
	Prof.	5.9	90.5	1.9	1.2	2.6	0.9	3.6	4.2	0.924	1.197
2000	Artists	23.3	72.4	4.5	3.0	1.1	0.6	1.0	0.0	1.239	1.355
	Prof.	5.3	92.2	1.8	1.2	1.5	0.7	1.4	0.0	1.090	1.275

*Notes:* Reporting of self-employment earnings, total earnings and total income allows for negative amounts. Prior to the 1970 Census, self-employment earnings included earnings from farm sources; from 1970 on, they are from non-farm sources only. Total earnings include all wage and salary and self-employment earnings. All income sources were capped at a maximum amount by the Census. The maximum reportable amounts in each income category for each Census year were: in 1940, \$5000; in 1950, \$10,000; in 1960, \$25,000; in 1970, \$50,000; in 1980, \$75,000; and in 1990, \$90,000. In 2000 maximum reportable earnings/income varied by type of income and were: \$175,000 for wage and salary earnings; \$126,000 for self-employment earnings; and \$310,000 for total earnings. Total earnings were not directly reported in the 1950–1980 Censuses but were calculated by the authors. Income from all sources was also calculated by the authors. For 1940, earnings variability measures refer to wage and salary earnings only. See the text for more detail.

*Source:* Authors' tabulations and calculations from the 1940–2000 Census PUMS.

<sup>1</sup>Self-employment earnings.

<sup>2</sup>Total earnings.

<sup>3</sup>All income. Labor earnings and non-labor income.

<sup>4</sup>Coefficient of variation.

\*Year of data is the calendar year prior to the census year.

Table 6  
Professional occupations with greatest earnings variability: 2000<sup>1</sup>

Occupation	Mean earnings (\$)	Earnings of \$0 or less (%)	Earnings greater than \$175,000 (%)	Earnings greater than \$310,000 (%)	Coefficient of variation		Mean/Median	
					Rank	Value	Rank	Value
Actor	35,545	4.9	3.9	3.1	1	1.774	1	2.031
Announcer	31,739	3.5	1.9	1.7	4	1.580	8	1.580
Artist and rel.	30,427	8.2	1.2	0.5	8	1.305	14	1.415
Athlete	26,160	5.9	1.6	1.3	2	1.727	2	1.869
Author	40,093	6.5	2.9	1.6	6	1.381	12	1.463
Chiropractor	80,646	3.6	12.6	4.5	15	1.123	6	1.613
Dancer	19,709	7.5	0.2	0.1	14	1.141	*	1.314
Entertainer, nec	25,463	6.1	1.2	0.8	5	1.526	7	1.591
Health diag.	32,483	7.6	1.4	0.3	10	1.282	9	1.547
Library tech.	10,731	4.4	0.0	0.0	*	1.076	3	1.788
Misc. health tech.	34,642	2.0	1.6	0.9	11	1.214	13	1.443
Musician	25,323	5.4	1.2	0.6	3	1.596	4	1.783
Other teacher	18,930	5.2	0.2	0.1	9	1.285	5	1.661
Photographer	30,404	4.8	1.4	0.7	7	1.351	10	1.498
Podiatrist	110,813	1.5	20.7	7.1	*	0.906	11	1.478
Producer/Director	53,916	2.5	3.6	3.2	13	1.145	*	1.390
TV, movie camera operator and editor	41,914	4.0	1.7	1.2	12	1.152	*	1.397
Veterinarian	77,158	1.7	8.6	3.7	*	0.967	15	1.403

Source: Authors' tabulations from 2000 Census PUMS.

\*Occupation not in the top 15.

<sup>1</sup>Year of data is for calendar year prior to Census year.

level using 2000 Census data. Any comparisons using Census data are subject to some degree of bias because of the Census practice of top-coding all income categories. In both tables, earnings at the high and low ends of the distribution are examined, and two measures of earnings inequality are provided, the coefficient of variation and the ratio of median to mean. The initial comparison is limited to all artists and all professional workers.

The data in Table 5 show that earnings inequality among artists was greater than among other professional workers in all seven census years. Excluding 1939, the year in which self-employment earnings were not reported, both the coefficient of variation and mean/median measures show increases in inequality over time for both artists (as predicted by Rosen) and other professional workers. Earnings inequality among artists has grown at a faster rate than for other professional workers. In terms of the size of the tails of the earnings distribution, artists were more likely in every year to have total earnings (net of expenses) of zero or less than the other professional workers. The

likelihood of having total earnings at the maximum or top-coded level in each year was roughly equal for the artists and the other professionals.

To account for the possibility that non-earnings sources, such as royalty income, further contributed to inequality, the pattern of income from all sources was examined. This made little difference in the results. The same tests were further performed on full-time year round workers (not shown). Not surprisingly, the difference in inequality between artists and professional and technical workers narrows but does not disappear when only full-time, year-round workers are compared.

A detailed examination of the fifteen professional occupations with the greatest earnings inequality in 1999 is shown in [Table 6](#). The occupations were ranked using the mean/median and the coefficient of variation.

Although there are 123 professional occupations, nine of the eleven artist occupations show up in the top fifteen when ranked by coefficient of variation, and seven show up when ranked by mean/median. Regardless of the measure, actors have the greatest variation in total earnings. The three artist occupations with lowest earnings inequality, regardless of measure, are dancers, designers and architects. It is also true that within this group of high variance occupations the arts occupations generally have the largest percentage of members with earnings of \$0 or less in 1999 and the largest percentage with earnings above the Census established top-code (\$310,000).

It is also interesting to note the nature of the non-artistic occupations found in [Table 6](#). They do not generally correspond to ones expected to be listed among those in superstar or winner-take-all markets. Perhaps the sole exception is athletes, which ranked second using either variability measure.

#### 4.5. Earnings functions

To gain a deeper understanding of whether, how and why artists are different from other workers, and to more accurately determine any earnings penalty from being an artist, earnings functions for artists and for other professional and technical workers are estimated.

This is by no means the first use of earnings functions to explore possible differences in the rewards to education, training and other labor market attributes between artists and a reference group. Using 1980 US Census data [Filer \(1986\)](#) compared the earnings of artists to those of the general work force. [Filer \(1990\)](#), using the same data source, focused on the return to years of education among artists. Also, [Withers \(1985\)](#) compared the earnings of artists (collected from a special survey of artists) to the earnings of the general work force (using Australian Census data). In addition, this approach has been used to examine artists' earnings recorded in special surveys, as in [Snooks \(1983\)](#), [Wasall and Alper \(1984\)](#), [Montgomery and Robinson \(1993\)](#) and [Throsby \(1992, 1996\)](#).

In the earnings studies comparing artists to a reference group, some consistent findings emerge. Artists do not seem to fit the standard earnings model as well as other workers, and earnings functions for them have poorer goodness-of-fit [[Filer \(1986\)](#)].

The return to education for artists is lower [Filer (1986); Withers (1985)], or even negative [Withers (1985); Throsby (1992)].

As discussed in previous sections, what makes this research unique is the estimation of earnings functions from a consistent data source over a period of sixty years. Earnings functions for artists and other professional workers are compared using samples from the 1950 through 2000 Census years.<sup>27</sup> Selected results from these sets of earnings functions are presented in Table 7 and the definitions for the variables in the model are presented in Table 8. In each census year, identical functions for artists and for professional and technical workers are constructed. Across census years, the variables in these functions are essentially the same; however, some variables were not available in all six census years.<sup>28</sup>

In general, the earnings function coefficients are consistent both across census years and between the two occupational groups. Conventional wisdom and prior research suggests that the determinants of success for artists should be harder to quantify, and this is confirmed here. In four of the six census years the goodness-of-fit for the estimated equation (as measured by adjusted  $R^2$ ) of the professional and technical workers earnings function is greater than that of the artist earnings function, but not that different.

Consistent with results from other studies, the return to an extra year of education is found to be greater for professional and technical workers in all years except for 1949. The interpretation of this result, based on Wassall and Alper (1984), Towse (1992a) and Rengers and Madden (2000), is that since Census data mixes artistic and non-artistic earnings in reported total earnings of artists, the resulting coefficient of years of education in the earnings function reflects no or negative correlation of education with artistic earnings and a positive correlation with non-artistic earnings.<sup>29</sup> This was also observed by Montgomery and Robinson (1993) using data from a different survey. The relative difference in 1999 is among the largest it has been.<sup>30</sup>

<sup>27</sup> Earnings functions for 1940 are not presented. The failure of the Census to collect information on self-employment income in 1940 biases results for those professions with substantial income from this source; prominent among such professions are most artistic disciplines.

<sup>28</sup> Other independent variables not reported in Tables 7 and 8 include whether the artist was: a member of another ethnic group (not for 2000); in school at the time of the census; a veteran; a federal, state or local government employee. It also included regional and occupational dummy variables.

<sup>29</sup> Using the New England artist data, it was found that years of schooling was significantly and negatively correlated with artistic earnings, but significantly and positively correlated with arts-related earnings and with non-arts related earnings [Wassall and Alper (1984)]. The finding of a significant negative relationship between years of schooling and artistic earnings may overstate the adversity of this relationship. Presumably artists are maximizing earnings at the margin among all jobs; those who are better educated receive a higher return per extra hour worked in non-artistic jobs, so they reduce their artistic labor supply and thus their artistic incomes. Little or no correlation between education and artistic earnings is consistent with the notion that artistic talent is innate and cannot be enhanced by general education.

<sup>30</sup> US Census data does not provide information that allows the researcher to control for an individual's mental abilities, though there are controls for physical abilities in the data and model. This is likely lead to an upward bias in the estimated coefficients and therefore on the estimated returns to formal schooling [Angrist and Krueger (2001); Griliches and Mason (1972)]. It also is not possible to control for artistic ability, but the impact of this is unknown.

Table 7  
 Selected results from earnings functions of artists and professional and technical workers in the experienced civilian labor force: 1950–2000\* (dependent variable is natural log of earnings; *t*-statistics in parentheses)

Variables	1950		1960		1970		1980		1990		2000	
	Artist	Prof.	Artist	Prof.	Artist	Prof.	Artist	Prof.	Artist	Prof.	Artist	Prof.
Education	0.079 (3.42)	0.060 (8.43)	0.030 (2.93)	0.053 (17.83)	0.059 (10.97)	0.081 (65.53)	0.051 (13.03)	0.080 (109.30)	0.104 (26.83)	0.116 (101.02)	0.080 (46.05)	0.116 (264.00)
Experience	0.166 (2.84)	0.157 (8.97)	0.140 (5.76)	0.129 (20.61)	0.165 (17.17)	0.112 (51.64)	0.077 (9.44)	0.117 (76.75)	0.174 (15.51)	-0.049 (-16.78)	0.127 (35.84)	0.114 (133.60)
Experience <sup>2</sup>	-0.008 (-2.29)	-0.008 (-6.73)	-0.007 (-4.23)	-0.006 (-14.34)	-0.007 (-11.55)	-0.004 (-27.82)	-0.003 (-5.01)	-0.005 (-46.36)	-0.007 (-11.15)	-0.010 (-58.40)	-0.005 (-21.05)	-0.004 (-65.93)
Experience <sup>3</sup>	0.0002 (1.95)	0.0001 (5.58)	0.0001 (3.10)	0.0001 (11.01)	0.0001 (8.16)	4.5E-5 (13.80)	3.6E-5 (2.53)	8.8E-5 (32.00)	0.0001 (7.97)	-0.0003 (-89.97)	5.5E-5 (10.53)	4.2E-5 (29.98)
Experience <sup>4</sup>	-1.0E-6 (-1.77)	-1.3E-6 (-5.33)	-7.6E-7 (-2.46)	-8.4E-7 (-9.91)	-7.1E-7 (-6.69)	-2.0E-7 (-8.30)	-2.4E-7 (-2.03)	-6.7E-7 (-29.37)	-7.6E-7 (-6.64)	2.8E-6 (101.01)	-2.2E-7 (-5.39)	-2.0E-7 (-17.49)
Head of household	1.124 (6.53)	0.858 (16.96)	0.894 (12.92)	0.873 (45.95)	0.943 (26.75)	0.602 (78.35)	0.769 (34.35)	0.489 (111.99)	0.692 (36.20)	0.664 (122.74)	0.391 (48.33)	0.281 (149.59)
Female	-0.589 (-3.13)	-0.537 (-8.74)	-0.781 (-10.2)	-0.332 (-14.48)	-0.767 (-22.53)	-0.416 (-51.37)	-0.622 (-26.92)	-0.288 (-60.70)	-0.457 (-22.16)	-0.310 (-50.46)	-0.419 (-51.58)	-0.226 (-113.24)
Black	-0.745 (-2.10)	-0.267 (-2.76)	-0.483 (-3.27)	-0.168 (-5.13)	-0.168 (-2.50)	-0.058 (-5.10)	-0.408 (-8.83)	-0.166 (-25.48)	-0.303 (-7.10)	-0.013 (-1.44)	0.000 (-0.01)	0.050 (17.44)
Asian	1.085 (0.97)	-0.900 (-1.76)	0.387 (1.25)	0.111 (-1.44)	-0.015 (-0.14)	-0.004 (-0.16)	0.094 (1.38)	0.116 (9.63)	-0.026 (-0.50)	0.028 (2.02)	-0.025 (-1.24)	0.032 (7.51)
Hispanic	-0.401 (-0.58)	-0.491 (-1.93)	0.200 (0.78)	-0.130 (-1.78)	-0.239 (-3.18)	-0.105 (-5.52)	-0.063 (-1.09)	0.018 (1.57)	-0.052 (-0.95)	0.007 (0.39)	-0.007 (-0.39)	0.047 (11.95)

(continued on next page)

Table 7  
(continued)

Variables	1950		1960		1970		1980		1990		2000	
	Artist	Prof.	Artist	Prof.	Artist	Prof.	Artist	Prof.	Artist	Prof.	Artist	Prof.
Married	0.072 (0.46)	0.310 (6.80)	0.195 (3.09)	0.447 (26.48)	0.137 (4.45)	0.211 (31.55)	0.147 (6.74)	0.197 (48.08)	0.095 (5.18)	0.107 (20.35)	0.027 (3.28)	0.068 (36.02)
Children under 6	-0.089 (-0.57)	-0.074 (-1.74)	-0.379 (-5.63)	-0.160 (-9.27)	-0.102 (-3.07)	-0.171 (-24.34)	-0.259 (-7.00)	-0.181 (-30.38)	-0.384 (-8.87)	-0.922 (-93.19)	-0.088 (-8.15)	-0.043 (-19.11)
Disability	-	-	-	-	-0.393 (-8.18)	-0.349 (-19.10)	-0.845 (-19.10)	-0.640 (-70.72)	-1.071 (-27.07)	-1.674 (-161.17)	-0.094 (-7.27)	-0.050 (-16.45)
Non-citizen	-0.623 (-1.13)	-0.204 (-2.45)	-	-0.037 (-0.73)	-0.322 (-3.67)	-0.046 (-3.22)	-0.484 (-7.13)	0.019 (1.88)	-0.351 (-6.92)	0.058 (5.11)	-0.112 (-5.44)	0.158 (34.97)
Immigrant	-0.006 (-0.02)	-0.891 (-4.33)	-0.048 (-0.26)	-	0.162 (2.91)	-0.291 (-12.93)	0.157 (3.10)	-0.360 (-26.90)	0.098 (2.39)	-0.617 (-41.02)	0.127 (7.49)	0.053 (14.32)
English spoken	-	-	0.033 (0.16)	0.191 (3.19)	-	-	0.055 (1.44)	0.146 (19.65)	0.088 (2.76)	0.087 (9.63)	0.034 (2.49)	0.037 (12.05)
Self-employed	-0.782 (-4.88)	0.005 (0.07)	-0.467 (-7.75)	0.122 (4.48)	-0.448 (-14.56)	-0.177 (15.54)	-0.933 (-42.00)	-0.487 (61.49)	-0.717 (-38.65)	-0.101 (11.08)	-	-
Self-incorporated	-	-	-	-	-	-	-	-	-	-	0.220 (17.98)	0.124 (22.33)
Self-unincorporated	-	-	-	-	-	-	-	-	-	-	-0.491 (-53.91)	-0.402 (-96.83)
$R^2$	0.148	0.197	0.253	0.242	0.265	0.225	0.163	0.171	0.139	0.196	0.294	0.366
$F$	8.87	91.68	60.98	552.61	212.76	2375.62	297.47	4231.35	363.54	7807.47	993.73	18,083.03
$n$	1447	14,010	5681	67,502	19,938	319,695	53,479	841,699	78,610	1,314,705	97,594	1,284,105

\*Year of labor market data is the calendar year prior to the census year.

Table 8  
Variables in the estimated earnings functions

Variable	Values	Explanation
Education	Years of schooling	Assigned from Census variable measuring highest level of education attained. See footnote 12.
Experience, Experience <sup>2</sup> , Experience <sup>3</sup> and Experience <sup>4</sup>	Years of experience and years of experience up to the 4th power	Calculated as: age – education – 5
Head of household	1–household head, 0–not household head	
Female	1–female, 0–male	
Black	1–race black, 0–not black	
Asian	1–race Asian, 0–not Asian	
White and other	Racial category excluded	
Hispanic	1–ethnicity is Hispanic, 0–ethnicity not Hispanic	
Married	1–married, 0–not married	Not married includes: separated, divorced, widowed and never married
Children under 6	1–children under six years old present in household, 0–otherwise	
Disability	1–employment related disability, 0–otherwise	
Immigrant	1–born outside the US, 0–born in the US	
Non-citizen	1–citizen of the US, 0–not a citizen of the US	
English spoken	1–speaks English (well or very well), 0–otherwise	
Self-employed	1–self-employed, 0–employee	
Self-incorporated	1–self-employed in an incorporated business, 0–otherwise	
Self-unincorporated	1–self-employed in an unincorporated business, 0–otherwise	

Until 1999 both estimated earnings functions for the artists and the reference group of professionals showed the usual adverse effects on earnings of being a woman or a member of most minority groups. In 1999 the difference in earnings associated with the artists' race and ethnicity seems to have disappeared, not so for their professional peers. The difference between male and female artists' earnings is still significant, as it

Table 9  
Statistical rewards and penalties of being an artist

Reward/penalty	Year*					
	1950	1960	1970	1980	1990	2000
Percent return to extra year of education: artists	7.9	3.0	6.0	5.1	10.4	8.0
Percent return to extra year of education: professional & technical workers	6.0	5.3	8.1	8.0	11.6	11.6
Percent earnings difference between artists and other professional & technical occupations: actual	58.2	56.3	43.6	90.3	12.1	21.9
Percent earnings penalty to artists for not working in other professional & technical occupations: calculated using regression coefficients	19.3	25.3	33.2	50.7	5.9	8.4

*Notes:* Returns to education and earnings penalties calculated from regression equations with selected coefficients shown in Tables 6 and 7. In calculating penalty to artists for not entering other professional and technical occupations, it was assumed that they would distribute among these occupation in the same percentages as the then existing occupational distributions.

\*Year of data is the calendar year prior to the census year.

is for the other professionals. This difference has shown little change since 1989 and is still considerable. In 1989 black professionals' earnings were not significantly different from their white peers' earnings, and in 1999 they were actually earning significantly more than white professionals.

Being a head of household affects artists' earnings with a larger positive affect than it does for professional and technical workers in general. Being married has a larger positive affect on professional and technical workers' earnings than on the artists' earnings for the entire period.

Using these earnings functions, the return to an extra year of education can be observed (at mean values) for both groups. This is shown in Table 9. As noted, in all years but 1949, the return to an extra year of education is greater for professional and technical workers.<sup>31</sup>

<sup>31</sup> Though not reported because of possible bias due to missing self-employment income, the calculated return to an extra year of education for artists in 1939 was also (2.3 percent versus 4.2 percent) less than that for other professional and technical workers.



The earnings functions can also be used to estimate a more sophisticated version of the earnings penalty associated with being an artist, referred to earlier. This is accomplished by comparing the predicted earnings of an artist possessing average levels of all characteristics used as explanatory variables in the earnings function to what that artist would earn, with the same average levels of all characteristics, as a professional or technical worker (i.e., using the estimated coefficients from the professional and technical worker earnings function).

Using this technique, the earnings penalty to artists is estimated to be less than that revealed by the direct comparisons of earnings in the previous section. In 1959, an artist with average characteristics is estimated to have earned 25 percent more as a professional or technical worker. In 1969, an artist with average characteristics would have earned 33 percent more as a professional or technical worker. In 1979, an artist with average characteristics would have earned 51 percent more as a professional or technical worker. In 1989 the calculated difference was only six percent. Most recently, in 1999, the calculated difference in the predicted earnings of the average artist in a professional or technical job is eight percent more than they would have earned as a professional or technical worker.

In comparison, Filer (1986, p. 72) calculated, using the same method, the earnings penalty to artists relative to members of the general work force. He finds that an artist with average characteristics would have earned 10.3 percent more as a member of the general work force. He was subsequently criticized by several authors for comparing artists to the general work force. Clearly one would expect a larger earnings penalty to be calculated if artists were simulated as members of occupations where more education is required and returns to additional education are greater. This is consistent with these findings.

Filer (1986) dismisses the notion that artists have a lifetime earnings penalty because the penalty, calculated at the mean, is offset over one's work years by the steeper age-earnings profile of artists than those for the general work force.<sup>32</sup> In this paper a larger earnings penalty is found to exist in each census year when artists are compared to professional and technical workers. Further, since artists' returns from extra years of experience are comparable to that for the overall professional and technical work force, this argument becomes moot when a more appropriate reference group is used.

## 5. US artists' careers: 1979–1998

In the United States several nationally representative panel surveys have been in existence for some time. Perhaps the best known is the US Department of Labor's National Longitudinal Survey of Youth. Initiated in 1979, it was designed to follow a sample of young people from high school throughout their working careers. Alper and Wassall

<sup>32</sup> Filer (1988) does not calculate an earnings penalty in his paper.

(2002) utilized these surveys to examine issues of longevity and career change among artists in the US.

The survey participants were 14 to 22 years of age when first interviewed in 1979. Two additional panels were added over the years to make the sample more representative of the population. One panel was a military sample that was later dropped. The other was a group of economically disadvantaged youth, some of whom were later dropped as well. In total 12,686 individuals were part of the survey at some point during the more than 20 years it has been in existence. The participants were interviewed annually from 1979 through 1994 and every other year thereafter. The number of artists in the survey is not large enough to provide detailed information by type of artist. Therefore, the artists are grouped into four artistic occupations: performers, architects and designers, visual artists and 'other' artists.

What follows presents a sampling of what can be learned about artists' careers through the use of panel data. It starts by examining the stock and flow of people into and through arts occupations (Section 5.1). Then there is a brief discussion of who they are at the start of their artistic careers (Section 5.2). The next section examines the transitions that occur during artistic careers with respect to the artistic and non-artistic jobs they held (Section 5.3). The last two sections describe who permanently leaves the arts (Section 5.4) and what they do afterwards (Section 5.5).

### *5.1. Flow and tenure*

Clearly evident from this panel survey is the fact that many people explore the arts as an occupation but very few remain as artists for significant periods of time. When comparing the number of people who indicated that they were artists in any given year, i.e., the stock of artists, to the total number of people who moved into and out of artistic occupations over the survey's 19 years, i.e., the flow of artists, a significant difference in these numbers is found. Throughout the period 766 of the 12,686 people indicated that they had worked as artists at some point in time. This is more than five times the number of people who were artists in any given year (Table 10).

Of those who were part of the artist population at some point over this period, almost fifty-seven percent were male. This is slightly greater than the overall male representation in the survey of just under fifty-one percent, suggesting that over their careers men are more likely to explore the arts as a means of earning income than women.

The average time spent working as an artist was only 2.2 of the survey years. In comparison, people who had been artists at some time during the survey period also worked at non-arts jobs in approximately 10.2 of the 16 survey years. The number of survey years working as an artist was slightly, but not significantly, higher for men than women (2.3 years versus 2.1 years). Almost 60 percent of those who worked as artists worked as artists in only one of the survey years. In comparison, more than 98 percent of those who were artists at some point during the period also spent time in more than one year employed in non-arts jobs. Just slightly more than two percent of those who were artists at some point were artists in 10 or more years while more than 60 percent

Table 10  
Number of artists by year: NLSY79

	Architect or designer	Performer	Visual artist	Other artist	Total
1979	2	15	10	8	35
1980	3	25	10	14	52
1981	8	23	16	17	64
1982	11	35	13	21	80
1983	13	38	20	23	94
1984	21	41	22	29	113
1985	25	32	19	25	101
1986	33	53	21	24	131
1987	36	39	24	43	142
1988	33	40	29	26	128
1989	20	39	45	5	109
1990	32	43	29	28	132
1991	24	38	28	25	115
1992	32	45	28	26	131
1993	22	42	33	30	127
1998	56	37	29	22	144

Source: Authors' tabulations and calculations from NLSY79, 1979–1998, Release 10.0.

worked in 10 or more years at non-arts jobs. The female artists' behavior was not that different from the male artists. Only 1.5 percent of the female artists worked as artists in 10 or more survey years, while 2.5 percent of the male artists did the same.

## 5.2. The start

Artists started working as artists, on average, at just about 25 years of age (Table 11). The difference between starting ages for men and women (approximately a half year) was not statistically significant.<sup>33</sup> Only for the visual artists was the difference in starting age significant (marginally), with women starting to work as painters and sculptors two years later than men.

Overall there is a significant difference in the ages at which the artists started working in their arts occupations when comparing across their first arts occupations. The average age for the architects and designers was significantly greater than any other artist group. This probably reflects the additional years of formal schooling required to become an architect.

<sup>33</sup> It is likely that the true average age at which these artists started working as artists is somewhat lower. Some of the artists were 14 years of age at the first interview and were not likely to have worked at all prior to their participation in the survey, while others were as old as 22 at the first interview. Those who were older may well have already been working as artists by the time they were first interviewed.

Table 11  
Average age started being an artist by gender and first arts occupation (years)

	Total	Male	Female
Architect or Designer	27.2	27.4	26.8
Performer	24.2	24.1	24.3
Visual artist	25.4	24.4	26.4
Other	24.9	24.6	25.1

Source: Authors' tabulations and calculations from NLSY79, 1979–1998, Release 10.0.

### 5.3. The transition

An examination of those employed in the arts at some time during any given year reveals that the proportion that had an arts occupation as their primary occupation tended to diminish over time. This was determined by looking at the employed artists' current occupations at the time of the interview.<sup>34</sup> While the decline in the proportion with their primary occupations being arts occupations is not constant as artists mature, the trend is clearly in the downward direction. It was at the time of the fourth round of interviews (1982), when the artists were 18 to 26 years old, that the largest proportion of people employed as artists reported having an arts occupation as their primary occupation. Almost 85 percent were working as artists (Table 12). By 1998, the proportion had decreased to approximately 70 percent.

A transition that occurred over the period is the change in variety of artists' non-arts primary work activities.<sup>35</sup> In 1979, the year in which the artists were the youngest, approximately 17 percent of those artists whose primary occupations were not in the arts were working in some other professional or managerial occupation and about one-half were working in sales and clerical occupations. By 1998 more than two-thirds of the artists with primary occupations not in the arts were working in other professional and managerial jobs while fewer than ten percent were working in sales and clerical jobs. Additionally, 17 percent were working in service occupations at the beginning of the survey period, all in food service, but by 1998 only five percent were doing so, none in food service.

There were transitions among the arts activities the artists participated in as their current or primary occupation. They may reflect factors including changes in the artists'

<sup>34</sup> When the artist indicated that s/he was working at more than one occupation, then the one in which s/he worked the most hours was determined to be his/her primary occupation.

<sup>35</sup> This was determined by looking at the primary work activities at the time of the interview for those who indicated that they had worked as artists within the last 12 months but did not indicate that they were working as artists at the time of the interview.

Table 12  
Occupation in the years working as an artist (percent)

	Prof. & mang. <sup>1</sup>	Sales & clerical	Craft, operative, laborer or farmer	Service	Total non-art	Arch. or design <sup>2</sup>	Perf. <sup>3</sup>	Visual artist	Other	Total art
1979	4.2	12.5	4.2	4.2	25.1	8.3	25.0	25.0	16.7	75.0
1980	2.6	5.2	5.2	7.9	20.9	16.3	47.4	7.9	21.1	79.0
1981	14.3	6.1	2.0	2.0	24.4	16.3	30.6	16.3	12.2	75.4
1982	6.7	3.3	3.4	3.4	16.8	11.7	38.4	13.3	20.0	83.4
1983	4.5	10.6	4.5	3.0	22.6	10.6	33.2	18.2	15.2	77.2
1984	10.3	4.6	4.5	8.0	27.4	17.2	20.6	18.4	16.1	72.3
1985	3.6	6.0	4.8	3.6	18.0	21.7	22.9	16.8	20.5	81.9
1986	13.3	3.6	12.4	8.8	38.1	19.4	20.3	11.5	10.6	61.8
1987	15.5	4.3	7.8	5.2	32.8	20.7	17.2	12.9	16.4	67.2
1988	8.3	11.9	10.1	6.4	36.7	23.9	12.8	16.5	10.1	63.3
1989	8.0	13.7	10.3	5.7	37.7	18.4	18.4	18.3	6.9	62.0
1990	14.1	6.2	5.4	2.7	28.4	23.0	21.3	15.0	12.4	71.7
1991	16.2	6.5	6.5	3.3	32.5	14.0	21.6	17.3	15.1	68.0
1992	11.2	10.3	0.0	4.7	26.2	23.4	16.8	18.7	14.9	73.8
1993	16.0	7.5	3.8	1.8	29.1	15.1	20.6	17.0	17.9	70.6
1998	20.8	2.5	5.8	1.7	30.8	34.1	9.1	15.0	10.9	69.1

Source: Authors' tabulations and calculations from NLSY79, 1979–1998, Release 10.0.

<sup>1</sup>Professional and managerial.

<sup>2</sup>Architect and designer.

<sup>3</sup>Performer.

labor markets, changes in the markets for their art, the end result of a search process for the “right” arts occupation or the completion of the requisite education and training. For example, in 1979 approximately 22 percent of the employed artists had occupations among the “not elsewhere classified” group. By 1998 the proportion had dropped to fewer than 15 percent. The performer occupations were the most popular for most of the early part of the period. They were replaced by people transitioning into the architect and designer occupations during the latter part of the period probably reflecting the time required to become trained as an architect.

An examination of the non-arts occupations held by the artists in the years they did not work as artists provides some insights into what the artists were doing to survive and what they do as they stop exploring the possibility of an artistic career. In part reflective of the ages and educational backgrounds of those surveyed in 1979, almost 30 percent of those people who did not work as artists were working at sales or clerical jobs with another 30 percent working in service jobs (Table 13). The proportion working at sales or clerical jobs peaked in 1982, when they were between 18 and 26 years old, and declined throughout the remainder of the period. There was an even greater decline

Table 13  
Occupation in the years not working as an artist (percent)

	Professional & managerial	Sales & clerical	Craft, operative, laborer or farmer	Service
1979	12.5	27.6	29.5	30.4
1980	12.6	33.5	30.7	23.2
1981	18.2	30.7	27.9	23.3
1982	22.3	33.6	23.4	20.5
1983	27.8	31.8	20.9	19.4
1984	30.3	30.1	24.2	15.4
1985	33.9	26.9	25.3	14.1
1986	32.0	27.1	28.9	12.0
1987	36.4	27.2	22.6	13.7
1988	40.3	24.2	24.7	10.8
1989	44.8	20.7	22.5	12.0
1990	43.4	20.2	23.0	13.4
1991	49.8	19.5	23.6	7.2
1992	45.3	24.0	22.8	9.9
1993	46.0	19.6	22.6	11.6
1998	57.2	18.3	17.7	6.6

Source: Authors' tabulations and calculations from NLSY79, 1979–1998, Release 10.0.

in the proportion working in the service occupations. By 1998 only seven percent held service jobs. The greatest growth in employment was found in the other professional and the managerial occupations with the proportion increasing fourfold over the period. This change is reflective not only of the artists aging, but in the investments they made in education and training.

#### 5.4. *The leavers*

Push and pull factors are both important in the decision to leave the arts as it is for any occupation. By 1998 only 42 percent of those who indicated that they had worked as artists at some time during the period were still working, to some degree, as artists.<sup>36</sup> This means that either the person's current job was among the arts occupations or one of the five other jobs they could have possibly held during the time since the previous

<sup>36</sup> Those who had been artists but who were no longer working as artists were anyone who did not report that s/he worked as an artist for at least three consecutive years. This definition was used to examine the former artists' post-artist careers and to assist in developing a better understanding of the factors that may have enticed them to leave the arts.

Table 14  
Artists' current status by first arts occupation (percent)

	Still artist	No longer artist	Status unknown
Architect or designer	53.1	39.6	7.3
Performer	39.8	50.3	9.9
Visual artist	38.7	51.4	9.9
Other	36.3	53.9	9.8
All	41.6	48.1	9.3

Source: Authors' tabulations and calculations from NLSY79, 1979–1998, Release 10.0.

Table 15  
First arts occupation by last arts occupation (percent)

First arts occupation	Last arts occupation			
	Arch./des.	Performer	Visual	Other
Architect or designer	100.0	0.0	0.0	0.0
Performer	1.8	89.0	0.09	8.3
Visual artist	0.0	0.0	95.3	4.7
Other	2.7	2.7	3.6	90.9
All	19.9	26.6	23.4	30.0

Source: Authors' tabulations and calculations from NLSY79, 1979–1998, Release 10.0.

interview was in the arts. Almost half were no longer working at all as artists. The status of the remaining 10 percent was not known.<sup>37</sup>

Based on their first-arts occupation, the architects and designers were the most likely to still be working as artists in 1998. Approximately 53 percent were still working in the arts (Table 14). Fewer than forty percent of those in the performing arts occupations (actor, dancer, musician, composer and announcer) were still artists. The proportion is basically the same for visual artists. The artists who started out in the 'other' arts occupations were somewhat less likely to still be artists than performers and visual artists.

For the most part artists who stopped working as artists did so from the same arts occupation that they had started in. By comparing the first-arts occupation to the last-arts occupation for those who were no longer working as artists, it was found that the vast majority were participating in the same occupation at both points in time (Table 15).

<sup>37</sup> After the 1990 survey a portion of the people who had been in the supplemental samples was dropped. Included among them were the economically disadvantaged non-black, non-Hispanic males and females who were part of the supplemental sample. They also included those who were lost due to natural attrition.

Architects and designers were the most likely to have persisted in the same occupation. Visual artists were second while the performers and 'other' artists were the least likely to persist in their starting occupations indicating that these artists have experimented the most in the arts.

### 5.5. *The post-artist period*

On average, the age at which artists stopped working in the arts was 24 (Table 16). This estimate, like the age at which artists started working, is not likely to be representative of all artists. In fact it is likely to be an underestimate of the actual age since the oldest members of the sample were only 41 years old at the time of the last survey, and the youngest were 33. Without observing the artists over their entire lifetimes, the estimate is smaller than the true value. The variation in the age at which artists stopped working as artists related to gender was not significant.

For the most part, those people who had been working as artists but who stopped working in the arts continued to work (Table 17). Almost three-quarters were still employed. Another ten percent were still part of the civilian labor force but were un-

Table 16  
Average age stopped being an artist by gender and last arts occupation (years)

	Male	Female	Total
Architect or designer	24.7	25.3	24.9
Performer	23.9	23.3	23.7
Visual artist	23.5	23.5	23.5
Other	23.4	23.2	23.3
All	23.8	23.7	23.8

*Source:* Authors' tabulations and calculations from NLSY79, 1979–1998, Release 10.0.

Table 17  
Post-arts employment status (percent)

	Male	Female	Total
Employed	74.6	70.8	73.1
Unemployed	8.0	12.6	9.9
Keeping house	0.9	7.3	3.6
In school	7.5	5.3	6.6
Other	5.6	4.0	4.9
Armed forces	3.3	0.0	1.9

*Source:* Authors' tabulations and calculations from NLSY79, 1979–1998, Release 10.0.



employed in the year immediately following the year in which they last worked as an artist. The remainder dropped out of the civilian labor force with almost seven percent attending school, nearly four percent “keeping house” and almost two percent joining the armed forces. The remaining five percent were doing something else or were unable to work for some unspecified reason. Some differences were found in what the male and female former artists were doing after leaving the arts.

Post-artist employment status did vary with the last-arts occupation. Architects and designers were the most likely to still be employed in the year after they stopped working as artists. The artists who comprised the ‘other’ category were the least likely to continue working in the year after they stopped working as artists. Those former artists in the ‘other’ group of arts occupations were the most likely to be unemployed. Unemployment rates for the former performers and visual artists were essentially the same. The visual former artists were the most likely to be “keeping house” in the year post-art work. Those who were most likely to stop working as artists to attend school were the artists in the ‘other’ arts group.

There were also differences in the post-arts jobs associated with the former artists' gender (Table 18). There was little difference between the proportion male and female former artists whose post-arts employment was in a non-arts professional or managerial occupation. A major difference in post-arts employment was the proportion of female artists working in sales or clerical occupations. The female former artists were almost twice as likely as their male peers to be working in these occupations. Females were also 50 percent more likely to be working in service occupations than males. Males were considerably more likely to have post-arts jobs in the craft, operative and laborer occupations than their female colleagues.

Reflective of many factors, including their education, training, and prior work experience, the former artists' post-arts occupations varied by their last-arts occupation. Those who had been architects and designers were by far the most likely to be working in non-arts professional and managerial occupations. Those who had been visual artists were the least likely to be working in professional and managerial occupations. The “other” artists, who include college art-teachers and authors, and performers were between the other two groups in the proportion working in professional and manager-

Table 18  
Post-arts occupation (percent)

Occupation	Male	Female	Total
Professional + managerial	38.3	34.5	36.9
Sales + clerical	19.5	35.4	25.9
Craft, operative, laborer + farmer	30.1	11.2	22.5
Service	12.0	18.6	14.7

Source: Authors' tabulations and calculations from NLSY79, 1979–1998, Release 10.0.

ial occupations. Overall, almost 25 percent of the former artists were working in either sales or clerical occupations with the former visual artists most likely to be doing so. Overall, almost 15 percent were working in service occupations with the visual former artists most likely to be doing so. Relatively few former artists worked in the crafts, operatives (including transportation) and laborer occupations.

## 6. Summary, conclusions and future research

The survey and research presented here is a continuation of the ongoing work of economists, and others, on artists' labor markets and their careers. It highlights the use of a quasi-panel obtained from census data to examine the employment and earnings of artists and compares them to those of all other professional and technical workers. It also provides a glimpse of what can be learned about artists' careers from a study of artists based on true panel data that follows the same individuals over a period of time.

The quasi-panel of data from the seven US censuses, along with research from other countries such as Throsby's work in Australia, provides a reasonably consistent set of findings in each census year. Artists work fewer hours, suffer higher unemployment and earn less than members of the reference group. Over the sixty year period, disparities in unemployment and annual hours worked shrink somewhat, but disparities in earnings do not. Artists earn less across all years even when only members working full-time year-round of each group are compared. The earnings of artists display greater variability than those of other professional and technical workers. The greater earnings inequality of artists is reduced when only full-time year-round workers from each group are compared.

Evidence on earnings and earnings distributions give credence to theories of artists as risk-takers and as participants in winner-take-all markets. Earnings inequality measures for artists (as well as for the general work force) increased over the 1949–1999 period. However, the consistently lower mean and median earnings of artists compared to their reference group over the 1939–1999 period is more in line with a risk-taking theory of artist behavior than one of superstardom, which does not predict that mean earnings should suffer.

Earnings functions for artists and for professional and technical workers are estimated for six of the seven census years. It is shown that the return to years of schooling is lower for artists in all but one year. This is consistent with findings by several authors that education does not significantly increase artistic earnings but does increase their non-artistic earnings.

Earnings penalties that artists face because of their career choices are found to be quite large, varying from 6 to 51 percent of annual artistic earnings across the six census years. In all but one year, the estimated earnings penalty is as great as or greater than the actual observed earnings difference.

The panel data that cover almost twenty years in the artists' lives provides a limited set of insights into their behavior. It does suggest that many people participate in the

artistic labor market, but that few succeed to the point that enables them to develop a career in the arts. In part due to their relatively high educational levels, artists are able to transition from their forays into arts occupations to jobs in other professional and managerial occupations, not into service occupations as the 'mythology' of the arts might suggest. It is true that when they are young and struggling to make it as artists they do work in various service occupations that tend to provide greater work schedule flexibility.

There are a number of questions which are not addressed in this paper. Does the greater amount of multiple job-holding by artists, both inside and outside their artistic occupation, explain some or all the observed differences in annual hours worked and unemployment rates? To what extent does holding jobs outside the artistic profession reduce risk and thus earnings inequality as well as supplement earnings? What triggers the artists' decision to leave the arts for good? How does the allocation of an artist's time to various income generating activities change over his/her career?

These and other questions raised in this paper can best be answered using more detailed data on the careers of artists. There is clearly a need for additional survey-based panel data on artists. The existing national panel surveys in the US, like the NLSY79, are too small. The sample of artists from these data sources are not large enough to provide reliable empirical models of their behavior without aggregating to a level where disparate types of artists are grouped together. Additionally, the information collected, while quite extensive in most cases, is not tailored to unearth information about the unique labor market activities of many artists, especially their multiple jobholding behavior, sources of earnings, allocation of time and costs of producing their art. While a great deal has been learned, there is a great deal more to learn.

## Appendix A

Table A.1  
Census public use sample artistic occupations and unweighted sample sizes

Occupations (with original titles)	Census year						
	1940	1950	1960	1970	1980	1990	2000
Actors <sup>a</sup>	209	153	179	898	3946	5694	2,568
Architects	201	252	313	2448	5842	7953	10,063
Artists & art teachers <sup>b</sup>	627	781	1289	5095	8905	11,995	14,352
Authors	156	132	322	1253	2720	5920	9327
Dancers & dance teachers <sup>c</sup>	492	144	331	460	875	1235	1566
Designers	246	919	1528	5204	19,945	34,205	42,235
Musicians & music teachers <sup>d</sup>	1480	1619	2293	4913	8251	8653	9770
Photographers	376	570	598	3095	5595	8022	7164
Showmen	76	–	–	–	–	–	–
Entertainers N.E.C. <sup>e</sup>	–	180	157	3161	3098	5717	2289
College art, drama & music teachers	–	–	–	1382	1630	1172	–
Radio and TV announcers <sup>f</sup>	–	–	–	960	2506	3157	3023
Producers and directors	–	–	–	–	–	–	7112
Total	3863	4750	7010	32,913	63,313	93,723	109,469

<sup>a</sup>Called Actors and Directors in 1980 and 1990. Excludes producers and directors in 2000.

<sup>b</sup>Called Painters and Sculptors in 1970; Painters, Sculptors, Craft-Artists, and Artist Printmakers in 1980 and 1990; and Artists and Related Workers in 2000, excludes art teachers.

<sup>c</sup>Called Dancers from 1970 to 1990; and Dancers and Choreographers in 2000, excludes dance teachers.

<sup>d</sup>Called Musicians and Composers from 1970 to 1990; and Musicians, Singers and Related Workers in 2000, excludes music teachers.

<sup>e</sup>Called Writers, Artists and Entertainers, N.E.C. in 1970; Artists, Performers and Related Workers, N.E.C. in 1980 and 1990; Entertainers and Performers, Sports and Related Workers, All Other in 2000.

<sup>f</sup>Called Announcers starting in 1980.

## References

- Adler, M. (1985). "Stardom and talent". *American Economic Review* 75, 208–212.
- Alper, N., Wassall, G. (1998). "Artists' labor market experiences: A preliminary analysis using longitudinal data". In: Heikkinen, M., Koskinen, T. (Eds.), *Economics of Artists and Arts Policy*. Arts Council of Finland, Helsinki.
- Alper, N., Wassall, G. (2000). *More Than Once in a Blue Moon: Multiple Jobholdings by American Artists*. National Endowment for the Arts, Washington, DC.
- Alper, N., Wassall, G. (2002). "Artists' careers-preliminary analysis from longitudinal data for the US". Rotterdam: 12th International Conference on Cultural Economics.
- Angrist, J., Krueger, A. (2001). "Instrumental variables and the search for identification: From supply and demand to natural experiments". *Journal of Economic Perspectives* 15 (4), 69–86.

- Baumol, W., Bowen, W. (1966). *Performing Arts: The Economic Dilemma*. Twentieth Century Fund, New York.
- Benhamou, F. (2000). "The opposition between two models of labour market adjustment: The case of audiovisual and performing arts activities in France and Great Britain over a ten year period". *Journal of Cultural Economics* 24, 301–319.
- Bielby, W., Bielby, D. (1987). *The 1987 Hollywood Writers' Report: A Survey of Ethnic, Gender and Age Employment Practices*. Writers Guild of America, west, Los Angeles.
- Bielby, W., Bielby, D. (1989). *The 1989 Hollywood Writers' Report: Unequal Access, Unequal Pay*. Writers Guild of America, west, West Hollywood.
- Bielby, D., Bielby, W. (1993a). "The Hollywood 'graylist'? Audience, demographics and age stratification among television writers". In: Cantor, M., Zollars, C. (Eds.), *Current Research on Occupations and Professions (Creators of Culture)*, vol. 8. Jai Press, Greenwich.
- Bielby, W., Bielby, D. (1993b). *The 1993 Hollywood Writers' Report: A Survey of the Employment of Writers in the Film, Broadcast, and Cable Industries for the Period 1987–1991*. Writers Guild of America, west, West Hollywood.
- Bielby, W., Bielby, D. (1998). *The 1998 Hollywood Writers' Report: Telling All Our Stories*. Writers Guild of America, west, Los Angeles.
- Citro, C., Gaquin, D. (1987). *Artists in the Workforce, 1950 to 1965*. National Endowment for the Arts, Research Division, Washington, DC.
- Ellis, D., Beresford, J. (1994). *Trends in Artist Occupations, 1970–1990*. National Endowment for the Arts, Research Division, Washington, DC.
- Federal Committee on Statistical Methodology (1985). *Statistical Policy Working Paper 13 – Federal Longitudinal Surveys*. Washington, DC.
- Felton, M.V. (1978). "The economics of the creative arts: The case of the composer". *Journal of Cultural Economics* 2, 41–61.
- Filer, R.K. (1986). "The 'starving artist' – Myth or reality? Earnings of artists in the United States". *Journal of Political Economy* 94, 56–75.
- Filer, R.K. (1988). *Labor Market Earnings of American Artists in 1980*. National Endowment for the Arts, Washington, DC.
- Filer, R.K. (1989). "The economic condition of artists in America". In: Shaw, D.V., et al. (Eds.), *Cultural Economics 88: An American Perspective*. Association for Cultural Economics, Akron.
- Filer, R.K. (1990). "Arts and academe: The effect of education on earnings of artists". *Journal of Cultural Economics* 14, 15–38.
- Frank, R.H., Cook, P.J. (1995). *The Winner-Take-All Society*. The Free Press, New York.
- Galenson, D. (2000a). "Age and the quality of work: The case of modern American painters". *Journal of Political Economy* 108, 761–777.
- Galenson, D. (2000b). "The careers of modern artists: Evidence from auctions of contemporary art". *Journal of Cultural Economics* 24, 87–112.
- Galenson, D. (2001). *Painting Outside the Lines*. Harvard University Press, Cambridge, MA.
- Galenson, D. (2002). "Quantifying artistic success: Ranking French painters-and paintings-from impressionism to cubism". *Historical Methods* 35 (1), 5–19.
- Galenson, D., Weinberg, B. (2001). "Creating modern art: The changing careers of painters in France from impressionism to cubism". *American Economic Review* 91, 1063–1071.
- Griliches, Z., Mason, W. (1972). "Education, income and ability". *Journal of Political Economy* 80 (3), S74–S103.
- Jeffri, J. (1988). *Information on Artists*. Columbia University, Research Center for Arts and Culture, New York.
- Karttunen, S. (1993). "How to identify an artist". *Conference on Social Theory, Politics and the Arts*.
- Karttunen, S. (1998a). "How to identify artists? Defining the population for 'Status-of-the-Artist' studies". *Poetics* 26, 1–19.
- Kingston, P., Cole, J.R. (1986). *The Wages of Writing: Per Word, Per Piece, or Perhaps*. Columbia University Press, New York.

- Kingston, P., Cole, J.R., Merton, R.K. (1981). "The Columbia economic survey of American authors: A summary of the findings". Columbia University Center for Social Science, New York.
- Kominski, R., Siegel, P. (1993). "Measuring education in the Current Population Survey". *Monthly Labor Review*, September.
- MacDonald, G. (1988). "The economics of rising stars". *American Economic Review* 78, 155–166.
- McNertney, E., Waits, C.R. (1989). "The incomes of cultural providers: A review of current research". In: Shaw, D.V., et al. (Eds.), *Cultural Economics 88: An American Perspective*. Association for Cultural Economics, Akron.
- Menger, P., Gurgand, M. (1996). "Work and compensated unemployment in the performing arts. Exogenous and endogenous uncertainty in artistic labour markets". In: Ginsburgh, V., Menger, P. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier/North-Holland, Amsterdam, pp. 347–381.
- Montgomery, S., Robinson, M. (1993). "Visual artists in New York: What's special about person and place?" *Journal of Cultural Economics* 17, 17–39.
- Montgomery, S., Robinson, M. (2003). "What becomes of undergraduate dance majors?" *Journal of Cultural Economics* 27, 57–71.
- Rengers, M. (1998a). "The careers of visual artists: A longitudinal analysis". 10th International Conference on Cultural Economics.
- Rengers, M. (2002). *Economic Lives of Artists: Studies into Careers and the Labour Market in the Cultural Sector*. Utrecht University, Interuniversity Center for Social Science Theory and Methodology, Utrecht.
- Rengers, M., Madden, C. (2000). "Living art: Artist between making art and making a living". *Australian Bulletin of Labour* 26, 325–354.
- Richardson, J. (1980). "Career patterns of American opera singers". In: Hendon, W., et al. (Eds.), *Economic Policy for the Arts*. Abt Books, Cambridge, MA, pp. 176–183.
- Rosen, S. (1981). "The economics of superstars". *American Economic Review* 71, 845–858.
- Santos, F.P. (1976). "Risk, uncertainty and the performing artist". In: Blaug, M. (Ed.), *The Economics of the Arts*. Westview Press, Boulder, CO.
- Smith, T. (2000). "Longitudinal analysis of the earnings of artists". 11th International Conference of Cultural Economics.
- Snooks, G. (1983). "Determinants of earnings inequality amongst Australian artists". *Australian Economic Papers* 22, 322–332.
- Stohs, J. (1989). "Factors affecting the career patterns of male fine and applied artists". *Journal of Social Behavior and Personality* 4, 327–346.
- Stohs, J. (1990). "Young adult predictors and midlife outcomes of male fine art careers". *The Career Development Quarterly* 38, 213–229.
- Stohs, J. (1991a). "Young adult predictors and midlife outcomes of 'Starving Artists' careers: A longitudinal study of male fine artists". *Journal of Creative Behavior* 25, 92–105.
- Stohs, J. (1991b). "Moving beyond women's career choices: Factors associated with career continuity among female former art students". *Journal of Career Development* 18, 123–138.
- Stohs, J. (1992a). "Intrinsic motivation and sustained art activity among male fine and applied artists". *Creativity Research Journal* 5, 245–252.
- Stohs, J. (1992b). "Career patterns and family status of women and men artists". *Career Development Quarterly* 40, 223–233.
- Throsby, D. (1986). *Occupational and Employment Characteristics of Artists*. Australia Council, Sydney.
- Throsby, D. (1992). "Artists as workers". In: Towse, R., Kahke, A. (Eds.), *Cultural Economics*. Springer-Verlag, Berlin/Heidelberg.
- Throsby, D. (1994a). "A work-preference model of artist behaviour". In: Peacock, A., Rizzo, I. (Eds.), *Cultural Economics and Cultural Policies*. Kluwer Academic, Boston, MA, pp. 69–80.
- Throsby, D. (1996). "Disaggregated earnings functions for artists". In: Ginsburgh, V., Menger, P.-M. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier/North-Holland, Amsterdam, pp. 331–346.
- Throsby, D., Hollister, V. (2003). *Don't Give Up Your Day Job: An Economic Study of Professional Artists in Australia*. Australia Council, Sydney.

- Throsby, D., Mills, D. (1989). *When Are You Going to Get a Real Job?* Australia Council, Sydney.
- Throsby, D., Thompson, B. (1994). *But What Do You Do For a Living?* Australia Council, Sydney.
- Throsby, D., Thompson, B. (1995). *The Artist At Work*. Australia Council, Sydney.
- Towse, R. (1992a). "The earnings of singers: An economic analysis". In: Towse, R., Kahke, A. (Eds.), *Cultural Economics*. Springer-Verlag, Berlin/Heidelberg.
- US Bureau of the Census (1975). *Historical Statistics of the United States, Colonial Times to 1970, Bicentennial Edition*. Washington, DC.
- US Census Bureau (2003). *Public Use Microdata Sample-2000 Census of Population and Housing: Technical Documentation*. Washington, DC.
- Waits, C.R., McNertney, E.M. (1980). "Uncertainty and investment in human capital in the arts". In: Hendon, W.S., et al. (Eds.), *Economic Policy for the Arts*. Abt Associates, Cambridge, MA.
- Wakin, D. (2004). "The Juilliard effect: Ten years later". *The New York Times*, December 12.
- Wassall, G.H., Alper, N.O. (1984). "Determinants of artists' earnings". In: Hendon, W.S., et al. (Eds.), *The Economics of Cultural Industries*. Association for Cultural Economics, Akron, OH.
- Wassall, G.H., Alper, N.O. (1985). "Occupational characteristics of artists: A statistical analysis". *Journal of Cultural Economics* 9, 13–34.
- Wassall, G.H., Alper, N.O. (1990). "When is an artist an artist?" *Journal of Arts Management and Law* 19, 37–50.
- Wassall, G.H., Alper, N.O. (1992). "Toward a unified theory of the determinants of the earnings of artists". In: Towse, R., Khakee, A. (Eds.), *Cultural Economics*. Springer-Verlag, Berlin/Heidelberg.
- Wassall, G.H., Alper, N.O., Davison, R. (1983). *Art Work: Artists in the New England Labor Market*. New England Foundation for the Arts, Cambridge, MA.
- Withers, G. (1985). "Artists' subsidy of the arts". *Australian Economic Papers* 24, 290–295.

## Further reading

- These references were used for background. For the most part, they are not explicitly included in the paper because they are primarily based on cross-sectional data.
- Abbing, H. (2002). *Why Are Artists Poor? The Exceptional Economy of the Arts*. Amsterdam University Press, Amsterdam.
- Alper, N., Galligan, A. (1999). "Recession to renaissance: A comparison of Rhode Island artists 1981 and 1997". *Journal of Arts Management Law and Society* 29, 178–204.
- Alper, N., Wassall, G. (1984). "Performing artists and their unions". Tenth Annual Convention of the Eastern Economic Association.
- Alper, N., et al. (1996). *Artists in the Work Force: Employment and Earnings, 1970 to 1990*. National Endowment for the Arts, Research Division, Washington, DC.
- Bardez, C., Throsby, D. (1997). *Similarity and Difference: Craftspeople and Visual Artists in Australia*. Australia Council, Sydney.
- Bielby, W., Bielby, D. (1996). "Women and men in film: Gender inequality among writers in a culture industry". *Gender and Society* 10, 248–270.
- Bielby, W., Bielby, D. (1999). "Organizational mediation of project-based labor markets: Talent agencies and the careers of screenwriters". *American Sociological Review* 64, 64–85.
- Bielby, W., Bielby, D. (2001). "Audience segmentation and age stratification among television writers". *Journal of Broadcasting and Electronic Media* 45, 391–412.
- Bradshaw, T. (1984). "An examination of the comparability of the 1970 and 1980 census statistics on artists". In: Hendon, W.S., et al. (Eds.), *The Economics of Cultural Decisions*. Association of Cultural Economics, Akron, OH.
- Burke, A. (1994–1995). "Employment prospects in the Irish popular music industry". *Journal of the Statistical and Social Inquiry Society of Ireland* 27, 93–114.
- Cerf, J.G. (1982). *Craft Artists in the United States*. Mathematica Policy Research, Princeton, NJ.

- Chartrand, H. (1994). *Architecture & Design Arts Occupations, 1940–1990*. National Endowment for the Arts, Research Division, Washington, DC.
- Chung, K., Cox, R. (1994). "A stochastic model of superstardom: An application of the Yule distribution". *The Review of Economics and Statistics* 76, 771–775.
- Cowen, T. (1996). "Why women succeed, and fail, in the arts". *Journal of Cultural Economics* 20, 93–113.
- Cowen, T., Grier, R. (1997). "Do artists suffer from cost-disease?" In: Towse, R. (Ed.), *The Arts, the Heritage and the Media Industries*. Elgar, Cheltenham, pp. 371–390.
- Cowen, T., Tabarrok, A. (2000). "An economic theory of avant-garde and popular art, or high and low culture". *Southern Economic Journal* 67, 232–253.
- Dowd, T., Blyler, M. (2002). "Charting race: The success of black performers in the mainstream recording market, 1940 to 1990". *Poetics* 30, 87–110.
- Ekelund, R. (2000). "The 'Death-Effect' in art prices: A demand-side exploration". *Journal of Cultural Economics* 24, 283–300.
- Faine, H.R. (1972). "Unions and the arts". *American Economic Review* 62 (2), 70–77.
- Galligan, A., Alper, N. (1998). "Characteristics of performing artists: A baseline profile of sectoral crossovers". *Journal of Arts Management, Law and Society* 28, 155–177.
- Galligan, A., Alper, N. (2000). "The career matrix: The pipeline for artists in the United States". In: Cherbo, J., Wyszomirski, M. (Eds.), *The Public Life of the Arts in America*. Rutgers University Press, New Brunswick, NJ.
- Galloway, S., Lindley, R., Davies, R., Scheibl, F. (2002). "A balancing act: Artists' labour markets and the tax and benefit systems". Arts Council of England, Research Report 29, London.
- Goldin, C., Rouse, C. (2000). "Orchestrating impartiality: The impact of 'Blind' auditions on female musicians". *American Economic Review* 90, 715–741.
- Grant, D. (1990). "Artists' consensus is they don't count". *Boston Herald*, March 27, 46.
- Gray, C.M. (1984). "Nonpecuniary rewards in the performing arts labor market: A case study of dancers and choreographers". In: Hendon, W.S., et al. (Eds.), *The Economics of Cultural Industries*. Association for Cultural Economics, Akron, OH.
- Gray, L., Seeber, R. (Eds.) (1985). *Under the Stars: Essays on Labor Relations in Arts and Entertainment*. Cornell University Press, Ithaca, NY.
- Greffé, X. (2002). *Arts and Artists from an Economic Perspective*. UNESCO Publishing, Paris.
- Hansmann, H., Santilli, M. (2001). "Royalties for artists versus royalties for authors and composers". *Journal of Cultural Economics* 25, 259–281.
- Heckathorn, D., Jeffri, J. (2001). "Finding the beat: Using respondent-driven sampling to study jazz musicians". *Poetics* 28, 307–329.
- Heikkinen, M. (1995a). "Evaluating the effects of direct support on the economic situation of artists". *Journal of Cultural Economics* 19, 261–271.
- Heikkinen, M. (1995b). "The Nordic model for supporting artists: Public support for artists in Denmark, Finland, Norway and Sweden". The Arts Council of Finland, Research Report 26, Helsinki.
- Heikkinen, M., Karhunen, P. (1993). "The focus and functions of public support for artists in Finland". *Conference on Social Theory, Politics and the Arts*.
- Heikkinen, M., Karhunen, P. (2000). "The importance of public support for creative and performing artists: A case study of authors and actors". 11th International Conference on Cultural Economics.
- Heilbrun, J., Gray, C.M. (2001). *The Economics of Art and Culture*, second ed. Cambridge University Press, Cambridge.
- Hill Strategies (2004). "A statistical profile of artists in Canada, statistical insights on the arts". Vol. 3, No. 1. Hamilton, Canada.
- Horowitz, H. (1983). "Work and earnings of artists in the media fields". *Journal of Cultural Economics* 7, 69–89.
- Horowitz, H. (1993). "The status of artists in the USA". *Journal of Cultural Economics* 17, 29–48.
- Jeffri, J. (1999). "The playwright's the thing!" *Conference on Social Theory, Politics and the Arts*.
- Jeffri, J. (2003). *Changing the Beat: A Study of the Work Life of Jazz Musicians*. National Endowment for the Arts, Washington, DC.



- Jeffri, J., Throsby, D. (1994). "Professionalism and the visual artist". *European Journal of Cultural Policy* 1, 99–108.
- Karhunen, P. (1996). "The interaction between artists' professional training and employment in the field of Finnish theatre". *Journal of Cultural Economics* 20, 165–175.
- Karhunen, P. (1998a). "Labour market situation of graduated artists". In: Heikkinen, M., Koskinen, T. (Eds.), *Economics of Artists and Arts Policy*. Arts Council of Finland, Helsinki.
- Karhunen, P. (1998b). "A portrait of a woman as an artist." 10th International Conference on Cultural Economics.
- Karhunen, P., Rensujeff, K. (2002). "Preliminary results for a survey on the economic and labour market situation of Finnish artists". 12th International Conference on Cultural Economics.
- Karttunen, S. (1998b). "Profile of the artistic labour force in Finland based on the population censuses 1970–1995". 10th International Conference on Cultural Economics.
- Karttunen, S. (2001a). "How to make use of census data in status-of-the-artist studies: Advantages and shortcomings of the Finnish register-based census". *Poetics* 28, 273–290.
- Karttunen, S. (2001b). *Cultural Employment in Finland 1970–1999*. Statistics Finland, Helsinki. In Finnish with English summary.
- Kaun, D. (1991). "Writers die young: The impact of work and leisure on longevity". *Journal of Economic Psychology* 12, 381–399.
- Landes, W. (2001). "What has the visual Artist's Rights Act of 1990 accomplished?" *Journal of Cultural Economics* 25, 283–306.
- Laplante, B., Hébert, B., Menger, P. (n.d.). "Collective and individual action on a labor market: The case of new French actors". University of Quebec, Montreal.
- McLain, J.J. (1978). "The income of visual artists in New Orleans". *Journal of Cultural Economics* 2, 63–76.
- Menger, P. (1999). "Artistic labor markets and careers". *Annual Review of Sociology* 25, 541–574.
- Menger, P. (2001). "Artists as workers: Theoretical and methodological challenges". *Poetics* 28, 241–254.
- Menger, P. (2006). "Contingent work, excess supply and occupational risk management". In: Ginsburgh, V., Throsby, D. (Eds.), *Handbook of the Economics of Art and Culture*, vol. 1. Elsevier, Amsterdam. Chapter 22 in this volume.
- Nagel, J.J. (1987). "An examination of commitment to careers in the musical performing arts: Implications for alienation from vocational choice". In: Shaw, D., et al. (Eds.), *Artists and Cultural Consumers*. Association for Cultural Economics, Akron, OH, pp. 101–109.
- Nagel, J.J. (1988). "Identity and career choice in music". *Journal of Cultural Economics* 12, 67–76.
- National Endowment for the Arts (1981). *Artists Compared by Age, Sex and Earnings in 1970 and 1976*. Research Division Report No. 12. Publishing Center for Cultural Resources, New York.
- Netzer, D., Parker, E. (1993). *Dancemakers*. National Endowment for the Arts, Washington, DC.
- Nichols, B. (2004). "Artist labor force by state, 2000". National Endowment for the Arts, Research Note #85. Washington, DC.
- Nixon, L., Robinson, M. (1992). "The earnings of male and female artists". Eastern Economics Association.
- O'Brien, J. (1997). *Professional Status as a Means of Improving Visual Artists' Incomes*. Arts Council of England, London.
- O'Brien, J., Feist, A. (1997). *Employment in the Arts and Cultural Industries: An Analysis of the Labour Force Survey and Other Sources*. The Arts Council of England, London.
- Rengers, M. (1998b). "Visual artists in the Netherlands: Civil servants, self-employed or both?" In: Heikkinen, M., Koskinen, T. (Eds.), *Economics of Artists and Arts Policy*. Arts Council of Finland, Helsinki.
- Rengers, M., Plug, E. (2001). "Private or public? How Dutch visual artists choose between working for the market and the government". *Journal of Cultural Economics* 25, 1–20.
- Rensujeff, K. (2003). "The status of the artist in Finland – Report on employment and income formation in different fields of art". Arts Council of Finland, Research Report 27, Helsinki. In Finnish with English summary.
- Robinson, M., Montgomery, S. (1993). "The earnings and labor supply of artists: The 'Myth of the Starving Artist' revisited". Xerox.

- Robinson, M., Montgomery, S. (2000). "The time allocation and earnings of artists". *Industrial Relations* 39, 525–534.
- Rottenberg, S. (1997). "The remuneration of artists". In: Towse, R. (Ed.), *The Arts, the Heritage and the Media Industries*. Elgar, Cheltenham, pp. 590–594.
- Rushton, M. (2001). "The law and economics of artists' inalienable rights". *Journal of Cultural Economics* 25, 243–257.
- Ruttenberg, Kilgallen, Gutchess, Associates (1981). *A Survey of Employment, Unemployment and Underemployment Among Performing Artists, 1980*. Labor Institute for Human Enrichment, Washington, DC.
- Singer, L. (1981). "Supply decisions of professional artists". *American Economic Review* 71, 341–346.
- Stoyle, J. (1970). *Economic and Demographic Characteristics of Actors' Equity Association Membership*. Temple University, Philadelphia, PA.
- Summerton, J. (1999). *Artists at Work, 1999*. Southern Arts Board, UK.
- Throsby, D. (1994b). "The production and consumption of the arts: A view of cultural economics". *Journal of Economic Literature* 32, 1–29.
- Throsby, D. (1997). "On the economic survival of actors and musicians". *Revista Economica de Catalunya* 31, 71–75.
- Throsby, D. (2001). "Defining the artistic workforce: The Australian experience". *Poetics* 28, 255–271.
- Throsby, D., Towse, R. (1994). "On labor markets in the arts: Some unresolved issues". *American Economic Association*.
- Towse, R. (1992b). "The labour market for artists". *Ricerche Economiche* XLVI, 55–74.
- Towse, R. (1992c). "Market value and artists' earnings". *International Center for the Arts*.
- Towse, R. (1993). *Singers in the Marketplace: The Economics of the Singing Profession*. Oxford University Press, Oxford.
- Towse, R. (1996a). *The Economics of Artists' Labour Markets*. The Arts Council of England, London.
- Towse, R. (1996b). "Economics of training artists". In: Ginsburgh, V., Menger, P. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier/North-Holland, Amsterdam, pp. 303–329.
- Towse, R. (1999). "Copyright, risk and the artist: An economic approach to policy for artists". *International Journal of Cultural Policy* 6.
- Towse, R. (2006). "Human capital and artists' labour markets". In: Ginsburgh, V., Throsby, D. (Eds.), *Handbook of the Economics of Art and Culture*, vol. 1. Elsevier, Amsterdam. Chapter 24 in this volume.
- Towse, R. (n.d.). "Survey of artists in Wales: Econometric results". Xerox.
- Waits, C.R., McNertney, E.M. (1984). "An economic model of artistic behavior". *Journal of Cultural Economics* 8, 49–60.
- Waits, C.R., McNertney, E.M. (1989). "The incomes of cultural providers: A review of current research". In: Shaw, D.V., et al. (Eds.), *Cultural Economics 88: An American Perspective*. Association for Cultural Economics, Akron, OH.

## HUMAN CAPITAL AND ARTISTS' LABOUR MARKETS

RUTH TOWSE

*Erasmus University Rotterdam, The Netherlands*

### Contents

Abstract	866
Keywords	866
1. Introduction	867
2. The theory of human capital	868
2.1. The basic theory	868
2.2. Sorting models	870
3. Earnings functions	871
3.1. Estimation	871
3.2. Artists' earnings functions	873
3.3. Methodological aspects	875
4. Superstars, talent and creativity	876
5. Artists' training and occupational choice	879
5.1. Expected income	879
5.2. The role of training	880
5.2.1. The content of training	881
5.2.2. Certification in artists' labour markets	881
5.2.3. Students' expectations	882
5.3. Does the human capital model explain artists' training decisions?	883
6. Human capital, social capital, cultural capital and their implications for training artists	884
7. Human capital theory, copyright law and reproducibility	888
8. Conclusion	890
References	892

**Abstract**

It is argued that human capital theory applies only weakly to artists' decisions about investment in schooling and training and about occupational choice. However, the same can be said about the sorting model. What is lacking in cultural economics is an understanding of talent and creativity, what economic factors motivate artists and how creativity can be encouraged as part of government cultural policy. Bringing social and cultural capital into the equation do not seem to add much in the way of understanding artists' labour markets. A novel argument is made that the reproducibility of works of art in combination with copyright law alters the established view that human capital cannot be separated from labour, in this case that of the artist.

**Keywords**

human capital, labour markets, artists' training

*JEL classification:* Z1, J24, I21, K11, H4

*Lady Bracknell.* (to her daughter's suitor) . . . Do you smoke?

*Jack.* Well, yes, I must admit I smoke.

*Lady Bracknell.* I am glad to hear it. A man should always have an occupation of some kind. There are far too many idle men in London as it is. How old are you?

*Jack.* Twenty-nine.

*Lady Bracknell.* A very good age to be married at. I have always been of the opinion that a man who desires to get married should know either everything or nothing. Which do you know?

*Jack.* I know nothing, Lady Bracknell.

*Lady Bracknell.* I am pleased to hear it. I do not approve of anything that tampers with natural ignorance. Ignorance is like a delicate exotic fruit; touch it and the bloom is gone. The whole theory of modern education is radically unsound. Fortunately in England, at any rate, education produces no effect whatsoever. If it did, it would prove a serious danger to the upper classes, and probably lead to acts of violence in Grosvenor Square. What is your income?

*Jack.* Between seven and eight thousand a year.

Oscar Wilde, *The Importance of Being Earnest* (Act I). First published in 1899.

## 1. Introduction

The role of human capital in labour economics has a long history, going back to Adam Smith. Smith recognised the effect of both training and talent in determining wages, the latter in his famous comment on the 'exorbitant' rewards of opera singers and dancers.<sup>1</sup> The influence of innate ability and knowledge acquisition on earnings has been much discussed in human capital theory and this discussion is particularly relevant to artists. The study of artists' labour markets is important in cultural economics because we need to understand what factors affect the supply of work by creative artists and performers since cultural policy, whatever its explicit aims, is ultimately designed to encourage creativity.

In this chapter, two basic questions are addressed: what contribution does human capital theory make to understanding creativity in the arts and culture and what contribution does cultural economics make to our understanding of human capital? Much of the analysis of human capital over the last 30 years has been about the econometric problems of identifying the specific contribution to earnings of innate ability rather than of ability acquired through 'schooling'. Because the role of innate ability or talent is far greater in the arts than it is in non-arts occupations, its influence is an area in which cultural economics can make a contribution to human capital theory. It also seems likely that on-the-job training and experience are more important in the arts than in other

<sup>1</sup> See Smith (1976, p. 124).

occupations. Another distinguishing feature of artists' supply behaviour is their concern with utility and reputation, which considerably modifies their desire for financial reward. Furthermore, the arts and cultural industries are areas in which there is dependence on copyright law for protecting artists' earnings, enabling them to obtain future as well as present income. There is, therefore, reason to believe that artists' labour markets differ from those of other workers and this raises the question whether human capital theory applies in them. However, while there has been a great deal of empirical work on the role of human capital in 'ordinary' labour markets, there have been relatively few econometric studies of artists' labour markets.

This chapter is organised as follows: an introduction to human capital theory precedes a brief summary of issues in the measurement of human capital using earnings functions, including those relating to artists. Then the role of talent and creativity in artists' labour markets is discussed, with particular reference to the superstar phenomenon. That is followed by a section on artists' training and occupational choice, after which we consider the analogy between human capital, social capital and cultural capital. We then turn to the relation in artists' labour markets between human capital, the ability to reproduce artists' work and copyright law, showing how that alters the 'in-alienability' problem in human capital.

## 2. The theory of human capital

### 2.1. *The basic theory*

Sherwin Rosen has defined human capital as "... the stock of skills and productive knowledge embodied in people. The yield or return on human capital investment lies in enhancing a person's skills and earning power, and in increasing the efficiency of economic decision-making both within and without the market economy" [Rosen (1987, p. 682)]. This definition captures two essential features of the theory: that human capital cannot be separated from the person, and that human capital embodied in an individual may be increased by investment. What it does not recognise, however, is the ambiguity of the concept of human capital as a combination of inherited characteristics, tacit knowledge, innate ability and acquired skills; each plays some role in the individual's productivity and earning power but how much influence is exerted by one or the other has proved difficult to pin down.

A thumbnail sketch of the theory is as follows: though the concept of human capital had been recognised by Adam Smith and Alfred Marshall, it was only in the 1960s that Theodore Schultz (1961) set the stage for the present day interest in human capital theory. The fundamental conceptual framework of analysis for virtually all subsequent work in this area was developed by Gary Becker (1964) in his path-breaking book *Human Capital*. In it, he introduced the distinction between specific and general labour training, arguing that schooling (formal education) was in fact a leading example of human capital formation by general training. Jacob Mincer (1958) then specified the now

standard human capital earnings function that has given rise to a huge literature on the measurement of lifetime income and wealth [Willis (1986)].

Adam Smith had understood the essence of the notion of human capital investment: the formation of human capital through costly education, the necessity for higher earnings to compensate those who had made the investment in human capital, and the accrual of these earnings over a lifetime.

A man educated at the expense of much labour and time to any of those employments which require extraordinary dexterity and skill, may be compared to [an expensive machine]. The work which he learns to perform, it must be expected, over and above the usual wages of common labour, will replace to him the whole expense of his education, with at least the ordinary profits of an equally valuable capital. It must do this too in a reasonable time, regard being had to the very uncertain duration of human life, in the same manner as to the more certain duration of the machine.

The difference between the wages of skilled labour and those of common labour, is founded upon this principle [Smith (1976, pp. 118–119)].

Marshall also pursued the analogy between physical and human capital and between the wage rate and the rental rate of capital. He emphasised the difference between physical and human capital on the grounds that there is no market in human capital but rather a market in the hire of labour embodying the investment. This is what Rosen (1987, p. 682) refers to as “differences in the nature of property rights between them”. Furthermore, Marshall made the case for subsidies to education based on the implications of social inequality of access to the capital market for the finance of education [Blaug (1970, pp. 3–6)].

Blaug (1970) raised the question whether the concept of human capital is perhaps no more than a metaphor, and moreover whether all education indeed has an investment motive or whether it is not to some extent a consumption good. Education and learning may yield utility directly to the individual rather than a deferred utility of potential higher earnings. Furthermore, occupational choice may not be determined solely by financial reward because people may choose an occupation for non-pecuniary motives such as a preferred lifestyle. Occupations that require a higher level of human capital investment and so pay more are also ones, contrary to Adam Smith, that may be more attractive on non-monetary grounds. The identification of consumption elements, utility and the pursuit of non-monetary rewards might therefore be difficult to disentangle. These observations are particularly relevant to artists' training and occupational choices. We return to a discussion of the usefulness of the capital metaphor in Section 6 below.

For Becker, however, the analogy between physical and human capital is central. According to Becker, individuals invest in human capital formation up to the point at which the discounted costs of formal education and on-the-job training equal the discounted future earnings over the individual's lifetime. The rate of return that equates these two streams must in equilibrium equal the rate of interest, that is, the cost of borrowing the outlay on the investment. The private rate of return, which accrues to the individual,

is calculated from the out-of-pocket (direct) costs of schooling and the indirect costs of earnings forgone during the investment period. The social rate of return takes into account the return to society via income taxes and the total cost of providing schooling. Empirical estimates of the two rates of return produce the result that the private rate typically exceeds the social rate, if only because the social cost of schooling exceeds its private cost.

Like Marshall, Becker recognised that all families do not have equal access to financial resources and that affects the human capital investment decision. Family background may exert other influences on the costs and returns to human capital investment; inherited characteristics and early advantages such as health, greater investment during childhood, above average ability and innate talent all mean that some individuals learn more easily than others and therefore achieve higher productivity from a given period of study or, *pari passu*, need to spend less to achieve the same learning outcome. These family background and interpersonal differences give rise to the so-called ‘ability bias’ in measuring the rate of return to investment in human capital. We shall see later that this is an important topic in artists’ labour markets.

Becker also considered the question of who pays for post-compulsory education and training, the individual or the employer. The employer has no incentive to pay for general training that can be transferred between firms but does have the incentive to pay for firm-specific training.<sup>2</sup> Firms may well offer general training to employees but then ‘charge’ them for it by paying them lower wages (as in an apprenticeship training).

## 2.2. *Sorting models*

Human capital theory in its Chicago School version (Schultz, Becker, Mincer and Rosen) is not accepted by everyone in the economics of education or labour economics. The most fundamental criticism comes from those who support an alternative explanation of the observed positive relationship between investment in schooling and higher earnings – the screening hypothesis. Screening, which is always linked to signalling (the education system screens and students signal), is a form of sorting; indeed Weiss (1995) recommends using the generic term ‘sorting’ to include both aspects. According to sorting models, employers use educational choice to draw inferences about unobserved attributes that are correlated with schooling. Employers use formal qualifications (a university degree, for example) as an information signal about worker quality, but information is asymmetric – workers know their own productivity but employers cannot tell which workers are the most productive. Workers signal their superior productivity to employers by acquiring paper qualifications which high ability students acquire more easily. This ‘self-selection bias’ is exacerbated by the finding that many highly educated students come from higher socio-economic family backgrounds. The extreme version

<sup>2</sup> It has, however, proved difficult to find convincing examples of specific training other than the necessary learning period that marks the first few days or weeks of employment for any worker.



of the sorting model combines screening and signalling to conclude that higher earnings do not reflect higher marginal productivity at all, but only society's institutional commitment to 'credentialism'. Ability bias and signalling are difficult to separate and in general it has proved impossible to identify them empirically, at least so as to persuade the sceptic. All agree that more time spent in schooling yields higher earnings but sorting models dispute precisely what it is that leads to greater productivity.

Becker has argued that a refutation of sorting models lies in the fact that the education system is an inordinately expensive screening device; such inefficiency is unlikely to develop spontaneously in a competitive economy and is even more unlikely to be sustainable over long periods of time. This argument carries greater force in the USA where students pay for further and higher education, but in many European countries both are either free or available at very low fees. So the direct cost of post-compulsory education in Europe is small, leaving only the cost of forgone earnings to be borne by students and their parents. Indeed, many European countries even offer students grants and rewards for completing a course. In the arts, we see courses being oversubscribed as hopeful students who do not have to pay the full cost of training crowd into art, music and acting colleges [Towse (1996)]. We also observe that employers in the arts place little reliance on certification based on formal schooling and often use their own screening devices, a refutation of Becker [Towse (1993)]; we return below to these and other differences between artists' and other labour markets.

Before leaving the general theory of human capital, it is worthwhile anticipating a later discussion of the way copyright law (in combination with 'reproducibility' by means of copying technologies) alters property rights in human capital in the cultural sector. Many writers [for example, Blaug (1970)] have dwelt on the absence of a market for human capital separated from the labour of the individual and have claimed that only a slave economy would permit direct purchase of human capital analogous to the purchase of physical capital. However, human capital embodied in works protected by intellectual property rights can be 'alienated' by the assignment or transfer of the right to use them. For example, take the case of a sound recording: copyright law provides a number of rights for composers and performers, who use their skill and labour to create works that are recorded in a CD; but once these rights have been transferred to the record company, they become its assets, which it can exploit or sell as it sees fit. Copyright is therefore intimately tied up with the appropriation of artists' human capital. Nor is this only a feature of artists' labour markets; the growth of the 'Information Society' or the 'Knowledge Economy' and the spread of intellectual property law into ever more sectors of the economy have created similar conditions in other labour markets.

### 3. Earnings functions

#### 3.1. Estimation

Earnings data are the single most important source of information about human capital as they represent both the returns to investment and the cost of the time taken to make the

investment. Lifetime earnings are typically represented in age-earnings profiles. Earning starts when compulsory education ceases – if the legal school leaving age is, say, 16 years of age, any schooling beyond 16 incurs the opportunity cost of the earnings of 16+ year olds. Thus the age-earnings profile begins at age 16 and continues to retirement age. Normal age-earnings profiles display a common pattern: from the age of entry into the labour force they rise, then flatten out at around mid-career and fall towards retirement age.

Investment in human capital does not cease with the start of work, however, since on-the-job training now begins. On-the-job training is a loose concept that includes experience or learning-by-doing, which increases with age. Mincer (1958) recognised that workers could also choose at each point to invest in formal on-the-job training as a substitute for years of schooling and that that would eventually be compensated by higher earnings. Workers may rationally choose different jobs that enable them to gain experience, accepting lower earnings while they are trained on the job. Mincer analysed these different choices by individuals as yielding a series of age-earnings profiles whose shape is determined by the different combination of direct and indirect costs of education and on-the-job training. Because they represent different combinations of earnings forgone and lifetime earnings, the age-earnings profiles must cross. Mincer called this the ‘overtaking point’, which he argued would depend on the reciprocal of the internal rate of return, the discount rate that equates the stream of lifetime earnings with the cost of the investment. This approach enables the researcher to take into account the effect of family background effects such as financial constraints, but it does not solve the problems of ability bias or self-selection bias.

The Mincer earnings function has become the standard model for statistical measurement of the supply of educated labour and for estimating the internal rate of return to education:

$$\log Y = \log Y_0 + rS + b_1X + b_2X^2 + u, \quad (1)$$

where  $Y$  is income,  $S$  is length of schooling,  $X$  is length of time in the workforce as a proxy for years of work experience and  $u$  is the error term; the constant term is the log of the equivalent annuitised income of initial human capital value (innate and family background effects),  $r$  is the rate of return to schooling,  $b_1$  and  $b_2$  capture the effects of experience [Rosen (1992, p. 162)]. Several econometric specifications of the earnings function have been tried out and innumerable empirical studies have been made; a survey of the early literature is to be found in Willis (1986).

The ideal data for measuring the effect of investment in human capital would consist of longitudinal information on individuals’ lifetime earnings combined with individual tuition expenses. Such data are rarely available and in practice aggregate cross section or panel data are used; however, they abstract from individual decision-making and give rise to bias in estimation. Simplifying assumptions are therefore needed to enable researchers to use the data available. Briefly these are: that the only cost of education is earnings forgone; that individuals enter the labour force immediately on complet-

ing their studies; and that the individual's working life is independent of her years of education.

The problem of estimating the ability bias is aggravated by the use of aggregate cross-section data; individual data would avoid much of this problem by providing information on family background and possibly even on early IQ and other measures of innate ability. In the attempt to separate out the influence of innate ability and human capital investment on earnings differentials, studies of identical and non-identical twins have been used. Rosen (1987) reports that such studies have found that around one-third of the difference in earnings was due to differences in human capital investment, one-third was due to 'person effects' (unmeasured ability, health and other such factors) and one-third was due to random events, luck and suchlike. These findings are particularly relevant to artists, for whom variations in talent and luck may well be higher than for other professionals. This is discussed in more detail below.

### 3.2. Artists' earnings functions

We now turn to empirical studies that have been done of earnings functions in arts occupations. These studies have tested the human capital model outlined above, treating artists as workers like any others making rational choices about investment in education and on-the-job-training and occupation. There are considerable difficulties concerning the definition of the artist population and of obtaining data.<sup>3</sup> Suffice it to say here that in most countries in which these studies have been attempted, there are severe problems in obtaining artists' earnings data. Census data either cannot be disaggregated to appropriate artistic occupational levels or are subject to severe bias due to multiple job-holding. In the USA, Filer (1986) used Census data to estimate earnings functions for a range of artistic occupations but the validity of his results have been widely disputed by cultural economists because he used aggregate income data [Towse (2001a, Chapter 3)]. Consequently, cross-section survey data have been used in preference to Census data but response rates may be low and sample sizes may be relatively small and possibly unrepresentative. Interestingly, however, Filer found only a very weak effect of human capital variables on artists' earnings. In a later study in which he analysed earnings functions separately for different arts occupations, still using Census data, he found that longer schooling even had a negative effect in the case of performing artists [Filer (1990)]. Wassall and Alper (1985) did one of the first earnings-function studies of artists using data from their survey of 3000 artists in New England that allowed them to separate arts from non-arts income. They found that education was not positively correlated with income from arts work though it was from non-arts work, a finding that has been since replicated in other studies.

Large-scale national surveys of artists' earnings have been carried out in Australia and enabled two authors Withers (1985) and Throsby (1992, 1994, 1996) to estimate

<sup>3</sup> See Chapter 23 by Alper and Wassall in this volume; also McNertney and Waits (1997).

earnings functions.<sup>4</sup> Withers varied the standard Mincer specification by making a distinction between formal education and other qualifications (trade and technical certificates), using the imputed hourly wage rate and reported hours of work to measure earnings. He noted that the standard assumption of continuous employment from entry into the labour market to retirement age overstates work experience for females, a higher proportion of whom are in the arts compared to the non-arts workforce. This, however, would apply to all artists since many artists work irregularly and have frequent periods of unemployment during job searches. Withers' results showed that, compared to all Australian workers, artists earned 40 percent less, which he interpreted as the 'subsidy of artists to the arts'. The 40 percent penalty can also be viewed as a compensating differential, the 'psychic income' for the net advantages of a preferred occupation – the interpretation depends on whether or not you assume that changing occupations is frictionless. On the latter point, Filer (1987) found that the penalty for the choice of artistic occupation in the USA was not high and that 'failed' artists were able to move into other occupations without a high earnings penalty – a striking testimony to the power of general training. Withers also found that human capital variables had only a very weak effect on earnings and concluded that innate characteristics (talent and motivation) and luck, though not identified, must play a considerable role in determining earnings in the arts.

Using the same data set, Throsby (1994) estimated an important variation on the standard form of the earnings function. The Australian survey (which he had directed) collected data on earnings and hours of work in arts and non-arts work. This was done because it was known from qualitative research that artists typically divide their work time between their chosen artistic occupation (arts work) and jobs outside the arts (non-arts work), mainly because they are unable to earn a living wage from the former. This approach was a means of overcoming the bias introduced in census data, which defines artists on the basis of their occupation in census week and attributes earnings from all sources to their arts work, even if a substantial part is due to non-arts work; this was a criticism that had been made of Filer's 1986 study (which he had defended as a 'market test' of who is and who is not an artist). Throsby therefore was able to estimate earnings functions for arts and non-arts work and he was able to separate out the private rate of return to education, training and experience in both arts work and in non-arts work. That had the advantage that he was able to compare rates of return for the same sample, unlike Withers (and others) who had made comparisons between aggregations of different sets of individuals. Throsby found that human capital investment *was* an explanatory factor of income differentials in both sectors, though the human capital model performed less well for arts work than for non-arts work, again due to the unspecified effect of talent and other innate ability factors. It should be noted, however, that the studies by both Withers

<sup>4</sup> Withers' estimation was done in 1984 and Throsby's in 1992; the references cited are reprinted in Towse (1997). The McNertney and Waits survey was done in 1988 and is also reprinted in Towse (1997), the earlier version now being difficult to obtain.

and Throsby had low  $R^2$  values, indicating that much about earnings differentials was left unexplained.

Subsequently, Throsby (1996) estimated another earnings function for Australian artists using data from a later survey and testing two different specifications, one a linear model using two-stage least squares and a second using the Mincerian earnings function (Equation (1) above), again breaking down earnings and hours worked, this time identifying arts, arts-related (such as art teaching) and non-arts work. His hypothesis was that income from arts and arts-related work are influenced by the level of professional arts training, whereas non-arts income is more likely to be influenced by the level of general education, and that time spent on-the-job as an artist is the appropriate explanatory variable for arts and arts related work, with age (as a proxy for experience) being more relevant to non-arts earnings. Dummy variables for level of training, education and gender were used. The models were tested with a further division between creative artists (writers, composers, choreographers and so on) and performers (actors, musicians, dancers and the rest). Interestingly, this elaboration of his earlier study did not yield very different results. The linear model performed better than the standard Mincer model and the hypotheses of the standard model were confirmed. Even so,  $R$ -squares for arts income were low, indicating that factors other than human capital were at work. The experience of testing human capital models in the arts has led Throsby to develop a work-preference model of artist behaviour [Throsby (1994)] as an alternative to the human capital investment model. Cowen and Tabarrok (2000) also develop a utility-based model. It remains to be seen how these models stand up to empirical testing.

### 3.3. Methodological aspects

By way of conclusion to this section on estimating earnings functions, it is worth considering the methodological aspects of human capital theory (methodology being the logic of different methods, not merely a comparison of methods of econometric estimation). Blaug (1976) subjected the theory of human capital to methodological analysis based on Lakatos' concept of a scientific research programme and asked what empirical tests would refute it? Is it 'a theory' or a set of theories – in Lakatosian terms, is there a 'hard core' of theory or just a 'protective belt' of ad hoc empirical generalisations? Though sorting models appear to provide a rival theory, Blaug believed they would eventually become a complement to the human capital hypothesis; moreover, as noted earlier, no discriminating test has been found that could refute one hypothesis and confirm the other, despite a huge battery of empirical work.<sup>5</sup> In the absence of a rival theory, Blaug concluded that the human capital research programme had to be evaluated on its own terms: the predictions of the theory cannot unambiguously be tested because of the unsolved question of the separate influence of innate ability and the assumption

<sup>5</sup> Thirty years later Blaug's insights have been confirmed by Weiss (1995), who states that sorting models are a refinement and an extension of human capital theory rather than an alternative.

of individual rational behaviour regarding the schooling decision. The question therefore is not so much whether schooling explains earnings, a fact commonly accepted, but why it does so.

Despite these objections, Blaug acknowledged that there has been empirical progress in human capital theory in the sense of better data, more sophisticated modelling and econometric analysis. That could also be said of empirical testing of earnings functions in the arts, though there are far, far fewer examples. The absence of reliable data sets still inhibits research in artists' labour markets but what studies there are point to fundamental difficulties. Apart from the obvious question of what is talent – to be discussed below – other problems are present in relation to artists' earnings that are more pronounced than in 'ordinary' labour markets. The distribution of artists' earnings is far more skewed than is found in other occupations and therefore estimates of mean earnings lack conviction.<sup>6</sup> The associated greater variance of artists' earnings points to a far greater risk for artists but whether that suggests that artists are risk-averse and have to put up with greater risk, or are in fact risk-takers is open to debate; however, that is also a question of the validity of the assumptions rather than the accuracy of predictions. Artists' labour markets are dominated by self-employment, with frequent job search and other information problems. Artistic output almost by definition is heterogeneous and demand for it is radically uncertain; *Caves (2000)* has shown that these circumstances lead to contracting problems in the creative industries. Persistent excess supply of artists is widely inferred from the prevalence of unemployment among them [*Towse (2001a)*]. Moreover, there has been little empirical research on the demand for artists, though such studies have increasingly been done for other occupations [*Acemoglu (2002)*].<sup>7</sup> These observations suggest that artists' labour markets may be fundamentally different from other labour markets and that the human capital model is therefore less likely to apply to them.

#### 4. Superstars, talent and creativity

In this section we discuss specific features of artists' labour markets that reinforce the view that they differ fundamentally from other labour markets. Chief among these features is the absence of a clear specification of talent and creativity and their role in artists' labour markets.

*Sherwin Rosen (1981)* in his seminal article 'The Economics of Superstars' focussed on talent as the cause of the skewed distribution of earnings in certain professions and the vastly higher earnings of the few superstars in them.<sup>8</sup> His explanation revolves

<sup>6</sup> See *Seaman (2003)* for recent analysis of variances in artists' earnings and for a comparison with sports-people's earnings.

<sup>7</sup> *Towse (1993)* collected information on the demand for classically trained singers.

<sup>8</sup> See further *Chapter 25* by *Adler* in this volume.

around two causes: on the supply side the development of media technologies, for example sound recording, which have considerably increased economies of scale, enabling artists to serve a far greater market; and on the demand side, consumer preferences for greater rather than for lesser talent when there is imperfect substitutability between suppliers such as artists and entertainers. Superstars, according to Rosen's definition, are people who 'earn enormous amounts of money and dominate the activities in which they engage'; they are highly talented and highly rewarded for their talents because, as Rosen shows, the net revenue earned from talent is a convex function, causing small differences in talent to be magnified into larger differences in earnings [Rosen (1981, p. 845)]. Rosen does not attempt to analyse talent other than by giving some examples – gifted surgeons, sportspeople, singers – but he does state that talent can be ranked; indeed, he specifically avoids the problem of measuring talent by saying 'a cardinal measure of quality or talent must rely on measurement of actual outcomes' [Rosen (1981, p. 848)].<sup>9</sup> Very Chicago!

We might try to analyse talent and creativity in the arts by analogy with the role of innate ability in human capital theory. In relation to education, innate ability reduces the cost of investment in schooling needed to achieve a given outcome, such as a university degree, as it is an argument in the educational production function; the greater the innate ability, the higher the productivity achieved by a given level of investment, or *mutatis mutandis* less investment is needed to achieve a given level of attainment. Innate ability therefore has similar effects to the fertility of land on cultivation. The analogy with the Ricardian theory of rent is a strong one. In that theory differential rents are explained in terms of differential natural fertility of land and the demand for corn; as the demand for corn shifts out (say, due to population growth), land of ever less fertility is brought into cultivation with lower yields of corn. But which is cause and which is effect in determining rents?

Talent can be thought of as akin to the fertility of land, assumed to be a free gift of Nature; it is an inborn asset, which often manifests itself early in life and it enables the 'owner' not only to acquire skills more easily but also to achieve a high level of 'artistic productivity' – great performances – that few competitors can supply. In the human capital model, rents to innate ability accrue on the supply side. By contrast, Rosen's model puts the explanation on the demand side in which the perception of small differences in talent between individual artists causes multiplicative effects to incomes. One might paraphrase the Ricardian argument as follows: 'is the price of opera high because singers' fees are high, or are singers' fees high because the price of opera is high?'

But there still remains the problem noted earlier of the inalienability of human capital. When the output is a personal service such as a surgical operation or a live performance, it can only be supplied in conjunction with labour. The performer must be present to supply her talent live to the audience and therefore limitations on the supply of her time

<sup>9</sup> See Towse (2001a, Chapter 3) for a more detailed discussion of this point.



lead to rewards like rent being determined by demand. Towse (1992) argued that this is the case for the fees of singers. The story changes, however, when the constraints of the performer's time are lifted by making her services available in a reproducible form like a sound recording; this is what Rosen analysed in his model. Recording technology has enabled performers and other artists to reach vastly greater markets, exacerbating the effect of small differences in talent and reducing the possibility of substitution between artists of lesser or greater talent.

It is time to consider what is meant by talent and creativity: 'You've either got it or you haven't' is something frequently said in conservatories of music and colleges of art. As with the inalienability problem in human capital theory, talent does not have a price because there is no direct market for it – you cannot buy yourself some genius to get through university or become a singer (except, perhaps, with a Faustian contract!).<sup>10</sup>

Talent and creativity are widely viewed as the *sine qua non* of art, so what has cultural economics to say about it? A quick answer is that it has been little discussed. Throsby (2001) has a short chapter on creativity, which he links to his concept of cultural value and to a utility model of artistic supply; Frey (2000) discusses the motivation and incentives to create and the role of public support for artists; Castañer and Campos (2002) deal with innovation by arts organisations, adopting a Schumpeterian approach; Caves (2000) deals with what could be called the industrial organisation of creativity, without however explicitly discussing the concept; Towse (2001a) analyses the reward of creativity, again with only a token discussion of creativity, and like Throsby, associates it with artistic supply. All these authors skate around the central issue, which is the contribution of creativity and talent to artists' productivity and earnings.

Creativity is often regarded as an individual activity, though there are many joint creative activities such as theatrical rehearsals and teams of scriptwriters for soap operas.<sup>11</sup> When we speak of creative artists such as composers, authors or painters, they are essentially envisaged as working alone, experimenting with ideas. Another feature of creativity is originality, a spark of novelty that comes to the artist 'out of nowhere' or from reworking existing ideas in new ways. Creativity in these terms clearly parallels innovation and invention in science and technology. Following Schumpeter (1942), the reception of creativity must also be considered [Wijnberg (1995)]. Is it recognised? Can it be marketed? Can it be motivated by financial reward? These are questions that are relevant to the exploitation of talent and creativity by arts organisations and the cultural industries, as well as to cultural policies aimed at fostering the production of art. The economic value of creativity and talent is that they are necessary inputs to satisfy consumer demand for novelty and new experiences and to create lasting works of art.

Enough has been said to indicate that artistic creativity and talent are perceived differently in cultural economics than in human capital theory. Bearing this in mind, we

<sup>10</sup> Whether or not it can be produced through investment in schooling is discussed in the next section on the economics of training.

<sup>11</sup> Seaman (2003) contrasts research in the economics of sport on sportspeople working in teams with the far more limited work in cultural economics on teamwork in the arts.



now turn to the question of artistic training and consider the role of investment in human capital as a preparation for artistic occupations.

## 5. Artists' training and occupational choice

### 5.1. Expected income

According to human capital theory, occupational choice is made on the basis of expected lifetime income. As ability bias reduces the cost of acquiring skills and qualifications, students sort themselves into occupations in which they have a comparative advantage. Taken over a whole society, individual rational behaviour is thought to lead to an optimal allocation of human capital. Equilibrium is achieved in the usual way: excess supply in one occupation reduces earnings and thus the private rate of return, causing workers – at least to some extent – to switch jobs to other occupations. Skill shortages encourage on-the-job training; jobs/professions requiring a greater investment in higher education and professional training offer higher lifetime earnings to compensate for higher costs of study. Experience adds to human capital over the workers' career and is rewarded by higher earnings.

The question is: could this theory apply to artists' labour markets or are they really different from other labour markets? Artists may not be rational wealth maximisers, something that is widely believed and frequently stated by artists themselves [Abbing (2002)], but that is a behavioural assumption that cannot be tested directly. The key question is whether the choice of an arts occupation is based on the private rate of return to investment in human capital that is determined by the costs of schooling and artists' earnings. As reported earlier, Throsby (1992) found some limited support for human capital theory in the arts, so it is worthwhile taking it at face value and seeing what resonance the theory has for artistic training and occupational choice.

Surveys have shown that median earnings in the arts are always lower than those of other equally qualified workers, even though the artist population has a higher than average level of educational attainment. The lower expected lifetime earnings in the arts, combined with the higher indirect costs, ensure that discounted costs exceed discounted benefits in arts occupations [Towse (1996)]. What may modify these results in the arts is the longer working life of some artists; cross-section studies assume a normal age of retirement and that is misleading because retirement may be very late in some artistic professions. Some artists continue to work until they die, and with royalties from copyrights they can continue earning even beyond the grave. Even in the performing arts, where there is a premium on strength and (sometimes) youth, many performers teach and adjudicate long after they retire from performing. On the other hand, the direct cost of training (schooling<sup>12</sup>) performing artists in specialised institutions is higher than the

<sup>12</sup> At the risk of confusing formal training in the sense of schooling with on-the-job-training, I have switched to the common way of speaking of the lessons and other forms of teaching that are provided in specialist

average cost of higher education in general [Towse (1993)]. However, as noted earlier, in many European countries higher education is provided free or at very low fees and even in the USA some students obtain scholarships that cover the cost of training; public finance therefore reduces the direct costs of training and, other things being equal, should increase the private rate of return to training.<sup>13</sup> But even if, following Mincer, we ignore the direct costs of training, the fact that courses in performing arts are longer than those for most other subjects raises the indirect cost of study and would therefore reduce the private rate of return.

Despite the unfavourable prospect of financial reward in arts occupations, higher education courses for artists are typically oversubscribed and that contributes to the oversupply of artists. The question of how many places should be offered in specialist colleges has been a hotly debated one for many years: should the number be restricted so as to ensure that only the most highly talented students receive artistic training, or should more be admitted in the hope that good quality training will enable them to reach a satisfactory standard of competence? It is widely accepted that all students who complete formal training in the creative and performing arts will not be able to make a living from their art. However, when it comes to the public finance of higher education, considerations of equity often override those of efficiency. Moreover it is not easy to define what efficiency would mean in the circumstances of artists' labour markets, given the uncertainty surrounding the chances of success.

### 5.2. *The role of training*

The question still remains whether it is possible to increase an artist's human capital by investment in formal schooling and whether training can add value in the case of less talented students. 'You can't make a silk purse out of a sow's ear' is a saying that resonates through specialist art, music and drama colleges. Even if they have strict entrance requirements, colleges cannot assess the quality of entrants with any certainty – heterogeneity and the absence of objective information about quality pervades all aspects of artists' labour markets. It is also difficult to assess colleges' success in preparing students for work in the arts; the demand for artists' services in the labour market is uncertain and difficult to define since there are few 'regular jobs', with most artists working in self employment on short term contracts [Gurgand and Menger (1996)]. These problems raise a number of points about artists' training that can be analysed

colleges as artists' 'training'. Not all arts training in that sense is confined to specialist art, music or drama colleges, even for performers: writers and composers often do academic courses in universities. There are different institutional arrangements in different countries. The point is that students receive formal education apparently dedicated to artistic occupations.

<sup>13</sup> The social rate of return in specific occupations is used to evaluate public policy decisions about the allocation of educational funding; Towse (1996) suggests the social rate of return to training artists is very low, even negative.

separately (though they may well interact in practice): the content of training; certification in artists' labour markets; and students' expectations about lifetime earnings, the nature of work opportunities and the duration of employment and career possibilities.

### 5.2.1. *The content of training*

Given the importance of talent and creativity in the arts, one might be tempted to conclude that the content of higher education courses in art and music colleges adds nothing to students' innate ability, the more so if only the most gifted students are able to gain entrance. In dance and music most students have already had years of specialist teaching, and in other art forms students are expected to show a portfolio or other such evidence of attainment as an entrance requirement. What therefore can colleges offer such students? Here it is useful to draw a distinction between the 'art' element and the 'craft' element of artistic training. Even very talented students need to learn how to present their work, study repertoire or the craft of drawing and writing, and so on. Colleges also provide facilities that are difficult or expensive for individuals to provide for themselves, such as studios, artists' models, participation in theatrical productions, orchestras and suchlike. Colleges also provide students with a forum for displaying their talents to the outside world in exhibitions, performances, etc., and enable them to develop networks, learn professional conduct and assess their own abilities. This last point is particularly important because information about one's own quality is needed in order to make career choices – whether to choose another occupation, whether to aim high or low, etc. MacDonald (1988) explains drop-out rates after the first few years of work in artistic occupations as the outcome of this search for information. Colleges may perform a preliminary sorting function by providing that information before entrance into the labour market takes place.

As a coda to this brief discussion of a complex subject, it is worth pointing out that surveys of artists' labour markets have revealed that a significant proportion of working artists did not receive specialist arts training.<sup>14</sup> There are two possible explanations: that they made the decision to be artists after having done another higher education course; and/or that they regarded the content of arts training courses as irrelevant, something that is frequently reported in surveys of working artists. Either way, the finding is hostile to the human capital model, as well as to the sorting model, as these artists were able to make their way without either formal training or a certificate and apparently with no earnings penalty [Towse (2001a)].

### 5.2.2. *Certification in artists' labour markets*

Although colleges training artists award degrees and diplomas, studies of artists' labour markets have found that certification apparently plays a less important role in the arts

<sup>14</sup> 30 percent is a not untypical figure; see Towse (2001a).

than in it does other labour markets. Certification is of course highly correlated with formal schooling, which tests students for their achievement in following the curriculum. However, that may not provide the kind of information wanted by employers or others hiring artistic labour for several reasons. Artists typically are self-employed and so need not signal to an employer. A reputation for professionalism and high quality talent/creativity are very important in artists' labour markets and certification by art colleges apparently does not provide adequate information about these characteristics. Employers may not trust colleges' certification because they produce too many graduates, not all of whom are sufficiently talented. Colleges have their own objective functions and they maximise their income from student numbers, and they may also be pressed by funding authorities to offer a mix of services to a mix of students; these factors combine to give higher education institutions training artists the incentive to 'oversupply' graduates. Indeed as stated earlier, many 'employers' who hire artists show no interest in their paper qualifications and set up expensive screening arrangements of potential employees for themselves [Towse (1993, 1996)].

### 5.2.3. *Students' expectations*

Students face problems in forming expectations about the probability of making an artistic career pay because there is little objective information available. Their subjective assessment of their own talent and creativity is likely to be overestimated due to 'the overweening conceit of the young', as Adam Smith put it in the *Wealth of Nations*, and objective information about future earnings and the ease of obtaining work is difficult to obtain in a labour market in which there is a wide dispersion of earnings and frequent job change. Following criticism that young people are let out into the real world without adequate preparation, colleges have made considerable efforts to offer courses on business methods and how to manage a career in the arts but often to little avail; students apparently fail to attend such courses, though they later complain that they 'should have been warned' about the difficulty of making and managing a career [Towse (2001a)]. These problems might seem to support the view that young artists are irrational. However, when 'nobody knows' about quality and demand for artistic output [Caves (2000)], irrationality may be confused with radical uncertainty. Rationality is inevitably bounded in artists' labour markets.

Students also form expectations about non-pecuniary rewards such as the opinion of peers, the desire to work on one's own account and other 'psychic rewards'. These are particularly important to artists: and they accept earnings lower than those available in alternative occupations as a price worth paying for the chance to work in the arts and, as long as they can earn enough to live on from other work (arts-related or non-arts jobs), they do not change occupations. Throsby's evidence shows that once a 'satisficing' level of income has been reached from earnings from all sources, artists devote more time to arts work, eschewing the opportunity to earn more from doing more hours of non-arts work [Throsby (1992)]. Abbing (2002) has reinforced this view with ample anecdotal

evidence from the visual arts world.<sup>15</sup> Artists often report choosing work that offers new challenges rather than repeating a former piece of work even if that would raise their earnings [Jeffri and Throsby (1994)]. This behaviour could be superficially interpreted as opting for utility rather than money but it may also be efficient in improving the artist's reputation, which is the best investment in her future, particularly for self-employed workers [Benhamou (2000)]. It is difficult to know when artists absorb these values – whether they do so as students or after they have entered the labour market. There have been few longitudinal studies of artists' careers by cultural economists that match students' expectations to their labour market experience.<sup>16</sup>

### 5.3. Does the human capital model explain artists' training decisions?

Finally, we come to the question of how well the human capital model performs in explaining artists' decisions on training and career development. If we were to take the extreme view that talent alone determines an artist's career and earnings, investment in schooling would not be worthwhile; by definition it could not raise productivity. However, casual evidence from the biographies of artists goes contrary to that view, as many highly talented artists have trained at art, drama or music college. Of course that could be for institutional reasons and we may question whether an alternative organisation of training such as private lessons could have yielded the same results. Apprenticeships in the performing and visual arts – the only source of training before the nineteenth century – are alternatives that are still sometimes available today, for example, for potters and for opera singers.<sup>17</sup>

However, as noted above, part of the experience of attending college is socialisation and professionalisation. Blaug (1985) has pointed out that the three Ss – skilling, screening and socialisation – are as important in the labour market as the three Rs. All artists need to learn the ropes; networking – forming working partnerships, meeting with other artists and with agents who may in future be in a position to offer work, getting recommendations from well-known teachers – may be more important than the schooling function. Tacit knowledge, trust and reputation, are also important in artists' labour markets. These are all features of social capital; are they also part of human capital formation? An important question is whether they are amenable to investment decisions. Tacit knowledge acquired in childhood for example cannot be regarded as an

<sup>15</sup> Abbing is writing mostly about the position of artists in the Netherlands, where a government support scheme for visual artists was tried in the 1970s, resulting in a vast oversupply of very large works of art. It also led to a considerable oversupply of artists with the result that art prices are low, creating a vicious circle of dependence on state basic income payments [Abbing (2002); see also Rengers and Velthuis (2002)].

<sup>16</sup> See Chapter 23 by Alper and Wassall in this volume for a discussion of longitudinal studies.

<sup>17</sup> Until the mid-twentieth century in Italy, the standard training of singers consisted of the pupil going to live with the Maestro and having daily lessons. Tito Gobbi trained that way and so did Cecilia Bartoli, whose singer parents taught her at home. Conductors often started as repetiteurs in opera houses teaching singers their parts or even, as in the case of Georg Solti, as pianists working for a singing teacher. Opera houses have fairly recently reintroduced apprenticeship schemes for trained young singers to acquire work experience.

investment as it is not deliberately fostered, though early childhood education within the family may be, and often is in the case of children's dance and music lessons [Seaman (2003)]. Acquisition of tacit knowledge and early training certainly seem to play some role in occupational choice in the arts; there are many instances of children following in a parent's footsteps (and not only in the days when there was only on-the-job training). Early acquisition of knowledge is probably easily confused with innate talent in the arts. However, there have been no systematic studies of the influence of family background on occupational choice in the arts and it is all too easy to generalise from the Placido Domingos and Vanessa Redgraves of this world. This topic would be a very interesting research project and could shed light on the role played by social as compared to human capital investment; I return to this matter below.

These qualitative arguments may explain the weak effect cultural economists have found for the influence of human capital on artists' earnings. Schooling helps the artist get her first work assignment and that may be a critical step on the career ladder. Schooling also provides general training that can be used outside the arts and it teaches good networking and other 'social capital' skills. But if investment in human capital only marginally explains the observed high demand for arts training, does that suggest that sorting models perform better? It seems from the earlier discussion that certification also plays an ambiguous role in artists' labour markets. Besides formal schooling, there are other screening devices available such as prizes and competitions, awards from Arts Councils and other forms of informal certification that offer information.<sup>18</sup> However, employers do take note of the fact that artists have attended college because the college screened them on entrance; they also treat attendance as a signal that arts students will have acquired some basic professional skills, even if these are marginal to the innate talent necessary for undertaking work in the arts. It seems that in a situation with oversupply of new entrants and the presence of a sea of amateurs, the position of the potential 'employer' in the arts (as compared to other labour markets) is especially difficult because of information problems; on the other hand, making the right choice matters less in a situation in which frequent job change and working on short term contracts is normal. And, as reported earlier, having a higher education qualification pays off in arts related and non-arts work even if not for artistic work and that is an important consideration for the majority of artists who inevitably hold multiple jobs. Thus, there is some supporting evidence for the sorting model.

In short, the jury is still out on what is an appropriate model to explain artists' training and occupational decisions.

## **6. Human capital, social capital, cultural capital and their implications for training artists**

We now return to the fundamental question of what purpose is served by the use of the term 'capital' in human capital theory. Originally raised by Blaug (1976), who asked

<sup>18</sup> See Wijnberg (2003) for a discussion of the role of awards in the arts.

whether human capital is in fact a useful metaphor, this question is being asked again in relation to social capital and it is also relevant to the notion of cultural capital that has appeared in the literature of cultural economics. Social capital is a concept that originated in social theory. It has provided an umbrella term under which a range of diverse topics has been investigated by economists. Bowles and Gintis (2002, p. F419) define it as follows: "Social capital generally refers to trust, concern for one's associates, a willingness to live by the norms of one's community and to punish those who do not". However, Putnam (2000, p. 19) defines it as "... connections among individuals – social networks and the norms of reciprocity and trustworthiness that arise from them ...". Both aspects have been related to labour markets in general and specifically to human capital. In a recent review article, Sobel (2002) states: "economists find the social capital *metaphor* useful in studies of economic development, transition economies, common-resource property use, and education" (p. 143, my italics). The concept of cultural capital was introduced by Throsby, who defines it as "an asset that embodies, stores or provides cultural value in addition to whatever economic value it may possess", where cultural value is defined as a combination of aesthetic, spiritual, social, historical, symbolic and authenticity values, typically produced by artists [Throsby (2001, p. 46)].<sup>19</sup> Thus we have three somewhat related concepts carrying the label 'capital'. What do these concepts contribute to an understanding of artists' labour markets and indeed how appropriate is the capital metaphor in all these cases?

Since the World Bank's adoption of the concept of social capital, a number of objections have been levelled against it by economists, notably by Kenneth Arrow (2000) and Robert Solow (2000), who judge it according to the standard notion of capital in economics due to Irving Fisher. Arrow's definition of (physical) capital identifies three features: capital is deferred consumption; it can be deliberately accumulated by investment; and it is alienable. Solow added two further criteria: capital has a rate of return that can be used as a measure of its value; and capital depreciates, whether from use or from technological obsolescence. Both Arrow and Solow conclude that social capital does not have these characteristics. As Solow has said "social capital is an attempt to gain conviction from a bad analogy" [quoted in Sobel (2002)]. Many other economists have criticised the concept from various points of view [Fine (2001); Durlauf (2002)].

Be that as it may, it seems that several features bundled together as social capital are relevant to artists' labour markets. How do these features help us understand artists' incentives to invest in acquiring skills and competences? If, as is the strong implication, social capital is a 'public good' (in the technical sense of that term, that is, non-rivalrous and non-excludable), the inference is that individuals would not have the incentive to invest in themselves and therefore artists' training would have to be collectively financed. However, if artists can appropriate the benefits for themselves, as they can with the human capital model, then they would have the incentive to invest

<sup>19</sup> For a full discussion, see Throsby (2001, Chapter 2).



in their own training. Though much of the research on social capital has been on its communal benefits, Glaeser, Laibson and Sacerdote (2002) make the case that standard microeconomic analysis can indeed be applied to social capital formation. Their model of individual investment in social capital, in which individuals choose optimal levels depending upon the opportunity cost of their time and time preference rate, is almost identical to the human capital investment model. Clearly they assume that social capital is not a public good, although it may have some degree of publicness. More important for our present purposes is their assertion that: "(T)he connection between social capital and human capital is one of the most robust empirical regularities in the social capital literature. Better understanding of this connection should be a key goal for future research" [Glaeser, Laibson and Sacerdote (2002, p. F455)]. The clear implication of this is that investment in social capital is analogous to that in human capital – indeed, the two are 'joint products'. But if that is so, there is an identification problem of distinguishing the private rate of return to one from that to the other. That muddies the debate about the relevance of human capital theory in artists' labour markets. It is clearly important for policy purposes to understand their separate contributions to lifetime earnings in view of the public/private dichotomy.

The public/private division also arises in respect of cultural capital. Throsby defines cultural capital as the stock of goods and services that constitute society's cultural assets, which have been created by artists (past and present). He distinguishes tangible from intangible cultural capital: tangible assets may be privately or publicly owned but intangible cultural capital (which overlaps significantly with social capital) is always a public good. Both yield a return of cultural value that Throsby regards as a communal rather than an individual variable. However, the motive for the artist in creating these assets, according to the human capital view (which Throsby has tested more than any other cultural economist), is the desire for private gain, whether pecuniary or non-pecuniary. Thus, artists' human capital creates cultural capital – Adam Smith would have liked the implied doctrine of the unintended consequences of private action, the coincidence of private incentive with public benefit. As with social capital, there are strong externalities present in cultural capital (if not true public goods characteristics) that call for communal rather than individual investment.

One way of identifying the differences between these three obviously related capital concepts is to focus on their implications for cultural policy. A central concern of cultural policy is how society can best invest in the supply of artistic creativity. According to human capital theory, we should encourage artists to raise their productivity through subsidies to formal training courses in colleges and possibly also by giving artists basic income support or price subsidies in order to raise their earnings. Social capital instead suggests that developing social skills, joining professional networks, acquiring a reputation and the rest are what is needed to pursue a career; acquiring the right experience for building a reputation for reliable professional behaviour takes precedence over schooling. Some social skills may be acquired in specialised colleges but investment in social capital on-the-job through work experience is likely to be more effective. The policy implications of the social capital model are therefore that there should be subsidies to



colleges providing hands-on experience with teachers who are professionally active, agreements with professional associations and trade unions to accept young entrants, and so on. It possibly also implies a policy of restricting the number of training places in college so as to raise the 'exclusiveness' and reputation of being accepted on a course. However, apprenticeship schemes and artist-in-residence placements would seem to be more effective in building social capital. As noted, social capital and human capital are likely to be formed side by side and for both there is a private return to the individual that is an incentive to investment.

The implications of the notion of cultural capital for cultural policy are complex and reach to the heart of artistic creativity and the difficulty of applying economic analysis when motivation and outcomes are not easily understood in terms of economic rationality. To say that artists are motivated by the desire to supply cultural value (as I interpret Throsby to do) is simply to 'pass the buck' by changing the language. It does not tell us how that translates into economic terms such as productivity or earnings. We may accept that cultural heritage (or accumulated cultural capital), whether tangible or intangible, is a public good formed by artists in pursuit of their own motives but that does not tell us how it influences successive generations of artists. It could be argued that the greater the stock of cultural heritage, the more difficult it is for artists to be creative, and training that makes students aware of that heritage runs the risk of frightening them off. Conversely, making your mark may best be achieved by shock tactics rejecting that heritage. Therefore, investment in cultural capital by preserving heritages could be either an incentive or a disincentive to individual creativity and furthermore may similarly influence consumers' taste or distaste for new works. It is certainly difficult in some art forms, notably music and opera, to get audiences to attend performances of contemporary creators. The public good nature of cultural capital also implies that it is difficult for individual artists to appropriate the full economic value of their work [Wijnberg (1995)]. This is one of the rationales for copyright law (see below).

The above points raise the question how artistic motivation may be stimulated by government policy. Frey (1997) has developed a general theory of economic motivation based on human psychology that includes the response to pecuniary incentives – extrinsic reward – but extends the maximand to the satisfaction of an inner intrinsic motivation. While other cultural economists have recognised this distinction, Frey's insight is what he calls the 'Crowding Effect', the proposition that inappropriate rewards can displace incentives; for example, monetary payment, an extrinsic reward, may crowd out intrinsic motivation and become a disincentive rather than an incentive for acts which are intrinsically motivated. A better response may be achieved by offering intrinsic rewards to inner-motivated output; an optimal system combines appropriate incentives and rewards. In applying this analysis to the arts, Frey (2000) asks how government support for the arts affects creativity. He distinguishes what he calls 'institutional creativity' from 'personal creativity': institutional creativity is motivated by extrinsic rewards and personal creativity is motivated by intrinsic rewards. Extrinsic rewards are what the market and the state can offer – the market via prices for artists' work and the state through direct financial measures such as subsidy, and indirect measures such as copyright law

[Towse (2001b)]. Personal creativity is clearly more closely related to intrinsic motivation, which in its extreme form is the Romantic concept of the driven genius pursuing art for art's sake at all costs. Its reward is intrinsic, for example, via recognition by one's peers; this is not something the state can offer. But even personal creativity is subjected by Frey to an economic interpretation by applying the all-powerful doctrine of opportunity cost: younger artists can 'afford' to be more creative than older established artists because they have less to lose artistically and financially.

Frey's theory seems to get us somewhat closer to the crucial question about creativity and cultural policy: can we 'create' creativity by investment, private or public? That obviously has important implications for the present-day policy in many countries that emphasise the role of the creative industries. However, whether we espouse the concepts of human, social or cultural capital as our guide to cultural policy towards artists, it is difficult to get away from the role of innate talent. In fact, none of these theories comes to terms with this issue. Even distinguishing intrinsic and extrinsic motivation does not solve the matter. All told, we must accept the fact that the probability of backing the right horse and choosing only highly talented artists to nurture and support is very low. 'Many are called but few chosen' has to be the motto for art colleges and artists' labour markets. At best, they filter out the least able and create conditions in which the best can make their way. Throwing money at arts training by way of investing either in human, social or cultural capital does not really solve the problem of how to create or improve creativity. So how should we attempt to create and improve creativity? The answer is quite simply that we do not know.

## 7. Human capital theory, copyright law and reproducibility

I now turn to an old problem in human capital theory – its inalienability – and argue that the combination of the ability to reproduce works of art ('reproducibility') and copyright law overcome previous objections to the capital analogy, namely that human capital cannot be separated from labour. There is a close relationship between human capital and copyright since both spring from the human mind. Copyright law protects authors and performers by establishing statutory property rights that enable them to control the exploitation of their works, granting them the exclusive right to authorise their use.<sup>20</sup> The economic purpose of copyright is to encourage creativity and the dissemination *inter alia* of works of art.<sup>21</sup>

The evolution of copyright law is inextricably connected to the ability to make mechanical copies that began with the invention of the printing press. The development of

<sup>20</sup> Copyright law in the Anglo Saxon tradition applies both to authors, performers and 'publishers' – companies in the cultural industries, such as producers of sound recordings and film. In the European civil law tradition, authors' rights pertain to human creators and neighbouring rights to the other groups. Here I use the term copyright loosely to refer to both types of rights.

<sup>21</sup> See Landes (2003) and Chapter 7 in this volume.

recording technologies – sound recording, motion picture making, photocopiers, home recording equipment (VCRs, CD burners) and now the Internet – that duplicate a work from a master copy (a performance, a book, a photograph) has vastly extended reproducibility. These inventions have created markets for copyrighted works embedded in CDs, videos, computer games etc., that have been alienated from the person of the artist or creator. Creators mostly have their work marketed by ‘publishers’ (record, film, TV, publishing companies, art galleries and so on – firms in the creative or cultural industries) who act on the assignment or licence of the copyright by the creator. The typical contract is a royalty contract, which may or may not include an advance payment, sharing the sales revenue of the publisher for a fixed percentage, often 10 or 15 percent.<sup>22</sup> Once economic rights have been assigned, however, the artist has little residual control over exploitation (though moral rights may not be alienated). When firms decide to delete works from the catalogue, artists can rarely do anything to stop them. Copyright enables artists to earn from their investment human capital but it does not ensure they do so and how much they earn depends on market forces. It was noted earlier that superstar earnings are disproportionately higher than ‘middle income’ artists. That is also the case with copyright royalty income. Because superstars have greater bargaining power with firms in the cultural industries, they are able to strike a better bargain than ‘ordinary’ artists [Caves (2000)]. ‘Average’ artists’ royalty earnings, by contrast, are typically low [Towse (2001a)].

One other feature of copyright that can be mentioned in this context is ‘works-for-hire’, according to which copyright is conferred on the employer in cases where the employee was directed to do the work; that is typically the situation for Hollywood script-writers and animators, for example. Therefore the control of copyright assets depends on the way the labour market for artists is organised – the less full-time employment there is, the more important copyright is for freelancers.

The exploitation of the author’s work embodied in reproducible form has a double-sided effect: it ‘alienates’ the author’s human capital input from her labour as the work can now reach the market without the necessity of her presence; and through copyright law the publisher acquires a durable asset, the master copy, which he can exploit independently of the author (who may even be dead since the copyright term is life plus 70 years).<sup>23</sup> Thus the joint effect of reproducibility and copyright law has been the creation of capital assets in the hands of the firms in the cultural industries that may be traded and transferred in mergers. The AOL/Time Warner merger, for example, involved the transfer of 1.5 m song titles. This effect is believed by writers on the cultural industries to be responsible for increased merger activity [Bettig (1996)].

Another effect of the combination of copyright law and reproducibility on artists’ labour markets is that an artist can decide to allocate her time to earn a spot price or a

<sup>22</sup> Watt (2000) analyses royalty contracts in detail from the economic point of view.

<sup>23</sup> It is often forgotten that an author’s work is protected for a longer period than the copyright term. If, for example, an author creates a work at the age of 25 and she lives to the age of 75, that work is protected for 120 years.

future return – for example, doing a sound recording in preference to a concert (since the concert pays a fee and the sound recording a royalty). Copyright therefore alters the duration of human capital and artists' supply decisions. In addition to dividing their time between arts and non-arts work [Throsby (1996)] or 'high art' and 'low art' [Cowen and Tabarrok (2000)], artists can optimise a portfolio of copyrights that form part of an inter-temporal decision about present and future earnings. Taking this into account, an artist's earnings at any point in time depend upon wages and fees for the hours of work done in that period *plus* copyright royalty payments (the royalty rate times the number of copyrights the artist holds). It is to be expected that the higher the royalty income, the fewer hours of work artists would do in any given period. A model along these lines could be tested using data from artists' surveys that asked for separate information on fees and wages and on royalties.<sup>24</sup>

The combination of copyright law and reproducibility therefore fundamentally alters two issues in human capital theory, the inalienability of human capital from labour and the period over which the worker can recoup the investment in human capital. As a result of these two features, human capital thus becomes conceptually far closer to physical capital. It is likely that these features are also present in other labour markets, especially those in the 'information' industries. Casualisation of labour, preferences for self-employment, the increased value of information and knowledge, and the increased value of protection through copyright and other intellectual property law are growing in the economy at large. Artists' labour markets may indeed be the forerunner of a more general trend in the evolution of labour markets.

It remains to consider whether a policy of 'strengthening' copyright law or 'increased copyright protection', both much touted by the cultural industries and their pressure groups as assisting artists (as well as themselves). That is a complex question that has been little researched. Strengthening copyright for artists, for example by lengthening its duration, is a two-edged sword: while increasing protection it also takes more work out of the public domain, thus also increasing the cost of creation [Landes and Posner (1989)]. It also benefits companies in the cultural industries more than individual artists since companies have better access to capital markets and a higher time preference rate, though it must be admitted that this is an assertion that remains unproven [Towse (1999)]. In Towse (2001b), however, I argued that copyright may well meet Frey's call for intrinsic motivation for artists by providing symbolic recognition of their status. This is something that could be investigated further.

## 8. Conclusion

In this chapter, I have argued that human capital theory applies only weakly to artists' decisions about investment in schooling and training and about occupational choice.

<sup>24</sup> For a preliminary attempt at such a model, see Towse and Watt (2005).

The same, however, can be said about the sorting model, though the case for it is possibly somewhat stronger. What is lacking in cultural economics is an understanding of talent and creativity, what economic factors motivate artists and how creativity can be encouraged as part of government cultural policy. Bringing social and cultural capital into the equation do not seem to add much in the way of understanding artists' labour markets. The case has been made in this chapter that reproducibility of works of art in combination with copyright law alters the view that human capital cannot be separated from the labour of the artist but that separation, while intellectually interesting, does not help artists to greater rewards in and of itself; indeed, it may well be a cause of increasing skewness of artists' earnings. The effect of copyright earnings on artists' supply decisions is something that must be tested empirically and I have sketched a model that could be used as a basis for further investigation. It could be linked to longitudinal studies of artists' careers, another piece of research that is badly needed.

By concentrating on the role of human capital in artists' labour markets, the focus has been on the supply side. So far there have been no systematic studies of the demand for artists or attempts to analyse skill-bias in the arts. This type of study has increasingly been done in labour economics in order to understand changes in rates of return to human capital over the last 30 years [Acemoglu (2002)]. Cultural economists who have studied artists' labour markets are certainly aware of the increasing demands that are made on artists' skills and competencies and also of the ever growing skills of certain kinds of artists – for example, singers and instrumentalists now routinely perform music that was considered unplayable 75 years ago and do so with little rehearsal. That is surely a sign of increased productivity.<sup>25</sup> This is another important topic for future research in artists' labour markets.

It is essential that cultural economics in general and the study of artists' labour markets in particular continue to apply standard economic ideas to the arts and to test them. The arts form part of the economy; they use resources and produce consumer goods. Artists are workers – they may be more like ministers of religion, inventors or creative engineers than accountants or travel agents but as a starting point it is right to look for similarities between artists and other workers using labour economics and human capital theory. That has been the approach of those who initiated research on artists' labour markets; however, experience so far suggests that the human capital model is not the way forward. The overall conclusion, then, is that there has been empirical progress in the analysis of artists' labour markets but there is much more to be done.

In closing, it is interesting to note that there are some strong parallels between aspects of artists' labour markets and those of sportspeople. In both fields the role of talent, innate ability and ability acquired early in life exert a strong influence on earnings and career success, and superstardom is probably even more marked in sport than in the

<sup>25</sup> One of the long run effects of the espousal of Baumol's Cost Disease in the performing arts has been the assumption that no technical progress is possible in the arts or on the part of artists. That is, however, a misunderstanding. See Cowen (1998) for a counter view; see also Chapter 11 by Baumol and Chapter 15 by Brooks in this volume.

arts. Seaman (2003) concludes that there is much to be gained by cultural economists from joint research between the two fields. Maybe artists are after all not so completely different!

## References

- Abbing, H. (2002). *Why Are Artists Poor? The Exceptional Economy of the Arts*. Amsterdam University Press, Amsterdam.
- Acemoglu, D. (2002). "Technical change, inequality and the labour market". *Journal of Economic Literature* XL (1), 7–72.
- Arrow, K. (2000). "Observations on social capital". In: Dasgupta, P., Serageldin, I. (Eds.), *Social Capital: A Multifaceted Perspective*. World Bank, Washington, DC, pp. 3–5.
- Becker, G. (1964). *Human Capital*. Columbia University Press, New York.
- Benhamou, F. (2000). "The opposition of two models of labour market adjustment. The case of the audiovisual and performing arts in France and in the United Kingdom". *Journal of Cultural Economics* 24, 301–319.
- Bettig, R. (1996). *Copyrighting Culture*. Westview Press, Boulder.
- Blaug, M. (1970). *An Introduction to the Economics of Education*. Allen Lane Penguin Press, London.
- Blaug, M. (1976). "The empirical status of human capital theory: A slightly jaundiced survey". *Journal of Economic Literature* 14 (3), 827–856.
- Blaug, M. (1985). "Where are we now in the economics of education?" *Economics of Education Review* 4 (1), 17–24.
- Bowles, S., Gintis, H. (2002). "Social capital and community governance". *Economic Journal* 112, F419–F436.
- Castañer, X., Campos, L. (2002). "The determinants of artistic innovation: Bringing in the role of organisations". *Journal of Cultural Economics* 26 (1), 29–52.
- Caves, R. (2000). *Creative Industries: Contracts between Art and Commerce*. Harvard University Press, Cambridge, MA.
- Cowen, T. (1998). In *Praise of Commercial Culture*. Harvard University Press, Cambridge, MA.
- Cowen, T., Tabarok, A. (2000). "An economic theory of avant-garde and popular art, or high and low culture". *Southern Economic Journal* 67 (2), 232–253.
- Durlauf, S. (2002). "On the empirics of social capital". *Economic Journal* 112, F459–F479.
- Filer, R. (1986). "The starving artist: Myth or reality? – Earnings of an artist in the United States". *Journal of Political Economy* 94 (1), 56–75.
- Filer, R. (1987). "The price of failure: Earnings of former artists". In: Shaw, D., Hendon, W., Waits, C. (Eds.), *Market for the Arts*. Akron University Press, Akron.
- Filer, R. (1990). "The arts and academe: The effect of education on earnings of artists". *Journal of Cultural Economics* 14 (1), 15–38.
- Fine, B. (2001). *Social Capital versus Social Theory*. Routledge, London and New York.
- Frey, B. (1997). *Not Just for the Money*. Edward Elgar, Cheltenham.
- Frey, B. (2000). *Art and Economics*. Springer, Heidelberg.
- Glaeser, E., Laibson, D., Sacerdote, B. (2002). "An economic approach to social capital". *Economic Journal* 112, F437–F458.
- Gurgand, M., Menger, P.-M. (1996). "Work and compensated unemployment in the performing arts. Exogenous and endogenous uncertainty in artistic labour markets". In: Ginsburgh, V., Menger, P.-M. (Eds.), *Economics of the Arts*. Elsevier, Amsterdam, pp. 347–381.
- Jeffri, J., Throsby, D. (1994). "Professionalism and the visual arts". *European Journal of Cultural Policy* 1 (1), 99–108.
- Landes, W. (2003). "Copyright". In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham, pp. 132–142.

- Landes, W., Posner, R. (1989). "An economic analysis of copyright law". *Journal of Legal Studies* 18, 325–366.
- MacDonald, G. (1988). "The economics of rising stars". *American Economic Review* 78, 155–166.
- McNertney, M., Waits, C. (1997). "The incomes of cultural providers: A review of current research". In: Towse, R. (Ed.), *Cultural Economics: The Arts, Heritage and the Media Industries*. Edward Elgar, Cheltenham, pp. 275–283.
- Mincer, J. (1958). "Investment in human capital and personal income distribution". *Journal of Political Economy* 66, 281–302.
- Putnam, R. (2000). *Bowling Alone*. Simon and Schuster, New York.
- Rengers, M., Velthuis, O. (2002). "Determinants of prices for contemporary art in Dutch galleries". *Journal of Cultural Economics* 26, 1–26.
- Rosen, S. (1981). "The economics of superstars". *American Economic Review* 71, 845–858.
- Rosen, S. (1987). "Human capital". In: Eatwell, J., Milgate, M., Norman, P. (Eds.), *The New Palgrave: A Dictionary of Economics*. Macmillan, London and Basingstoke, pp. 681–689.
- Rosen, S. (1992). "Distinguished fellow: Mincering labor economics". *Journal of Economic Perspectives* 6 (2), 157–170.
- Schultz, T. (1961). "Investment in human capital". *American Economic Review* 51, 1–17.
- Schumpeter, J. (1942). *Capitalism, Socialism and Democracy*. Harper, New York.
- Seaman, B. (2003). "Cultural and sports economics: Conceptual twins?" *Journal of Cultural Economics* 27 (2), 81–126.
- Smith, A. ([1776] 1976). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Campbell, R.H., Skinner, A.S., Todd, W.B. (Eds.). Clarendon Press, Oxford.
- Sobel, J. (2002). "Can We Trust Social Capital?" *Journal of Economic Literature* XL (1), 139–154.
- Solow, R. (2000). "Notes on social capital and economic performance". In: Dasgupta, P., Serageldin, I. (Eds.), *Social Capital: A Multifaceted Perspective*. World Bank, Washington, DC, pp. 6–10.
- Throsby, D. (1992). "Artists as workers". In: Towse, R., Khakee, A. (Eds.), *Cultural Economics*. Springer, Berlin. Reprinted in Towse (1997, pp. 261–268).
- Throsby, D. (1994). "A work preference model of artists' behaviour". In: Peacock, A., Rizzo, I. (Eds.), *Cultural Economics and Cultural Policies*. Kluwer Academic, Dordrecht, pp. 69–80.
- Throsby, D. (1996). "Disaggregated earnings functions for artists". In: Ginsburgh, V., Menger, P.-M. (Eds.), *Economics of the Arts: Selected Essays*. Kluwer Academic, Dordrecht, pp. 331–346.
- Throsby, D. (2001). *Economics and Culture*. Cambridge University Press, Cambridge.
- Towse, R. (1992). "Earnings of singers: An economic analysis". In: Towse, R., Khakee, A. (Eds.), *Cultural Economics*. Springer, Berlin. Reprinted in Towse (1997, pp. 218–226).
- Towse, R. (1993). *Singers in the Marketplace: The Economics of the Singing Profession*. Clarendon Press, Oxford.
- Towse, R. (1996). "Economics of artists' training". In: Ginsburgh, V., Menger, P.-M. (Eds.), *Economics of the Arts*. North-Holland, Amsterdam, pp. 303–329.
- Towse, R. (1997). *Cultural Economics: The Arts, Heritage and the Media Industries*. Edward Elgar, Cheltenham.
- Towse, R. (1999). "Copyright and economic incentives: An application to performers' rights in the music industry". *Kyklos* 52 (3), 369–390.
- Towse, R. (2001a). *Creativity Incentive and Reward*. Edward Elgar, Cheltenham.
- Towse, R. (2001b). "Partly for the money: Rewards and incentives to artists". *Kyklos* 54 (2/3), 473–490.
- Towse, R., Watt, R. (2005). "A model of author time allocation". Paper presented at SERCI conference 7–8 July Montreal, [www.serci.org](http://www.serci.org).
- Wassall, G., Alper, N. (1985). "Occupational characteristics of artists: A statistical analysis". *Journal of Cultural Economics* 9 (1), 13–34.
- Watt, R. (2000). *Copyright and Economic Theory. Friends or Foes?* Edward Elgar, Cheltenham.
- Weiss, A. (1995). "Human capital vs. signalling explanations of wages". *Journal of Economic Perspectives* 9 (4), 133–154.

- Wijnberg, N. (1995). "Selection processes and appropriability in art, science and technology". *Journal of Cultural Economics* 19 (3), 221–235.
- Wijnberg, N. (2003). "Awards". In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham, pp. 81–84.
- Willis, R. (1986). "Wage determinants: A survey and reinterpretation of human capital earnings functions". In: Ashenfelter, O., Layard, R. (Eds.), *Handbook of Labor Economics*. Elsevier, Amsterdam, pp. 525–602.
- Withers, G. (1985). "Artists' subsidy of the arts". *Australian Economic Papers* 24, 290–295. Reprinted in Towse (1997, pp. 269–274).



## STARDOM AND TALENT\*

MOSHE ADLER

*Columbia University, New York, NY, USA*

### Contents

Abstract	896
Keywords	896
1. Theories of superstardom	897
1.1. Why does superstardom occur?	897
1.2. The emergence of superstars	898
1.3. The efficiency of stardom	899
1.3.1. The effect on consumers	899
1.3.2. Publicity rights	900
2. Superstars and other artists	901
2.1. Are there too many artists?	901
2.2. Superstardom and the psychic income of artists	902
2.3. Government policy and the optimal number of artists	902
3. Empirical testing of superstardom theories	903
3.1. Testing superstardom in the arts	903
3.2. Testing superstardom in sports	904
4. Conclusion	905
References	905

\* The author would like to thank David Throsby and Victor Ginsburgh for their invaluable suggestions that greatly improved this chapter. Tony Bryant and Ellen Adler also provided important comments. In addition, the author would like to thank the participants at the conference of the authors at Princeton in the Fall of 2004.

**Abstract**

The Economics of Superstars sets out to explain the relationship between talent and success in the arts, but there is no agreement about what this relationship is. But whatever its other features may be, superstardom means that market output is concentrated on just a few artists. Concentration always raises the question of efficiency. Superstardom may be inefficient not only because it raises prices for consumers but also because it deprives other artists of the opportunity to practice art. Artists who do not practice art lose psychic income. Because psychic income cannot be transferred from one person to another, the loss of this income may be inefficient. This chapter reviews theories of superstardom and theories about the emergence of stars. The efficiency of superstardom is discussed in terms of effects on consumers and the use of publicity rights by the star. The chapter goes on to deal with the loss of opportunities to practice art that are caused by superstardom and suggests ways to alleviate the problem. Finally the empirical literature that tests the different theories of superstardom is reviewed.

**Keywords**

superstars, stardom, talent

*JEL classification:* Z1, Z10

## 1. Theories of superstardom

### 1.1. *Why does superstardom occur?*

Superstars in the arts, sport etc. are individuals who attain considerable prominence and success in their field and whose earnings as a result are significantly greater than the earnings of their competitors. According to Rosen (1981), superstardom in the arts is due to two factors: a hierarchy of talent, and the perfect or nearly perfect reproducibility of art. A good illustration of Rosen's argument comes from the music or theater markets. In the past a singer or an actor, no matter how good she or he was, could serve only a limited number of people. Hence singers and actors of all levels of talent could find audiences. But now that music can be perfectly reproduced on recordings, and theater has been supplemented or displaced by movies and videos, every consumer can inexpensively consume the performances of the best artist.

But if the best artist is significantly better than her competitors, then in fact "each consumer consuming the best" is a special case. Consider singers producing and selling CDs. The best artist in these circumstances is a monopolist, and whether profit maximization calls for a low price and selling to all consumers, or a high price and selling only to a few, depends on the elasticity of the demand for her product. Only if demand is highly elastic will it be profit-maximizing to serve the whole market. Particularly interesting in Rosen's model is the case in which there are several artists with the same top-level talent. Because the average cost of producing CDs is decreasing (the cost of producing the music recorded on a CD is a fixed cost), there will still be only one star even in this case. But since competition among these top-level artists to be the star would be fierce, the star would be able to charge only the average cost of production in selling her CDs. The star would, therefore, be poor.<sup>1</sup>

Hence in Rosen's model there are two extreme possibilities. If there is an artist who is significantly more talented than the rest, this artist, unless the demand is nearly perfectly elastic, sets a high price for her art and sells it to only a fraction of consumers. If there are several artists of equal talent, one of them serves the whole market, but she is poor. Thus according to Rosen's model, if a star is extremely popular and extremely rich, her talent must be greater than the rest by just the right amount.

Adler (1985) argues that the existence of superstars is not due to differences in talent. He suggests that there are in fact many artists who possess stardom-quality talent; what produces superstars is the need on the part of consumers to consume the same art that others do. This need arises from the fact that the consumption of a piece of art is not a momentary experience but a dynamic process in which "the more you know, the more you enjoy". Consumers build "consumption capital" in art, and the larger the capital

<sup>1</sup> Rosen (1981, p. 852).

the greater is the enjoyment from each encounter with the art and the artist [Stigler and Becker (1977)].<sup>2</sup>

The acquisition of this knowledge occurs in three ways. It can result from exposure to the art itself, from discussions about the art with friends or acquaintances, or from reading about the art in newspapers and magazines. When the artist is popular, it is easier to find discussants who are familiar with her or to find media coverage about her. This is why consumers prefer to consume what others also consume.

Because the number of artists who can be popular at any one time is limited, not all talented artists can be successful. However, the frustration of an unsuccessful artist does not end with not having an audience. She must also suffer consumers' judgment that she deserves her fate. It is easy to conclude that an artist who is not as popular "is not as good". Thus, the hierarchy of success manufactures a hierarchy of pseudo-talent in its own image, tending to reinforce the spurious perception of a talent differential.

### *1.2. The emergence of superstars*

Rosen's model relies on known differences in talent and it therefore includes a straightforward process by which superstars emerge. MacDonald (1988), considering only the performing arts, defines talent differently, and describes a dynamic process through which stars emerge. He argues that every performing artist is capable of producing either a good or a bad performance. The difference in talent between artists is seen not in the quality of their good or bad performances, but in the probability that a particular performance will be good. This probability is the same throughout the artist's career. But from the vantage point of audiences, the probability of a good performance is lower for a new performer than for a known performer. This is because many artists try their luck at performing and those who perform poorly drop out. Those who perform well, on the other hand, stay, and their probability of performing well in the future is higher. Therefore artists with a good track record can command higher ticket prices and entertain larger audiences. In these circumstances, artists of equal talent do equally well.

Adler (1985) also describes a dynamic process for the emergence of a star, but in this process the star emerges from among several artists who are all equally talented. In this model the emergence of the star arises from a chance event: consumers select an artist at random when they add a new artist to their consumption basket, and it is simply by pure chance that one of these artists ends up with more patrons than the rest. This initial advantage makes the lucky artist the most popular, and since consumers prefer popular artists, other consumers will switch to her as well. An initial advantage can thus snowball into superstardom.

Pure luck is just one possible mechanism by which consumers initially choose a particular artist. Indeed artists themselves do not usually entrust this choice to chance.

<sup>2</sup> Harvey Leibenstein (1950) described a "bandwagon effect" that causes consumers to consume the same goods that other consumers do. Leibenstein attributed the effect to the need of consumers to conform or to be stylish.

An author, for instance, may purchase copies of her own book in order to try to push it onto the bestseller list. A musician may pay “payola” (even though it is illegal) to convince a disk jockey to play her music on the radio. And, most importantly, artists use publicity such as appearances on talk shows and coverage in tabloids and magazines to signal their popularity. None of this detracts from the possibility, however, that the allocation of publicity resources to artists is independent of differences in artistic talent.<sup>3</sup> We return to this issue in a following section.

### 1.3. *The efficiency of stardom*

#### 1.3.1. *The effect on consumers*

Are the prices that stars charge too high? Economists do not have a yardstick for determining whether prices are too high or too low. But they distinguish between prices that are determined in markets with free entry and markets with barriers to entry. We can analyze the issue of market power via the example of the market for music CDs.

Entry into a market is said to be free if an artist who is as talented as the star can offer her CDs for a slightly lower price and capture the market. According to Rosen, this is indeed the case, and the art market is therefore efficient. However in Adler’s model, consumers prefer the most popular artist and therefore even an artist who is as talented as the star cannot entice audiences away from the star, not even by offering a lower price. Hence in this model entry is not free, and there are theoretical grounds for government control of prices.

Borghans and Groot (1998) explain why the prices that the stars charge are too high. They first reject Adler’s theory because it is based on the assumption that the stars may not be more talented than other artists. They argue that if there are several artists who have the potential of being a superstar, each would invest resources in enhancing her chances to be the one, and as a result the total income of the superstar would decrease. While this is of course true, Borghans and Groot acknowledge that the maximum that an artist would be willing to invest in promotion is the expected value of becoming a superstar. If there are many artists with the same superstar-potential, the probability that any particular one will become the winner will be low, and the expected value of becoming a superstar will therefore be low as well. Thus each artist’s investment in promotion will be small and the lucky superstar’s income, even after accounting for the promotion expenses, will be large.<sup>4</sup>

Borghans and Groot begin their analysis of Rosen’s theory by proving that talent differences and the reproducibility of art are not sufficient on their own to produce superstar earnings. Their proof is via an example of  $n$  consumers and  $n$  artists, but the case of two consumers and two artists is just as illustrative.

<sup>3</sup> This raises the possibility that the star could be an artist of a lesser talent; see Adler (1985), Bonus and Ronte (1997).

<sup>4</sup> Borghans and Groot (1998, p. 557).

Assume that art cannot be replicated, and that the cost of producing art is zero. Assume also that artist *A* is less talented than artist *B*, that the consumers *a* and *b* are identical, and that the consumers' reservation prices for the art of *A* and *B* are \$1 and \$2, respectively. Each consumer consumes just one unit of art. Finally assume that artist *A* charges \$1 for his art, which he sells to consumer *a*, and artist *B* charges \$2 for her art, which she sells to consumer *b*; with these prices, each consumer is equally satisfied. In other words, the lower price that artist *A* charges constitutes full compensation for his lower talent. Now suppose that all of a sudden art can be replicated costlessly. Intuition may lead one to believe that artist *B* would be able to increase her income by selling to both consumers, but Borghans and Groot show that this is actually not the case. If *B* were to charge more than \$1.00 for her art say \$1.50, then artist *A* will charge a price that is at least \$1.00 lower than *B*'s price, say \$0.25, and capture the whole market. Hence, competition would force the superstar to charge only \$1.00 and as a result the superstar would earn exactly the same income that she would have earned had replication not been possible.

Borghans and Groot conclude that something else must explain the high incomes of superstars. It must be that consumers stay with the superstar even when there is another artist who charges a price that is so low that it is sufficient to compensate them for his smaller talent. According to Borghans and Groot the reason for this loyalty is that consumers prefer to "watch the performance of someone known to be 'the best'".<sup>5</sup> The problem with this argument is that it is contradictory. If the lower price already constitutes full compensation for the smaller talent, why would the consumer still prefer to watch "the best"?<sup>6</sup> Substitute "most popular" for "the best" in their statement, however, and the argument, now identical to Adler's, is consistent. When replication is not possible a price difference compensates consumers for the talent difference, and each artist is equally popular (or equally unpopular). When replication becomes possible the same price difference continues to compensate consumers for the talent difference, but it does not compensate them for the difference in popularity, which exists when all consumers patronize the same artist. Of course, if consumers value popularity no difference in talent is required to produce it.

### 1.3.2. *Publicity rights*

The most novel challenge to the efficiency of the stardom system comes from the literature about publicity rights. The seminal article in this literature is by Madow (1993) who argues that the existence of these rights is inefficient because they restrict the public's use of symbols. Madow cites as an example a greeting card that carries a photograph of John Wayne wearing lipstick. The card's message is probably that masculinity and

<sup>5</sup> Borghans and Groot (1998, p. 561).

<sup>6</sup> There must be a price differential that constitutes full compensation for the talent differential, because otherwise there would be no equilibrium in a world in which replication is not possible.

homosexuality are not contradictory. But whatever the exact message is, it is clear that John Wayne was not its creator. Had the greeting card company had to pay for the use of the photograph it would have paid for a symbol that was in large part its own creation. Furthermore, Wayne might well have refused to sell his image for this use. In fact, during hearings on a bill to create publicity rights in New York State in 1989, Wayne's children cited this "abuse" of their father's image as a justification for the bill. Thus publicity rights amount to censorship.

While the distance between a star's intentions and the actual symbol that she becomes may be larger in this particular case than it is in most other cases, Madow argues that a large distance is nevertheless typical. Even the U.S. military does not treat Wayne as a positive hero, according to Madow, because when it warns soldiers against taking foolish risks, it admonishes them not to "John Wayne it". Nevertheless, do stars deserve to have publicity rights in those cases in which they are indeed the creators of the symbols that they become? If the stars had not created these symbols, consumers would have created them themselves, Madow argues, perhaps through the snowballing process discussed above.

As an example for this process Madow cites the emergence of Albert Einstein as the symbol of genius. According to Missner (1985, p. 268), Niels Bohr and Werner Heisenberg made equally great contributions to science. It was serendipity that favored Einstein. In 1921 he came to New York as a member of a Zionist delegation that was headed by Chaim Weizman, the head of the Zionist movement at the time. Thousands of New York Jews went to the port to greet the Zionist delegation, but the newspapers reported that the enthusiasm was not for Weizman and Zionism but for Einstein and his theory of relativity. This made Einstein a subject of newspaper interviews, and propelled his ascendance as the symbol of genius.

The moral justification for the existence of publicity rights notwithstanding, the question is whether stars can charge prices that are inefficiently high for these rights. Can a star who is not a symbol offer to serve as a symbol at a lower price, and enter the market? Even if a star could turn herself into a symbol at will, she could not do so overnight, since she would need the active participation of the public. Hence there are barriers to entry into the market for symbols, and the prices charged for the use of publicity rights are probably inefficiently high.

## 2. Superstars and other artists

### 2.1. *Are there too many artists?*

According to Frank and Cook (1995), the large incomes that superstars earn cause too many artists to attempt to be the winners. While seeking stardom these "surplus" artists forgo income from non-artistic jobs and some may even neglect the normal education that would have permitted them to earn high incomes doing regular jobs. Hence false

dreams of success may cause poverty. Frank and Cook call on the government to limit the remuneration of artists in order to make superstardom less attractive.

In his review of Frank and Cook's book, Rosen (1996) was skeptical about the empirical relevance of their argument, arguing that artists who seek success learn what their personal odds are quickly; when these odds are low they quit the field rapidly. Rosen also suggested that the effect of false dreams on the rate of poverty is negligible; in other words, the poor are not poor because they dream of being rich.

### *2.2. Superstardom and the psychic income of artists*

It appears that artists derive psychic income from practicing art, because their monetary incomes from producing art are consistently lower than in equivalent alternative occupations [Jeffri (1991); Throsby (1992); Wassall and Alper (1992); Menger and Gurgand (1996)]. When artists do not practice art this psychic income is lost. In regular markets when a business fails, its loss is another business's gain. But when an artist cannot practice art because consumers flock to superstars, her loss of psychic income is not transferred either to the superstar or to her audiences.<sup>7</sup> It is simply lost. This loss would only be efficient if there were other gains from superstardom that exceeded this loss. Whether there are depends in part on what gives rise to superstardom in the first place. If artists are displaced because they are less talented than the superstar, as in Rosen's model, then their displacement would be efficient. Why encourage anyone to produce an inferior product? But if the displaced artists are just as talented as the stars, and the only reason they are displaced is that consumers prefer popular artists, then the loss of benefits to the artists no longer practicing art must be weighed against the benefits to consumers from increasing the concentration in the art market.

There is no doubt that a certain degree of concentration in the arts is desirable. If every consumer patronized a different artist there would be no common culture. But is more concentration always better than less concentration? As we have noted, increased concentration results in an uncompensated loss of psychic income as unsuccessful artists stop practicing their art. But a government policy measure to increase the number of artists as a means of reducing concentration is likely to be costly. These costs need to be balanced, suggesting that there will be an optimal level of concentration, i.e. an optimal number of artists.

### *2.3. Government policy and the optimal number of artists*

If concentration is regarded as excessive, it could be reduced by means of a tax. What sort of tax might this be? The loss of psychic income amongst artists when superstars emerge arises because of an externality; a consumer who flocks to a star instead of

<sup>7</sup> This assumes that artistic psychic income, while perhaps requiring a minimum audience in order to exist, does not increase with audience size.



patronizing a lesser-known artist disregards the loss of psychic income that she inflicts on that artist. This externality can be internalized by a Pigouvian tax, but implementing the tax poses a unique problem. While a consumer should pay the tax if she flocks to a star who is already overly popular, she should not pay the tax if she were one of the star's first patrons. How could the "first" patrons be distinguished from the "rest"?

A way to implement a discriminatory tax is as follows. Let the total number of consumers be  $C$  and let the optimal number of artists be  $N$ . Assume that each consumer buys one CD, on which they would have to pay a tax. At the same time, however, the government would issue each person who claims to be an artist a book containing  $C/N$  tax-rebate certificates that bear her name. The artist would distribute these certificates free of charge to consumers who would then present their certificates to the government for payment. The number of artists who receive the rebate books may be large, far exceeding  $N$ , but the number of artists who end up having patrons would nevertheless be the optimal number, because consumers will prefer the most popular artist who is not yet "full". Thus a successful artist would sell  $C/N$  CDs. The level of the tax will be set sufficiently high to discourage consumers from paying the tax and flocking.

It should be emphasized that under this system it is consumers, not the government, who determine who the successful artists are, and that no consumer actually pays the tax, because each gets a rebate. Nevertheless, the artists are still being subsidized, because consumers end up spending more on collecting information about "their" artists than they would under the unencumbered superstardom system. Of course, like with any other tax policy the cost of implementing this policy would have to be taken into account in any assessment of its desirability.

### 3. Empirical testing of superstardom theories

Is stardom the reward for superior talent or does stardom arise because of consumers' need for a common culture? This section reviews the studies and makes a suggestion for an additional test.

#### 3.1. Testing superstardom in the arts

Whereas talent in general is not measurable, the harmonic quality of a singer's voice is. Hamlen (1991, 1994) measured the relationship between it and record sales and discovered that record sales do increase with the quality of the voice, but that the differences in talent far exceed the differences in sales. Hamlen interprets this result as being inconsistent with Rosen's "reward for talent" explanation, but Schulze (2003) argues that "it is by no means clear that the harmonic content of voice is the relevant measure for artistic quality for singers of non-classical music (rock, folk and so on)". There is no

doubt that the quality of the songs, not just the quality of the voice, must be measured as well, but this may not be possible.<sup>8</sup>

Chung and Cox (1994) take another approach to testing the two theories. They show that the distribution of success among artists follows a snowballing process (a Yule distribution). In that process the probability that a consumer would buy a particular CD increases with the number of previous sales of that CD. There always remains a small probability that a consumer will choose a new CD that no other consumer has yet bought. When this happens, other consumers may follow suit and an initial small advantage may snowball into success. Chung and Cox believe that this lends support to Adler's theory over Rosen's, but Schulze (2003) argues that the process is also consistent with consumers' choices that are based on talent.

Regardless of whether or not one agrees with Chung and Cox, their study shows clearly how important it is for artists to have an initial advantage. But how do artists acquire this advantage? Do they employ techniques that emphasize their talent, or do they choose entirely unrelated means? The proverbial "casting couch" comes to mind, but the singer Britney Spears showed that it is also possible to capture headlines by claiming to be chaste. Madonna captured headlines at the beginning of her career when her husband "protected" her by routinely beating up photographers. What is required is a systematic study of a sizeable sample of stars' (unauthorized) biographies in order to determine how important non-talent factors were in the early stages of their careers. Ideally these biographies would be compared to the biographies of artists who have not been successful. However, because there are few biographies of unsuccessful artists, the ideal may be impossible to achieve.

One example of how an initial advantage is generated is provided by Ginsburgh and van Ours (2003) in an article about the Queen Elizabeth Piano Competition. Pianists who achieve high success in the competition are rewarded by subsequent success. While this may appear to be as it should be, Ginsburgh and van Ours show that the order in which the pianists perform in the competition – which is assigned randomly – affects the results of the competition. Since success in the competition is random, why does it influence subsequent success in the market place? The answer may be that success in the competition serves not as an indicator of the artist with the most talent (consumers may even believe that all those who make it into the finals are equally talented) but as a focal point for consumers who wish to listen to artists that others listen to.

### 3.2. *Testing superstardom in sports*

Theories of superstardom have also been tested empirically in sports. Talent differences are easier to measure in sport than in art, and Seaman (2003) has asked whether it is possible to learn from the relationship between talent and income in the two areas. He found

<sup>8</sup> Hamlen also finds that success in the singles market leads to success in the regular album market, but this finding is consistent with both theories of stardom and is therefore not informative. A consumer may buy the full album either because she liked the single or because the success of the single is a signal that the artist is popular.

only two studies about talent and income in sports. [Lucifora and Simmons \(2003\)](#), in a study of Italian soccer, showed that the distribution of soccer players' incomes is more skewed than the distribution of talent, attributing this finding to audiences' preference for watching star players over watching equally talented but less well-known players. This may be a confirmation of Adler's theory that consumers prefer to watch players or artists with whom they and other spectators are already familiar. The preference for familiar athletes may also explain the finding of [Blass \(1992\)](#) that the income of baseball batters increases with experience rather than with productivity.

It is important, however, to note some differences between art and sports that diminish the value of sports economics for the understanding of art economics. Sports are competitive and in a competition every participant must accomplish the same task. For example, in a country where sumo wrestling does not exist an athlete who is a great sumo wrestler and only a mediocre football player will have to play football. In art, however, there are no such limitations and therefore an artist has more opportunities to display his or her idiosyncratic talent. This is why in art, unlike in sports, there are no measurable standards.

#### 4. Conclusion

As the debate about the international enforcement of copyrights makes clear, globalization intensifies the phenomenon of superstardom. A global culture, with a global set of superstars, is replacing local cultures with local stars, and it is therefore important to know what this means for consumers, artists and art. Economists have started to examine these questions only very recently. As this chapter has shown, there are those who believe that the global superstars will simply be the best artists on the planet. From their vantage point, there is no reason for concern. However, this chapter has also shown that a single global culture could possibly destroy local cultures not because it is better but simply because it is global. If the emergence of a global culture cannot be stopped, and if this culture does not have to be superior to be triumphant, the question becomes how to democratize the process that builds this culture. "The Economics of Superstars" is thus rife with open questions.

#### References

- Adler, M. (1985). "Stardom and talent". *American Economic Review* 75, 208–212.
- Bonus, H., Ronte, D. (1997). "Credibility and economic value in the visual arts". *Journal of Cultural Economics* 21, 103–118.
- Blass, A.A. (1992). "Does the baseball labor market contradict the human capital model of investment?" *Review of Economics and Statistics* 7, 261–268.
- Borghans, L., Groot, L. (1998). "Superstardom and monopolistic power: Why media stars earn more than their marginal contribution to welfare". *Journal of Institutional and Theoretical Economics* 54, 546–571.
- Chung, K., Cox, R. (1994). "A stochastic model of superstardom: An application of the Yule distribution". *Review of Economics and Statistics* 76, 771–775.

- Frank, R.H., Cook, P.J. (1995). *The Winner-Take-All Society*. The Free Press, New York.
- Ginsburgh, V., van Ours, J. (2003). "Expert opinion and compensation: Evidence from a musical competition". *American Economic Review* 93, 289–296.
- Hamlen, W. (1991). "Superstardom in popular music: Empirical evidence". *Review of Economics and Statistics* 73, 729–733.
- Hamlen, W. (1994). "Variety and superstardom in popular music". *Economic Inquiry* 32, 395–406.
- Jeffri, J. (1991). *The Artists Training and Career Project: Painters*. Columbia University, Research Center for Arts and Culture, New York.
- Leibenstein, H. (1950). "Bandwagon, snob, and Veblen effects in the theory of consumers' demand". *Quarterly Journal of Economics* LXIV, 183–207.
- Lucifora, C., Simmons, R. (2003). "Superstar effects in sport: Evidence from Italian soccer". *Journal of Sports Economics* 4, 35–55.
- MacDonald, G. (1988). "The economics of rising stars". *American Economic Review* 78, 155–166.
- Madow, M. (1993). "Private ownership of public image: Popular culture and publicity rights". *California Law Review* 81 (1), 125–152.
- Menger, P.-M., Gurgand, M. (1996). "Work and compensated unemployment in the performing arts: Exogenous and endogenous uncertainty in artistic labor markets". In: Ginsburgh, V.A., Menger, P.-M. (Eds.), *Economics of the Arts: Selected Essays*. North-Holland, Amsterdam, pp. 347–381.
- Missner, M. (1985). "Why Einstein became famous in America". *Social Studies of Science* 15 (2), 267–291.
- Rosen, S. (1981). "The economics of superstars". *American Economic Review* 71, 845–858.
- Rosen, S. (1996). "Book review: The winner-take-all society". *Journal of Economic Literature* 34, 133–136.
- Schulze, G.G. (2003). "Superstars". In: Towse, R. (Ed.), *Handbook of Cultural Economics*. Edward Elgar, Cheltenham, pp. 431–436.
- Stigler, G., Becker, G. (1977). "De gustibus non est disputandum". *American Economic Review* 67, 76–90.
- Throsby, D. (1992). "Artists as workers". In: Towse, R., Khakee, A. (Eds.), *Cultural Economics*. Springer-Verlag, Heidelberg, pp. 201–208.
- Seaman, B.A. (2003). "Cultural and sport economics: Conceptual twins?" *Journal of Cultural Economics* 27, 81–126.
- Wassall, G.H., Alper, N.O. (1992). "Towards a unified theory of the determinants of the earnings of artists". In: Towse, R., Khakee, A. (Eds.), *Cultural Economics*. Springer-Verlag, Heidelberg, pp. 187–200.

## ART AUCTIONS

ORLEY ASHENFELTER

*Princeton University, Princeton, NJ, USA*

KATHRYN GRADDY

*University of Oxford, Oxford, UK*

### Contents

Abstract	910
Keywords	910
1. Introduction	911
1.1. The mechanics of art auctions	912
2. The price of art	914
2.1. Art price indices	914
2.1.1. Art as an investment	921
2.1.2. The masterpiece effect	923
2.2. The life cycle of an artist	925
2.3. The law of one price	926
3. The auction mechanism and the price of art	928
3.1. Competition (or collusion) between auction houses	929
3.2. Role of estimates and experts	931
3.3. Is there evidence that paintings are “burned?”	933
3.4. Sales rates and reserve prices	935
3.5. Why secret reserve prices?	937
3.6. The declining price anomaly	938
4. Conclusion	941
Appendix A	941
References	942

**Abstract**

This paper contains a new review of the research of the last decade that has been designed to shed light on how the art auction system works, what it indicates about price formation, and how well it performs. We begin with a short description of the mechanics of the auction system and then organize the remainder of our discussion around two major topics. The first topic concerns how researchers have used auction prices. We begin by discussing the returns to holding art and whether certain classes of art make a better investment. We then discuss studies that have used auction prices to determine the importance of individual works of art and individual artists. We conclude this topic by discussing whether different auction houses achieve systematically different prices. The second topic focuses on studies that examine the influence of the auction mechanism on prices. We begin with a discussion of the Christie's and Sotheby's price fixing case and the role of the competitive behavior of auction houses in the determination of prices. We move on to discuss the role of experts and price estimates in auctions. We then look at whether items that fail at auction subsequently return less than items that have not failed and furthermore discuss sale rates and reserve prices. We conclude this topic by reviewing the extensive literature that has documented and tried to explain why the prices of identical objects are more likely to decline than to increase when multiple units are sold. Subsequent empirical research has tended to document declining prices even when the objects are imperfect substitutes.

**Keywords**

auctions, art, price anomalies, asset returns, price fixing

*JEL classification:* D44, G11, L12, Z11

## 1. Introduction

The value of most important works of art is established by public auction, either directly, by an actual sale, or indirectly, by reference to other sales. How the auction system works is thus a critical determinant of how the public's preferences are translated into the evaluation of high-quality artistic work. The auction system is an important factor in the determination of the incentives for artistic work, and the efficiency of the auction system is a determinant of the cost of creating and distributing works of art. This paper contains a review of the burgeoning research of the last decade that has been designed to shed light on how the art auction system actually works, what it indicates about price formation, and how well it performs. We begin the paper with a description of the mechanics of the auction system. We organize the remainder of our discussion in sections around two major topics.

In section two, we discuss how researchers have used auction prices. We begin by discussing the returns to holding art and whether certain classes of art make a better investment. We find that in recent years returns on art assets appear to be little different from returns on other assets. Some researchers have recently found that, because of the weak correlation between art asset returns with other returns, there may be a case for the inclusion of art assets in a diversified portfolio, though investing in art is inherently risky. Furthermore, the evidence clearly suggests that, contrary to the view of the art trade, "masterpieces" under perform the market, although the precise interpretation of this finding is still open for study.

We move on in section two to discuss other studies that have made use of auction prices to determine the importance of individual works of art and individual artists. Several recent studies have used auction prices to test alternative models of the innovative process in artistic work, thus bringing some of the ideas of classical aesthetics into the area of testable hypotheses. One primary finding indicates that artists in different periods have produced their greatest works at very different ages. This has led some authors to theorize that conceptualists produce their greatest works when young and experimentalists produce their greatest works when older. We conclude this section by discussing whether different auction houses achieve systematically different prices. The evidence suggests that there are fairly long periods in which art prices may diverge across geographic areas and even auction houses.

There is now considerable theoretical research on strategic behavior in auctions, much of it in response to empirical findings. Section three focuses on studies that determine the influence of the auction mechanism on prices. We begin with a discussion of the Christie's and Sotheby's price fixing case and the role of the competitive behavior of the auction houses in the determination of prices. We review the public record of the criminal trial of Sotheby's former Chairman, who was accused of price fixing, to show how the collusion with Christie's, the other great public auction house, was actually engineered. Contrary to the way the proceeds from the settlement of the civil suit in this case were distributed, we show that actual buyers were almost certainly not injured by

the collusion, but that sellers were. In addition, based on the public record of settlement, it appears that the plaintiffs in the civil suit were very handsomely repaid for their injury by the auction houses.

We continue in section three to discuss the role of experts and price estimates in auctions. Some evidence suggests that art experts provide extremely accurate predictions of market prices, but that these predictions do not optimally process the publicly available information. Next, we look at whether items that fail at auction subsequently return less than items that have not failed. We then discuss sale rates and reserve prices. High reserve prices, and the resulting high unsold (“buy-in”) rates are best explained as optimal search in the face of stochastic demand. We also discuss why reserve prices are secret. Finally, extensive research has documented that the prices of identical objects are more likely to decline than to increase when multiple units are sold, and this has led to a large body of theoretical research. Subsequent empirical research has tended to document declining demand prices even when the objects are imperfect substitutes, although the empirical analysis required in this case must be much more sophisticated.

In section four, we conclude our review of studies on art auctions. Because of the unique nature of many art objects and the effect of the auction mechanism on price, the interpretation of market prices requires great care. However, art auctions provide key information for the evaluation of artistic work, and they also provide a key laboratory for testing and refining economic models of strategic behavior.

### *1.1. The mechanics of art auctions*

Historically, the major auctioneers of art have been the English houses of Sotheby’s and Christie’s. These firms, along with other smaller or regional houses such as Phillips in England or Butterfields in California, have developed and refined the rules of what have now come to be called “English” or “ascending price” auctions. Almost all art is auctioned in this ascending price format. Bidding starts low, and the auctioneer subsequently calls out higher and higher prices.<sup>1</sup> When the bidding stops, the item is said to be “knocked down” or “hammered down”, and the final price is the “hammer price”.<sup>2</sup> Not all items that have been put up for sale and “knocked down” have been sold. Sellers of individual items will set a secret reserve price, and if the bidding does not reach this level, the items will go unsold. Auctioneers say that an unsold item has been “bought-in”. Furthermore, the art trade refers to “bought-in” items as having been “burned”. That is, their value has been hurt because they failed to sell at auction. As we show below, sale rates vary tremendously across time and across different types of auctions.

<sup>1</sup> What is called an English auction is, in fact, Roman. The word auction comes from the Latin “auctio”, which means to ascend.

<sup>2</sup> For an interesting history of the development of auction rules in late seventeenth century London, see de Marchi (2004).



An item that has not been sold is rarely, if ever, actually bought by the auction house. It may be put up for sale at a later auction, sold elsewhere, or taken off the market. It is a part of the auctioneer's art to "get the bidding started", and this may involve accepting fictitious bids ("off the chandelier" or "from the order book") so long as the bidding has not exceeded the reserve price. Legally, the auctioneer is bidding on behalf of the seller when this occurs, but must refrain from accepting further bids on behalf of the seller once the bidding exceeds the reserve price.

Auction houses differ with respect to whether they announce during the sale whether an item has been "sold" or is merely "knocked down" and is unsold. In New York, all the auction houses have been compelled by law since the early 1980s to announce whether the bidding has resulted in a sale. The practice elsewhere varies by location and auction house, but there has clearly been a slow movement toward adopting the practice originally enacted by law in New York. While difficult, it is sometimes possible during an auction, if one listens carefully, to determine whether an item has been sold or "bought in".

Prior to an auction, it is common for a pre-sale catalogue to be published with information on the individual items coming up for sale. Common information included in the pre-sale catalogue is the title of the painting, the artist, the size of the painting, the medium, whether the painting is signed, monogrammed or stamped, the provenance, number of exhibitions, and a bibliography of published critiques of the painting. The auction houses also publish a low- and a high-price estimate for the work. The auction house does not publish, and indeed is very secretive about, the seller's reserve price for the work of art. The auction houses do commonly observe an unwritten rule of setting the secret reserve price at or below the low estimate, but the auctioneer is very careful about revealing anything about the reserve price during the bidding process.

Auction houses earn income primarily from commissions charged to buyers and sellers. The commission charged to buyers is called the "buyer's premium". The total sale price to the buyer is thus the sum of the "hammer price" and the buyer's premium. In recent years published buyers' premiums have typically hovered around 10 to 20 percent of the hammer price of an object, with large purchasers paying the lower amount. Although buyers may attempt to negotiate special arrangements regarding buyers' premiums, it is our impression that the typical buyer purchases such a small fraction of the objects on sale at a particular auction house that special terms for buyers are unusual.

Sellers also pay a commission to the auction house called the "seller's commission". Although the seller's commission is often stated as a percentage of the hammer price (typically 10 percent), it is our impression that actual sellers' commissions are often negotiated arrangements that differ by seller. In some cases, sellers pay no commission and may even be guaranteed a minimum sale price. Some key issues related to the negotiation of sellers' commissions, and the extent of competition and collusion in the setting of commission rates have recently surfaced in the trial of Alfred Taubman, former Chairman of Sotheby's, who was convicted of price fixing. We discuss issues related to competition among auction houses in more detail below.

## 2. The price of art

Art objects are generally unique, so that determining time-series movements in their prices requires careful measurement and extensive data. We begin the discussion below in Section 2.1 with a review of the methodology used in constructing price indices. A primary goal of the measurement of time-series movements in art prices is to evaluate the benefits of including art assets in a balanced investment portfolio, and we review the key findings on this topic in Sections 2.1.1 and 2.1.2. Since the key parameters for making this decision are difficult to estimate, this issue deserves far more research. In Section 2.2 we review some of the work by Galenson that looks at when, in an artist's career, his most important works were produced. By first identifying regularities regarding age of artist in different genres of art, Galenson develops a theory as to why these regularities exist. Finally, in Section 2.3, we focus on studies that have shown persistent differences in prices at different auction houses and in different locations.

### 2.1. Art price indices

A key feature of art auctions is that the items on sale are typically unique, or nearly so. The result is that there will be some ambiguity in the construction of a single index of the movement of prices over time. One concern about simply using average prices is that price rises may be exacerbated during booms as "better" paintings may come up for sale. For example, Wynne Kramarsky, whose family formerly owned Van Gogh's *Portrait of Dr. Gachet*, said of the London market prior to the unsuccessful sale of May 15, 1990: "I did not think that London was poor in terms of performance; I thought that the pictures were not up to it" [Watson (1992, p. 10)]. In general average prices will indicate variability over time in art prices that is better described as movements in the heterogeneity of the quality of the objects offered, rather than movements in prices for the same objects.

The extent of heterogeneity, and thus the ambiguity in the construction of auction price indices, differs across the items typically offered for sale by auction. Identical prints may be offered for sale monthly, while identical Impressionist paintings, such as the *Portrait of Dr. Gachet* may not be offered at all in a single decade.

Most art auction indices are based on a model where the price of the  $i$ th object sold in time period  $t$  is

$$p_{it} = p_i + p_t + \varepsilon_{it}, \quad (1)$$

where  $p_i$  is the fixed component of the price that reflects the unique and fixed character (or "quality") of the object,  $p_t$  reflects the index of aggregate movements in prices, and the remainder is an idiosyncratic error term. The key distinction in the construction of price indices is whether the fixed component is treated as determined by a small number of hedonic characteristics,  $x$ , that may be controlled by regression, or whether it is treated as a parameter that must be controlled explicitly.

“Hedonic models” control for the fixed effect  $p_i$  with the assumption that  $p_i = \beta x_i + \varepsilon_i$ , where  $\varepsilon_i$  is an error term independent of the  $p_i$ 's, and estimate

$$p_{it} = \beta x_i + p_t + \varepsilon_i + \varepsilon_{it}. \quad (2)$$

Alternatively, “repeat sale” models include a dummy variable for each painting.

The great attraction of hedonic models is that all the data may be used in the estimation, including data on objects that are only offered for sale once in the sample period. The disadvantage of these models is the strong assumption that a (typically small) set of  $x$  variables captures much of the variability in the fixed components of price (important if the estimates of the time effects are to be precise) and that the characteristics of the objects offered do not vary systematically over time (important for unbiased estimates of the time effects). An example of a painting where this may be a problem is Van Gogh's *Two Rats*. By hedonic characteristics such as artist and medium, this should be one of the most expensive paintings, equivalent to *Portrait of Dr. Gachet*. However, *Two Rats* is widely perceived to be a painting in much less demand than *Dr. Gachet*. A hedonic model would rarely be able to pick this up. If less important paintings such as *Two Rats* tend to come up when the art market is struggling and more important paintings such as *Dr. Gachet* tend to come up for sale during boom times, then an index constructed from a hedonic model will be biased. Thus, repeat sales indices rigorously control for differences in quality of paintings, whereas hedonic models can only control for effects that are observable to the econometrician.

Although the repeat sale method overcomes the primary disadvantages of the hedonic model, it does so at the cost of discarding much data. There must be at least two observations on a painting's price or it provides no information to help identify the time index. Indeed, depending on the frequency at which repeat sales occur, it may not be possible to identify all the time effects in the model.

The problem of sparse data when using repeat sales models has been extensively explored with regard to real estate sales.<sup>3</sup> One well-known problem is that if only a small percentage of paintings change hands in a given period, this will result in near-collinearity of the dummy variables for the time effects. Near collinearity results in imprecise coefficient estimates and an extremely variable series. Furthermore, sparse data can result in spurious negative autocorrelation in the estimated returns. This effect diminishes as the number of observations increases, but this effect is often present in the beginning of the period in which the returns are being estimated. As a result of the problems described above, Goetzmann and Spiegel (2003) recommend constructing indices for most major styles of art on at most a semi-annual frequency and for some styles, at most at an annual frequency.

For use in computing sub-indices in areas where the data is sparse but where there is ample data elsewhere, Goetzmann and Spiegel (1997) have developed a “Distance-Weighted Repeat-Sale” procedure where the distance is defined as the distance in characteristic space. They apply this method to calculating housing indices for a particular

<sup>3</sup> See Goetzmann (1992) and Goetzmann and Spiegel (1997).

neighborhood, but the method could be easily modified for use on particular sub-indices of art.

One other issue to be aware of in repeat sales data is identifying that the paintings are actually the same. Many paintings can have the same title, size, author, painting date, and medium, but yet are different paintings. Depending on the way in which the dataset was constructed, this may or may not be a problem. For example, [Mei and Moses \(2002\)](#) construct their dataset by looking at the provenance (that is, ownership history) of a particular painting, and then determining whether a prior sale is listed in the provenance. This method avoids false matches. However, when [Goetzmann and Spiegel \(2003\)](#) requested all existing repeated-sales from Gabrius, a company that tracks art auction sales, they found a number of unusual observations and found them to be mismatches. One sure way to avoid mismatches is to go back to the art catalogues and compare photographs of paintings.

Comparisons of the results from repeat sale and hedonic models have been reported by [Chanel, Gérard-Varet and Ginsburgh \(1996\)](#). The overall results indicate that both hedonic and repeat sales regressions yield estimates of real rates of return in art assets over long intervals that are the same magnitude. Hence, in some cases the hedonic model may also provide adequate estimates of time-series movements in aggregate prices. However, the danger remains that systematic movements in the unobserved characteristics of the objects being offered for sale may bias the results.

The nature of possible systematic movements is made clear when we do a detailed comparison using our data on Impressionist and Modern Art, as we did in [Ashenfelter and Graddy \(2003\)](#). When yearly price indices are constructed, the two types of indices at first appearance are very similar. [Figure 1](#) presents a graph of the hedonic and repeat sales price indices for Impressionist and Modern Art from 1980 to 1991. The correlation between the two estimates is 0.9559, the standard deviation of the hedonic price index is 1.024, and the standard deviation of the repeat sales index is 1.166. However, because of movements in the very last year, the two indices give very different internal rates of return. The hedonic index gives a real return of about 4 percent, while the repeat sales price index results in a real return of about 9 percent! Which is correct? For 1991, our data ends in May. The “major” impressionist sales are generally held in October. One explanation is that the hedonic index has underestimated the returns for this short period of time, because it was unable to correct for quality differences that occur during sales in the early part of the year. An alternative explanation is that because the repeat-sales index is based on such a small number of paintings during that period, these paintings were unrepresentative (i.e. their price held up better in poor market conditions) of the market as a whole. Our hedonic model incorporates as many as 8792 observations, while the repeat-sale estimates are based on only 474 observations.

It is not necessary to discard a great number of observations for all types of art. For example, the number of observations that are discarded when using repeat sales is smaller when studying the market for prints, as many prints are virtually identical. [Pesando \(1993\)](#) excludes less than one percent of the realized prices on modern prints

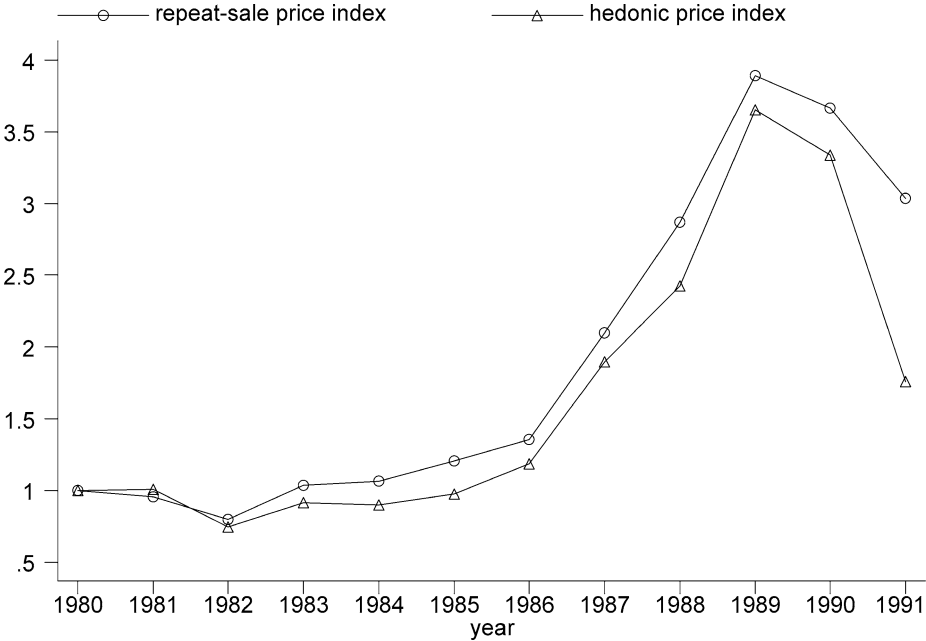


Figure 1. Repeat-sales and hedonic indices for Impressionist Art.

when using the repeat-sale methodology to construct the print price index. In this case, repeat-sales models would dominate hedonic models.

One can also measure the extent to which one type of index deviates from the other. Suppose, for example, that the repeat sales index,  $x$ , is the true index, and  $x^*$  is a measured hedonic index, and  $v^*$  is independent of  $x$ , so that

$$x^* = x + v^*.$$

If  $x$  were measured exactly, then a regression of  $x^*$  on  $x$  would give a slope of unity. If  $x$  is measured with error, the difference from unity provides an estimate of the measurement error as a fraction of the total variance in the repeat-sales index. (A more complex model would have  $x$  measured with error, but uncorrelated with  $v^*$ , say as a result of sampling error.) Computing the above regression, we find that  $x$  has a coefficient of 0.8400, and a standard error of 0.086, thus indicating statistically significant measurement error. The implication is that about 16 percent of the variance in the repeat sales measure of prices is measurement error.

Many of the recent papers that study art as an investment use regressions on repeat sales to construct an index of returns with the purpose of determining the suitability of art for investment. Rather than using a dummy variable for each painting as discussed above, these papers in effect difference prices for two identical paintings. Thus, taking Equation (1) above, and letting  $s$  refer to date of sale,  $b$  refer to date of purchase, the

individual painting characteristics ( $p_i$ ) drop out and we are left with

$$p_{is} - p_{ib} = (p_s - p_b) + \varepsilon_{is} + \varepsilon_{ib}. \quad (3)$$

Price is expressed in logs, and these studies interpret the above equation as follows.  $p_{is} - p_{ib}$  is the continuously compounded return for a particular art asset  $i$  between the purchase and the sale (often expressed as  $r_i = \sum_{t=b_i+1}^{s_i} r_{i,t}$ , where  $r_{i,t}$  represents the returns in each of the individual periods between the purchase and sale)  $p_s - p_b$  is the sum of the average return (that is average return for all paintings in the portfolio) in each period  $t$  between the purchase and sale of painting  $i$  (often expressed as  $p_s - p_b = \sum_{t=b_i+1}^{s_i} \mu_t$ ) and  $\varepsilon_{is} - \varepsilon_{ib}$  is the sum of the error terms in each period.<sup>4</sup> Our derivation suggests that the dummy variables for each pair should equal 1 at the time of sale,  $-1$  at the time of purchase, and 0 in all other periods. Goetzmann (1992, 1993) shows it is more efficient to allow the dummy variables to equal 1 during the periods between purchase and sale, zero otherwise, and then do a generalized least squares regression using the weights suggested by Case and Shiller (1987), which are based on the time the painting is held. The intuition behind the weighted regression is that the longer a painting is held, the larger the error term.

In the first stage of Case and Shiller's (1987) method, the log of the ratio of the sale price to purchase price is regressed on time dummy variables. In the second stage, a regression of the squared residuals from the first stage is run on a constant term and the time between sales. In the third stage, a generalized least squares (weighted) regression is run that repeats the stage-one regression after dividing each observation by the square root of the fitted value in the second stage.

A recent study by Biey and Zanola (2004) has combined the hedonic and repeat sales approaches. They use a method developed by Carter Hill, Knight and Sirmans (1997) for the real estate market, in which a hedonic and a repeat sales equation is jointly estimated. Case and Quigley (1991) pioneered this methodology for real estate. The coefficients on the time dummy variables are then restricted to be the same in the two regressions. They show that the standard errors on the regression estimates are decreased by using the combined method but, of course, only under special assumptions does this procedure avoid bias.

In this volume, Ginsburgh, Mei and Moses (2006) provide a thorough analysis of the different methods of constructing art indices. Other authors that have calculated price indices for art include Anderson (1974), Stein (1977), Baumol (1986), Frey and Pommerhne (1989), Buelens and Ginsburgh (1993), Pesando (1993), Goetzmann (1993, 1996), de la Barre, Docclo and Ginsburgh (1996), Chanel, Gérard-Varet and Ginsburgh (1996), Pesando and Shum (1996), Czujack (1997) and Mei and Moses (2002). The

<sup>4</sup> This methodology was developed by Bailey, Muth and Nourse (1963) and used by Case and Shiller (1987) and Hosios and Pesando (1991) for the real estate market, and subsequently used by Goetzmann (1993), Pesando (1993) and Mei and Moses (2002) for the art market. In these papers  $\varepsilon_{i,t}$  is assumed to be uncorrelated over time and across paintings.

details of these studies and the estimated rates of return on art assets they contain are presented in Table 1. In addition to art, economists have calculated price indices for other items that are sold at auction. Graeser (1993) has estimated a price index for American antique furniture, and Ross and Zondervan (1989) have estimated an index for Stradivari violins.

The estimated returns to holding art are quite dependent upon the time frame actually studied, which is not unexpected. For example, Goetzmann (1996) estimates real returns from 1907–1977 on auction data to be 13.3 percent, and even after correcting for survivorship problems, the returns remain at 5 percent. In Goetzmann (1993) he finds nominal returns for the period 1766–1986 to be only 3.2 percent. Even among authors looking at similar time frames, the returns can vary. The variation reflects differences in data, along with differences in method. It is difficult to come to any broad conclusions about the differences in estimates when using repeat-sales or hedonic indices. Anderson (1974) finds a real return of 2.6 percent using hedonic indices and 3.0 percent using repeat sales on art data from 1780–1960, and Chanel, Gérard-Varet and Ginsburgh (1996) find real returns of 4.9 percent and 5.0 percent for hedonic and repeat sales indices, respectively, for the period 1855–1969.

Commissions are another issue that various studies treat differently and can clearly affect the levels of returns. Ideally, to properly measure financial returns one would like to account for trading costs and thus to increase the purchase price by the level of the buyer's premium but then decrease the sale price by the level of the seller's commission. However, sellers' commissions are negotiable and not easily available. Therefore, one is left with information on buyers' premiums which different authors treat differently. For example, Goetzmann and Spiegel (2003) do not include premiums. However, Mei and Moses (2002) include buyers' premiums in the data they use to compute returns by increasing the recorded price of all items by the amount of the buyer's premium. This is problematic in that before 1979, auction houses did not charge buyers' premiums. After 1979, buyers' premiums were stable at approximately 10 percent from the mid-1970s to 1992, but then started to rise thereafter, mostly for the less expensive works of art. As premiums have increased over time, including buyers' premiums in repeat sales data in this manner may slightly increase returns.

Almost all studies of repeat sales data only use information on actual sales rather than information on all items, including those that were bought in. Goetzmann and Spiegel (2003) assume the "price" for bought-in items is 80 percent of the low estimate. They then use this "price" to compare indices in which all items were used to calculate the index with indices in which only sold items are used. They find that the indices calculated with and without bought-in data were fairly close to each other. However, they found that including bought-in data increased the explanatory power of the repeat sales regressions by about 20 percent. They conclude that bought-in data should be used in estimation of the repeat-sales indices.

Finally, survivorship bias or sample selection issues are a caveat that should be acknowledged after calculation of an index. As noted by Goetzmann (1996), "the use of repeat-sale data implicitly conditions upon artworks being in demand, and thus un-

Table 1  
Estimated returns to art from various studies

Author	Sample	Period	Method	Nominal return	Real return
Anderson (1974)	Paintings in General	1780–1960	hedonic	3.3%	2.6%*
	Paintings in General	1780–1970	repeat sales	3.7%	3.0%*
Stein (1977)	Paintings in General	1946–1968	assumes random sampling	10.5%	
Baumol (1986)	Paintings in General	1652–1961	repeat sales		0.6%
Frey and Pommerehne (1989)	Paintings in General	1635–1949	repeat sales		1.4%
		1950–1987	repeat sales		1.7%
Buelens and Ginsburgh (1993)	Paintings in General	1700–1961	hedonic		0.9%
Pesando (1993)	Modern Prints	1977–1991	repeat sales		1.5%
Goetzmann (1993)	Paintings in General	1716–1986	repeat sales	3.2%	2.0%*
De la Barre, Docclo and Ginsburgh (1996)	Great Impressionist Other Impressionist	1962–1991	hedonic	12.0%	5%*
		1962–1991	hedonic	8.0%	1%*
Chanel, Gérard-Varet and Ginsburgh (1996)	Paintings in General Paintings in General	1855–1969	hedonic		4.9%
		1855–1969	repeat sales		5.0%
Goetzmann (1996)	Paintings in General	1907–1977	repeat sales		5.0%
Pesando and Shum (1996)	Picasso Prints	1977–1993	repeat sales	12.0%	1.4%
Czujack (1997)	Picasso Paintings	1966–1994	hedonic		8.3%
Mei and Moses (2002)	American, Impressionist, and Old Masters	1875–2000	repeat sales		4.9%
Goetzmann and Spiegel (2003)	Contemporary, Impressionist, Old Masters	1985–2003	repeat sales	–1.2%	
Graeser (1993)	Antique Furniture	1967–1986	neither**	7.0%	
Ross and Zondervan (1989)	Stradivari Violins	1803–1986	hedonic		2.2%

\*As many of the surveys only report nominal returns, the authors calculated the real return rates as follows. For the Anderson and Baumol studies, an inflation rate of 0.7 percent a year was used. This number is based on Baumol's estimate of inflation during the 300 year period of his study using the Phelps-Brown and Hopkins price index. Goetzmann's estimate of inflation during the period of his study (also based on Phelps-Brown and Hopkins) is 1.2%. French price inflation between 1962 and 1992 according to OECD statistics was 7%.

\*\*Assumes random sampling within a portfolio of fixed furniture types.



derstates the risk of obsolescence” (p. 71). Goetzmann (1996) attempts to estimate the degree of this bias. He creates eight hypothetical portfolios of paintings, one for each decade, from 1907 through 1977. He then appraises the pieces as of 1987 using sales data for comparable art works. He then assigns a small residual value to artworks by those painters who did not show up at least once between 1977–1987. He deems these painters to be obsolete. Using this methodology, he finds his estimates to be 8.3 percent below his estimates of the real annualized rate of return for a similar period when the survivorship bias is not taken into account.

However, it is not clear that sample selection bias only serves to decrease returns to art. Many dealers insist that the very best paintings are purchased by museums, and therefore also never show up at auction in repeat sales data. We do not know of any work that has attempted to estimate this possible downward bias. Both survivorship bias and “museum” bias can also apply to hedonic price indices.

The research reviewed above has focused on the accurate construction of indices that measure aggregate price movements of unique objects. As discussed below, these indices are crucial for answering the question of whether art is a good investment.

### 2.1.1. Art as an investment

A primary concern of many of these papers is whether art outperforms or underperforms stocks and bonds and the correlation of art investment returns with other investment portfolios. Once a rate of return on art assets is calculated, it is possible to use this return to decide whether it may be sensible to include art investments in a diversified portfolio. Generally, art investments are more attractive as investments (using the standard capital asset pricing model – CAPM) the greater is their return relative to the return on a risk free asset and the weaker the correlation (or beta) between art investment returns and the return on other assets. Pesando (1993) has used the standard market model to assess these two characteristics of art investments in the case of modern prints. Pesando estimates the model:

$$R_t^P - r_{f,t} = \alpha + \beta(R_{m,t} - r_{f,t}) + u_t, \quad (4)$$

where  $R_t^P$  denotes the return on the print portfolio,  $R_{m,t}$  denotes the return on the market portfolio (Pesando uses the S&P 500 stock index), and  $r_{f,t}$  denotes the risk free rate (Pesando uses 180-day Treasury Bills). Pesando estimates a  $\beta$  for the entire print portfolio of 0.315 and estimates negative, but insignificant, risk adjusted returns. This implies that print investments tend to reduce the riskiness of a portfolio comprised of stocks only.

Determining whether art outperforms or underperforms a market portfolio is not an easy question to address. First of all, as pointed out above, there are many problems with the calculation of the returns to art, beginning with selection bias in the data. As all of the sales prices are drawn from auction records, with repeat sales regressions only paintings that have been re-auctioned are included. This excludes both the high end and the low end of the return distribution. Paintings that fall drastically in value or are not generally

in demand are usually not resold at auction; in addition, paintings that are donated to or bought by museums do not reappear. Furthermore, whether or not an owner decides to sell a painting at auction may be determined by whether or not the painting has increased in value. Other problems with estimating returns are that transaction costs are excluded and in contrast to stocks and bonds, these can be quite high (as much as 25 percent of the value of the object considering both buyers' premiums and sellers' commissions). Finally, there is significant theft and fire risk (and hence insurance costs) and cleaning costs involved in investing in art.

On the other hand, unlike stocks and bonds, art also pays some dividends in the form of pleasure the viewer (and owner) receives. In principle, the value of these dividends could be measured by the rental cost of similar art assets, but we are unaware of any study that has attempted to do this. Moreover, it seems unlikely that these returns would be significant for a large, diversified art portfolio that is not displayed.

Baumol (1986) and Goetzmann (1993) tend to concur that art is dominated as an investment vehicle. Goetzmann (1993, p. 1370) writes "While returns to art investment have exceeded inflation for long periods, and returns in the second half of the 20th century have rivaled the stock market, they are no higher than would be justified by the extraordinary risks they represent". Goetzmann does not formally estimate a CAPM, but simply reports correlations of art returns with inflation, the Bank of England Rate, consol bond returns, and the London Stock Exchange.

Although their estimates of the return to art are not significantly different from previous estimates, Mei and Moses (2002) take a different view. They argue that "a diversified portfolio of artworks may play a somewhat more important role in portfolio diversification than discovered in earlier research". They base their conclusions on the finding that their art price index has lower volatility and a much lower correlation with other asset classes than reported in previous research. They report that these differences are partly due to sample selection and partly due to a different time frame studied. Although Mei and Moses (2002) estimate a more sophisticated form of CAPM than has previously been estimated for art, they primarily base their conclusions on their estimates of the art index and simple correlations with bond and stock portfolios.<sup>5</sup>

Some authors have looked at the financial returns to holding other collectible items. For example, Ross and Zondervan (1989) estimate the real returns to holding Stradivari Violins between 1803 and 1987 to be 2.2 percent, and Graeser (1993) estimates

<sup>5</sup> For the CAPM, Mei and Moses follow Campbell (1987) and estimator

$$r_{i,t+1} = E_t[r_{i,t+1}] + \sum_{k=1}^K \beta_{ik} f_{k,t+1} + \varepsilon_{i,t+1},$$

where  $r_{i,t+1}$  is the excess return on asset  $i$  held from time  $t$  to time  $t + 1$ .  $E_t[r_{i,t+1}]$  is the conditional expected return on asset, conditional on information known to market participants at the end of time period  $t$ . It is allowed to vary over time [see Mei and Moses (2002) for details].  $f_{k,t+1}$  are excess returns on  $k$  different asset classes.

returns to holding antique furniture between 1967 and 1986 to be 7 percent. For a very good survey of papers calculating the rate of returns in various markets, see Frey and Eichenberger (1995).

It is clear that some collectors have benefited enormously from holding art. As documented in Landes (2000), Victor and Sally Ganz are a nice example. The Ganz's operated a family costume-jewelry business, but collected art on the side. Through clever purchases, they managed to build a collection worth over \$250 million over a fifty year period. For example, one of their successes was a Picasso that they acquired in 1941 for \$7000 which they subsequently sold at auction for \$48.4 million. Overall, Landes estimates that they achieved between about a 12 percent and 21 percent per year annual real return, depending upon the auction in which they sold their collection. Landes argues that their high return occurred because of their skill in investing, not because of pure luck. His claim is that the return is not based on a few big winners, but they made consistently good purchases. This argument is still open to debate as their high returns may statistically have resulted from them being on the upper tail of a random distribution in returns to art. Nonetheless, as in stock-picking, it may be possible that some individuals have skills that allow them to consistently outperform the market.

McAndrew and Thompson (2004b) study the loan quality of fine art. As with other investments, they find that the value of fine art in case of bankruptcy is quantifiable, and thus art can be used as quantifiable collateral on loans.

What can we conclude from these studies on the returns to art? On the positive side, the studies reviewed in Table 1 mostly report positive returns and many of the studies show that the returns to art may outperform bonds. Furthermore, the correlation to other investment portfolios may be low. On the negative side, returns to art generally appear to be less than the real rate of return on common stock. In addition, as is clear from Figure 1, investing in art over a short period of time can be risky.

From current research, it is difficult to conclude whether art should be included in a diversified portfolio. It appears that different views about the financial benefits of investments in art assets are primarily based on empirical issues that revolve, in part, around the temporal instability and sensitivity of the estimates of key parameters related to the market performance of art investments. This suggests that an important area for additional research is the development of a more general empirical model that will provide an explanation for temporal instability and thus lead to better-informed decisions.

### 2.1.2. *The masterpiece effect*

Pesando (1993) describes the “Masterpiece Effect” by quoting art dealer Edward Merin: “...it’s always better to buy one \$10,000 object than ten \$1000 objects, or one \$100,000 object – if that is what you can afford – than ten \$10,000 ones”.<sup>6</sup> There have now been several authors who have tested for the masterpiece effect. Pesando tests for

<sup>6</sup> Quote originally taken from “Antiques”, *Art and Auction*, September 1988, p. 131.

the effect by constructing a portfolio of the top 10 or 20 percent of prints by price, where price is determined during the first few years of his sample. If the “art trade” view is correct, the estimated price indices for these “Masterpieces” should uniformly outperform the general portfolio. He finds no support for this view and in fact finds that in part of his sample, masterpieces provide the lowest cumulative return. [Mei and Moses \(2002\)](#) measure a similar negative effect for masterpieces, and in fact estimate this effect to be uniform across American, Impressionist and Old Master samples. Other authors have found no masterpiece effect [[Goetzmann \(1996\)](#); [Ginsburgh and Jeanfils \(1995\)](#)].

As pointed out by [Pesando \(1993\)](#), there should not be a positive masterpiece effect. An efficient art market should capitalize favorable properties into their prices, and their risk-adjusted rates of returns should not exceed that obtained on other art objects in the same class. However, the art market and its participants may not always be efficient. [Pommerehne and Feld \(1997\)](#) conclude that museums outside the US pay above average prices in auction markets. If museums generally purchase masterpieces and do not resell them, this could contribute to a positive masterpiece effect.

Various plausible reasons exist as to why there should be a negative masterpiece effect. [Mei and Moses \(2002\)](#) speculate that it may be due to overbidding and then mean reversion. This explanation appears quite reasonable given the way that various studies above have defined “Masterpieces” as the highest price paintings that were sold. If a “Masterpiece” is defined purely by price, there may be some paintings in the “Masterpiece” sample that randomly commanded a higher price, perhaps because two or more bidders had high private valuations for the paintings. At a later auction, the prices on these paintings revert to the mean, thus resulting in a negative “Masterpiece” effect.

A different explanation for the negative “Masterpiece Effect” may be what [Goetzmann \(1996\)](#) terms “survivorship bias”. It is likely that the more expensive paintings remained in the sample throughout, even if they decreased in value, whereas less expensive paintings have dropped out of the sample. Hence it may appear that “Masterpieces” have underperformed in the sampled data, but in actuality less expensive paintings that have underperformed are no longer in the sample. A summary of papers estimating a “Masterpiece Effect” is presented in [Table 2](#).

A very nice addition to the literature on Masterpieces would be a classification of Masterpieces based on something other than price. For example, [Galenson](#), in a number of studies, uses a tabulation of illustrations in published surveys of art history [see, for example, [Galenson \(2002, 2004\)](#)]. He uses these tabulations for two primary purposes. First, to determine in which point of an artist’s career his most important work was created (see below), and secondly, to simply rank different artists. Using these tabulations, he ranks painters and paintings and compares rankings using tabulations from different textbooks. A very interesting use of these rankings would be to determine whether or not a “Masterpiece Effect” exists for the highest ranked artists.

[Schoenfeld \(2004\)](#) uses the *Konstkompass* rankings to determine whether or not certain artists can be classified as “Masters”. The *Konstkompass* is a publication that ranks artists based on where and how often an artist’s work is shown and written about. For example, an artist would receive 650 *Konstkompass* points for an exhibition in the Tate

Table 2  
The “Masterpiece Effect”

Author	Sample	Period	Result
Pesando (1993)	Modern Prints	1980–1992	Return of 11% less for “masterpieces”
Ginsburgh and Jeanfils (1995)	Impressionist, Modern and Contemporary European Masters, Other minor European painters, Contemporary US painters	1962–1991	No effect
Goetzmann (1996)	Paintings in General	1897–1987	No effect
De la Barre, Docclo and Ginsburgh (1996)	Impressionist	1962–1991	Great impressionists return 4% more than other impressionists
Mei and Moses (2002)	American, Impressionist, and Old Masters	1875–2000	A 10% increase in price reduces returns by 1%
Schoenfeld (2004)	Contemporary Art	1982–1985	No effect

Modern, London, or 800 points at the Guggenheim Museum, New York. Schoenfeld defines a “Master” by using the top twenty artists as ranked in 1983 and then follows the prices they receive at auction. She also defines “Master” by price by following the top twenty artists when ranked by gallery price. She does not find a Masterpiece effect in her study, whether using rankings or gallery prices.

We have summarized the findings of papers estimating a masterpiece effect in Table 2. Out of the six studies reported, only one study, [de la Barre, Docclo and Ginsburgh \(1996\)](#), finds a positive masterpiece effect. Interestingly, they have constructed their subsamples by choosing masterpieces by reputation of painter, rather than by price. More studies where price is not the primary determinant of a masterpiece are clearly needed in order to separate out any masterpiece effects from mean reversion in price.

## 2.2. *The life cycle of an artist*

Galenson and Weinberg have used art auction prices to determine at what point during his career an artist produces his most important work. For example, [Galenson \(2000\)](#) shows that in a sample of 42 of the most important Contemporary artists (broadly defined by auction houses as those who have become known since World War II), more than 90 percent of the artists born before 1920 did their most important work after the age of 40, whereas more than three-quarters of those born after 1920 did their most

valuable work before the age of 40. Galenson reconciles this fact by saying that the nature of modern art changed during the 20th century.

In Galenson and Weinberg (2000), the authors find that within two successive cohorts of American modern painters, the second cohort produced their most valuable works at a younger age than the first cohort, and in Galenson and Weinberg (2001), the authors document that painters born at the very end of the 19th century tend to produce their most valuable paintings at a younger age than other 19th century painters. Using this information, Galenson has gone on to develop a controversial hypothesis as to why some artists have produced their best works early in their career, and others have produced their best work later in their careers.<sup>7</sup>

Galenson and Weinberg's evidence is interesting. In this research auction prices have been used to discover the empirical regularities in which different artists produce their more important works at different ages. This empirical regularity has thus spawned a debate as to why this is so.

While in many of his papers Galenson uses prices to test for age and importance of work, as discussed above, in several of his papers he uses number of illustrations in art history texts. There are possible advantages to using measures other than price. For example, in Galenson (2002) he constructs a variable (a Gini ratio) that measures how unequally a given artist's paintings are illustrated. He then uses this measure to determine the extent to which an artist has one or two great works, or a large number of important paintings.

As in some of his work Galenson uses critical evaluation and in other works he uses auction prices, a natural question to ask is how well do results using auction prices match up with critical evaluation. In Galenson (2000, 2004) he demonstrates that critics' judgment of when in an artist's career his most important works were executed coincides very closely with market valuation.

### 2.3. *The law of one price*

The "law of one price" dictates that in the absence of different transactions costs, no systematic price differences should exist between distinct markets. Several authors have tested for price differences in different auction houses and in different geographical locations and have found that the law of one price does not hold.

Pesando (1993) focused on the sale of identical prints in different markets which occur within 30 days of each other for the period 1977–1992. For the entire period, he found that prices were 7 percent higher in New York than in London, and 10 percent higher in New York than in Europe. However, these differences were not statistically significant for the period 1977–1989. For the remaining period of the sample, 1989–1992, he found that prices were 11 percent higher in New York than in London and 17 percent higher in New York than in Europe. Pesando (1993) describes the trade explanation as being the presence of Japanese buyers in the New York market during that

<sup>7</sup> See Galenson (2004) for a review.

period, though one would expect any systematic price differences to disappear when buyers respond to incentives. Pesando also measures significant differences among auction houses. For the entire period, Pesando found that prices average 14 percent higher at Sotheby's New York than at Christie's New York, but there was no difference in the prices of prints at Sotheby's and Christie's in London.

Mei and Moses (2002) find mixed evidence on the law of one price. When they do find price differences, these differences tend to be small.

Pesando and Shum (1996) look specifically at the law of one price using prints only by Picasso. As this is a narrow and homogeneous segment of the market, this can provide a good test. They further restrict their sample to the 100 prints that comprise Picasso's Volland Suite, and test for price differences on this sample. They compare the prices of identical prints that were sold in two different international markets or auction houses within 30 days of each other. Using the entire sample of Picasso prints, they find no significant difference between New York and London, but they find that prices are significantly higher in Europe (excluding London) than in the US. Furthermore, they find higher prices at Sotheby's than at Christie's in New York, and at Kornfeld's than at the rest of the world. Using only the Volland suite, the price differences between New York and Europe disappear, as do the price differences between Kornfeld's and the rest of the world (though the Kornfeld price difference is again significant if the window is increased to 90 days). For the Volland suite only, prices remain significantly higher at Sotheby's than at Christie's New York.

De la Barre, Docclo and Ginsburgh (1996) compare prices of Great Masters at Drouot in Paris with Christie's and Sotheby's prices in New York and London. They use the coefficient on the auction house in a hedonic regression to test for price differences. They find that prices fetched in Paris are significantly lower than prices fetched for similar items in New York or London. They give two reasons for the differences. First, they argue that it may simply be unobserved quality differences that are driving the price differences, as Christie's and Sotheby's, being more well-known, may attract better paintings. Secondly, they point out that the French government has the right to prevent "important" works from leaving France, if they are sold by collectors living in France. This may discourage foreign collectors from purchasing and thus have a downward effect on price.

Ashenfelter (1989) studied differences in prices for wine between auction houses that were attributed to changes in buyers' premiums. In the spring of 1986, buyers' premiums were 10 percent at Sotheby's London (and at other locations), but Christie's in London had no buyer's premium. In the spring of 1986, prices at Sotheby's in London were 12 percent less than prices at Christie's in London, perhaps reflecting the difference in buyers' premiums. In the fall of 1986, Christie's had instituted a 10 percent buyer's premium in the London auctions. In auctions held in the fall of 1986, there was no difference in prices, while in an auction held in the spring of 1987, prices at Sotheby's in London were 5 percent higher than at Christie's, and in the fall of 1987, prices at Sotheby's in London were 4 percent lower than at Christie's. These price movements may provide empirical support for the observation that the incidence of the buyer's

Table 3  
The law of one price

Author	Sample	Period	Result
Ashenfelter (1989)	Wine	1986	Differences in prices reflect differences in commission rates
Pesando (1993)	Modern Prints	1977–1992	Prices average 14% higher at Sotheby's NY than at Christie's NY; prices were 7% higher in New York than in London; prices were 10% higher in New York than in Europe
Pesando and Shum (1996)	Picasso Prints	1977–1993	Prices average 7% higher at Sotheby's NY than at Christie's NY; no significant differences in price between NY and London; prices were 2% higher in New York than in London
De la Barre, Docclo and Ginsburgh (1996)	Old Masters	1962–1991	Prices were significantly lower at Drouot's in Paris than at Sotheby's or Christie's in New York or London
Mei and Moses (2002)	American, Impressionist, and Old Masters	1875–2000	Mixed evidence on differences between Sotheby's and Christie's; differences when they do exist are small

premium should fall on the sellers. As described below, significant changes to sellers' commissions and buyers' premiums occurred during the 1990s. These changes may provide an interesting subject for study by economists.

A summary of papers testing for the law of one price is presented in [Table 3](#).

### 3. The auction mechanism and the price of art

We begin our discussion in Section 3.1 with a review of the Sotheby's–Christie's price fixing case. This case shows just how important it is to understand the influence of the auction mechanism on price, and how this influence can easily be misunderstood.

In the remainder of Section 3, we discuss various aspects of auctions and the influence of each of these aspects on the price achieved. We begin by discussing whether auctioneer's pre-sale estimates are biased, and whether or not they take account of all information available. Most work suggests that these estimates are unbiased, but other work suggests that there may be systematic under and over predictions. We also discuss implications of the width of the spread between the low estimate and the high estimate. In the following three sections we examine the role of reserve prices in art auctions. We



begin by reviewing work that tests whether paintings that fail at auction are “burned” (lose value), as is claimed by many individuals in the art trade. We then move on to discuss sale rates across different types of auctions and why they may differ, and we end our examination of reserve prices by discussing why they are secret. In the last section we review the large body of work on declining prices.<sup>8</sup>

### *3.1. Competition (or collusion) between auction houses*

Prior to 1995, Sotheby’s and Christie’s were in fierce competition for consignments. At times, they would drastically cut commission rates paid by sellers, in many cases to nothing, make donations to their favorite charities, and even extend financial guarantees. In March of 1995, this competition abruptly ended. Christie’s announced that it would charge sellers a fixed nonnegotiable sliding-scale commission on the sales price, and a month later Sotheby’s announced the same policies. Detailed documents kept by Christopher Davidge, Christie’s former chief executive, show that the abrupt change was due to a price-fixing conspiracy. By admission, the conspiracy involved at least Christopher Davidge and Diana Brooks, Sotheby’s chief executive, and it was alleged to have involved Sir Anthony Tennant and A. Alfred Taubman, the chairmen of Christie’s and Sotheby’s, respectively. In fact, after a lengthy criminal trial, Taubman, a US citizen, was convicted of price fixing, which is a felony in the US. Although Tennant, a UK citizen, was also indicted in the US, price fixing is a civil offence in the United Kingdom and thus he was not extradited or tried. Christopher Davidge (and in some cases Sir Anthony Tennant) had kept detailed records describing the conspiracy. A civil suit, which has been settled, also alleged that Christie’s and Sotheby’s conspired since 1993 to fix buyers’ premiums. Because it ended in a public trial, this lawsuit provides an extraordinary window for viewing the operation of successful price conspirators.

The case progressed as follows [see especially [Stewart \(2001\)](#) for a detailed description]. The Justice Department agreed in January of 2000 not to prosecute Christie’s in return for its cooperation in the case. Diana Brooks, former president and chief executive of Sotheby’s, pleaded guilty to one felony count of price-fixing on October 5, 2000, and promised to cooperate fully in the government’s investigation. In September of 2001, Sotheby’s pleaded guilty to conspiring with Christie’s to fix sellers’ commissions, and agreed to pay a fine of forty-five million dollars over five years. Sotheby’s maintains their innocence with respect to fixing buyers’ premiums. Also in September of 2001, a civil suit was settled where Sotheby’s and Christie’s consented to a payment of two hundred and fifty-six million dollars each to the plaintiffs. This class-action suit comprises anyone who had bought or sold items through the auction houses since 1993.

<sup>8</sup> Our discussion is primarily motivated by empirical regularities (or irregularities!) that researchers have uncovered in their study of art auction prices. For an excellent review of auction theory and the effect of the auction mechanism on prices for telecommunication licenses, please see [Klemperer \(2004\)](#).

From an economist's point of view, the settlement of the civil suit is interesting, but appears to be misguided. Although Sotheby's did not admit to fixing buyers' premiums in the criminal settlement of the case, both Christie's and Sotheby's agreed to each pay \$256 M to both buyers and sellers. This amount was calculated taking the price-fixing of buyers' premiums into account. According to *In Re Auction Houses Antitrust Litigation* (2001), "The proposed plan of allocation estimated the overcharges to sellers as 1 percent of the hammer price, and those for buyers to be 5 percent of the hammer price up to and including hammer prices of \$50,000, and \$2500 for buyers at hammer prices exceeding \$50,000. The net settlement fund would be distributed to class members pro rata based upon each class member's overcharges during the relevant period."

Even if Sotheby's and Christie's admitted to colluding on buyers' premiums, the usual theory of private value auctions implies that, to first order, buyers deserve no compensation! The following is the reason why. When a buyer decides to bid in an ascending price auction, his strategy should be to bid up to his reservation price, if necessary. The price that the winning bidder has to pay is essentially (epsilon above) the reservation price of the second highest bidder. When buyers' commissions are raised, each buyer should reduce his reservation price by an equivalent amount, resulting in a reduction in revenue to the seller by the amount of the buyer's commission. Hence, the entire increase in buyers' commissions should fall on the seller. Thus, the standard model of private value auctions implies that the entire settlement arrangement in the civil suit was misguided!

There are several caveats to this argument. If sellers' supplies are elastic, some sellers may not offer their objects for sell due to the increased commissions. This could result in more buyers competing for the same item, and the increase in the number of bidders for each item may push up the price paid by the winning bidder. Furthermore, in art auctions, the private value assumption doesn't strictly apply. If some bidders are factoring into their value estimate the likely future market value of the piece of art, bidders' values are correlated. Bidders may increase their reservation prices if they believe market values will increase in the future. However, it is unlikely that either of these effects justify buyers receiving 2/3 of the 512 million. In addition, it can be argued that the real losers in this case were the buyers and sellers who did not manage to transact because of the price-fixing. Welfare was clearly reduced, as these transactions were lost.

Ashenfelter et al. (2003) provide a more detailed explanation of the incidence of the settlement and Ginsburgh, Legros and Sahuguet (2004) provide a general theoretical exploration of the incidence of commissions in auction markets with reserves. In this paper, they show that under sufficient assumptions, while commissions indeed make sellers worse off, some buyers may actually gain. Their intuition is that buyers who actually purchase can be better off since the impact of commissions is compensated by less participation and a lower degree of competition.

The criminal trial of Alfred Taubman, previously Chairman of Sotheby's Board, in the fall of 2001 provided some dramatic revelations about the details surrounding the price fixing. Many of these details were provided in the testimony of Diana Brooks, previously President of Sotheby's. First, although remarkably candid about her role in

the fixing of sellers' commissions, Ms. Brooks did not provide any evidence of collusion with respect to buyers' premiums. Second, Ms. Brooks estimated that the collusion on sellers' commissions resulted in higher profits to Sotheby's of some \$10 to \$15 million per year. Assuming that Christie's received the same increased profits implies that a total per year increase in profits would be on the order of \$20 to \$30. Assuming the conspiracy lasted 5 years (approximately the time period involved) suggests a total increase of \$100 to \$150 million. While profits may not be equivalent to the damages suffered by a group for a variety of reasons, it appears that the plaintiffs were more than amply compensated for the harm they incurred, especially in view of the fact that they did not have to proceed to the uncertainty of a trial.<sup>9</sup>

Finally, the details of the arrangement for price fixing revealed by Ms. Brooks suggest that great care was ensured to keep virtually all other employees of the auction houses from learning of the conspiracy. Taubman met solely with Tennant, and Brooks solely with Davidge in arranging the details of the conspiracy. For a full discussion of the price-fixing conspiracy including other economic and legal implications, please see [Ashenfelter and Graddy \(2005\)](#).

### 3.2. Role of estimates and experts

Before an auction takes place, in their pre-auction catalogues, auction house experts provide a low and a high price estimate for each item. Determining the accuracy of these estimates raises some important questions for the study of the role of expert opinion in economic decisions.<sup>10</sup> Of especial interest is the motivation of the auctioneer in choosing the high and low estimates. The theoretical literature stresses that auctioneers should provide truthful information about the items being sold.

[Ashenfelter's \(1989\)](#) results generally show that auction houses are truthful; the average of the auctioneer's high and low estimate is very highly correlated with the price actually received. Furthermore, [Abowd and Ashenfelter \(1988\)](#) find that auctioneer's price estimates are far better predictors of prices fetched than hedonic price functions.

The details of the arrangements for price fixing revealed by Diana Brooks during the Christie's–Sotheby's price-fixing trial provide further insight into the role of experts at auction houses. Brooks reported that at one point her boss, Alfred Taubman, proposed that the auction houses collude in providing clients with similar estimates of the value of their art. Brooks reported that this was impossible because she could not simply tell Christie's departmental experts, who produce the estimates, to do a dishonest job.

While the regressions in [Beggs and Graddy \(1997\)](#) generally uphold these results, they do find systematic under and over predictions. For example, they find that for Contemporary Art, more recently executed artworks are overvalued and longer and wider

<sup>9</sup> Price fixing damages are, by statute, tripled.

<sup>10</sup> [Ashenfelter \(2000\)](#) defines expert opinion as efficient if it incorporates all of the publicly available information that is useful in making predictions. He also provides one example of inefficient expert opinion.

paintings are undervalued. For Impressionist and Modern Art, they find that wider, signed, and monogrammed paintings may be underestimated relative to their value. One explanation for these findings may simply be that auction houses are unwittingly overestimating consumer demand (and hence willingness to pay) for recent Contemporary Art, and underestimating consumer demand for size! Many people in the trade express surprise at the strong correlations that many economists have found between size and price.<sup>11</sup>

Other authors have also found that ex-ante valuations cannot be considered unbiased predictors of market prices, although it is our impression that biases are not quantitatively large when they are precisely estimated. [Bauwens and Ginsburgh \(2000\)](#) study 1600 lots of English silver sold between 1976 and 1991 by Christie's and Sotheby's. They find that Christie's has a tendency to underestimate systematically, while Sotheby's overvalues inexpensive pieces and undervalues expensive ones. [Bauwens and Ginsburgh \(2000\)](#) also show that experts do not take into account all available information. They do this by running a regression of the difference between the actual price and the estimated price on observable characteristics. They strongly reject the hypothesis that the coefficients on the characteristics are jointly equal to zero.

[Chanel, Gérard-Varet and Vincent \(1996\)](#) studied jewelry auctions, and found that experts have an ex-ante valuation that is lower than the hammer price for all types of jewels, except for some watches. They speculate that some strategic undervaluation is occurring. These results are interesting, in part because, as [Milgrom and Weber \(1982a\)](#) show, in general, for auctioneers, "honesty is the best policy".

If price estimates are biased, this raises some interesting questions about the reason for the bias. One possibility is simply that the "experts" make systematic errors because they are not as "efficient" as the linear predictors they are being tested against. Evidence in favor of this hypothesis would be the finding that observed biases are not stable and vary from one sample to another or from one time period to another. Judging from the results reported above, there is certainly some evidence to support this view. [Mei and Moses \(2005\)](#) have recently taken up this issue. They conclude that estimates for expensive paintings are biased upwards. However, [McAndrew and Thompson \(2004a\)](#) find evidence for unbiased estimates.

A related question is, "what motivates the auctioneers when they determine the spread between the high and the low estimates that are published in the pre-sale catalogues?" One explanation of how the spread is determined is by the auctioneer's estimate of the uncertainty or possible variance in the price of the painting. In this case, the high estimate might reasonably be interpreted as the estimate of the mean price plus a multiple of the estimated standard deviation ( $H = \mu + r\sigma$ ). Likewise, under this interpretation, the low estimate would be the mean minus a multiple of the standard deviation ( $L = \mu - r\sigma$ ). With this interpretation the high estimate minus the low estimate divided by 2 is proportional to the estimated standard deviation ( $(H - L)/2 = r\sigma$ ) and

<sup>11</sup> For example, see [Anderson \(1974\)](#) and [Beggs and Graddy \(1997\)](#).

the average of the high estimate and the low estimate would be the estimated mean  $((H + L)/2 = \mu)$ . A large difference in the high estimate and the low estimate would therefore signal a high estimate of price variance or a lot of uncertainty. However, as the seller's secret reserve price, by convention, lies below the low estimate, it is very likely that the spread between the high and low estimate is not simply a reflection of the auctioneer's uncertainty surrounding the possible price. If the seller wishes to set a high reserve price, the auctioneer may increase the low estimate. [Ashenfelter, Graddy and Stevens \(2002\)](#) study the plausibility of these two explanations with regard to sales rates in Contemporary and Impressionist Art Auctions. A summary of papers addressing the role of estimates is presented in [Table 4](#).

A recent work by [Beggs and Graddy \(2004\)](#) suggests that auctioneers may not behave fully rationally when setting the pre-sale estimate, but instead may engage in partial information processing. Building on a methodology developed by [Genesove and Mayer \(2001\)](#) for loss aversion in the housing market, they show that experts may engage in "anchoring".<sup>12</sup> Auctioneers may anchor the pre-sale estimates using the price the painting fetched at auction in its previous appearance. This is somewhat surprising given that it may have been years (and in very different market conditions) since a particular painting appeared at auction. Of particular interest, their estimates show similar effects for paintings in which there are prospective (i.e. estimated) gains and losses and no diminishing effects with the size of the loss or gain.

In summary, there is evidence that pre-sale auction estimates are highly correlated with price outcomes, but there is also evidence that these estimates might well be improved by the use of appropriate statistical methods.

### 3.3. *Is there evidence that paintings are "burned?"*

As [Ashenfelter \(1989\)](#) noted, it is often claimed that when an advertised item goes unsold at auction, its future value will be affected. Such items are said to have been "burned". A perceived loss in value after a failed auction has acted as part of the basis for legal proceedings such as "Cristallina, S.A. vs. Christie, Manson and Woods, International, Inc." Cristallina S.A. was a Panamanian corporation engaged in the sale and purchase of art. In 1981, Cristallina consigned eight impressionist paintings to Christie's. During the auction that took place in May of 1982, seven out of the eight paintings failed to meet their reserve price. The prosecution alleged that Christie's did not use sufficient care in marketing and auctioning the paintings, and this lack of care resulted in a loss of value. The suit was eventually settled out of court.

[Beggs and Graddy \(2006\)](#) test whether failed paintings return less than other paintings by constructing a new dataset of repeat sales. In this dataset, some of the items that were

<sup>12</sup> [Kahneman and Tversky \(1979\)](#) proposed a theory of decision making under uncertainty. One of the components of "prospect theory" is that individuals exhibit riskier behavior when holding losses than when holding gains.

Table 4  
Role of estimates

Author	Sample	Period	Result
Milgrom and Weber (1982a)			Honesty is the best policy
Abowd and Ashenfelter (1988)	Impressionist Art	1980–1982	Auctioneer's price estimates are far better predictors of prices than hedonic models
Ashenfelter (1989)	Impressionist Art	1980–1982	Auction houses are truthful
Chanel, Gérard-Varet and Vincent (1996)	Jewelry	1993–1994	Pre-sale estimates undervalue most types of jewelry, with the exception of some watches
Beggs and Graddy (1997)	Impressionist Art Contemporary Art	1980–1991 1980–1994	Systematic over and under valuations (recently executed works of art tend to be overvalued, longer and wider paintings are undervalued)
Bauwens and Ginsburgh (2000)	English Silver	1976–1990	English silver: Christie's systematically underestimates Sotheby's overvalues inexpensive pieces and undervalues expensive pieces
Ashenfelter, Graddy and Stevens (2002)	Impressionist Art Contemporary Art	1980–1991 1982–1994	Examines whether spread between high and low estimate is indication of auctioneer's uncertainty or reflects seller's wish to set a high reserve price
Mei and Moses (2002)	American, Impressionist and Old Masters	1973–2002	Higher priced paintings are systematically overvalued
Beggs and Graddy (2004)	Impressionist Art Contemporary Art	1980–1994	Anchoring plays a significant role in estimate formation

sold twice appeared at auction between repeat sales and failed to sell, whereas others did not appear at auction (and thus could not fail) between the two repeat sales. They then estimate a variation of Equation (3) above,

$$r_i = \ln\left(\frac{P_{i,s}}{P_{i,b}}\right) = \sum_{t=b_i+1}^{s_i} \mu_t + \beta \text{fail} + \sum_{t=b_i+1}^{s_i} v_t. \quad (5)$$

$r_i$  is the return for painting  $i$  between the date of purchase ( $b$ ) and the date of sale ( $s$ ).  $\mu_t$  is the average return for all paintings in period  $t$ , fail is a dummy variable equal to 1 if a paintings fails and 0 if not, and  $v_t$  is an error term.

The return is thus conditioned on whether or not a painting appeared at auction and failed. Their regressions indicate items that fail to sell at auction end up returning about 30 percent less than other paintings.

A natural question to ask is what actually drives the burning effect. Or more specifically, is the burning effect causal, in that a painting failing to sell causes the price to fall, or is failing simply correlated with a fall in price? [Beggs and Graddy \(2006\)](#) show within a general auctions model of affiliated values that there can be a variety of reasons why “burning” occurs including trend effects, reserve price effects and of course common values. Trend effects occur when a lower price is observed after the painting fails at auction simply because that failure is correlated with a downward trend in common tastes for a specific painting. Reserve price effects could occur if a seller lowers the reserve price after a failure because, for example, of an urgent need to sell. Common value effects occur if failing causes a drop in price, because bidders take into account other bidder’s valuations of the painting.

[Beggs and Graddy \(2006\)](#) also find that failing has a significantly negative effect on the pre-sale estimates, indicating that auctioneers anticipate the effect of failing on price. This is not surprising, given the widespread belief that failed items are “burned”.

### 3.4. Sales rates and reserve prices

As we noted above, items that are put up for sale at auction often go unsold because the bidding in the auction does not meet the reserve price. Sale rates vary tremendously over time and they also vary systematically across different types of auctions. [Table 5](#) shows sale rates in different departments at Christie’s in London in 1995 and 1996 along with average value of a lot sold. As can be seen from the table, 96 percent of items put up for sale in auctions of arms and armor were sold, 89 percent of wine at auction was sold, and 71 percent of impressionist and modern art items were sold.

[Ashenfelter, Graddy and Stevens \(2002\)](#) provide a study of sale rates through time and across different types of auctions. Based on the observation that an item is bought-in if and only if it does not meet or exceed its reserve price, they develop a model of optimal reserve prices. The seller of a painting faces the following problem: if he participates in an auction the highest bid for the painting can be regarded as a random draw from some price distribution. When a seller sets a reserve price, he must decide at what price he would be indifferent between selling now and waiting for the next auction. The optimal policy is to set a reserve price that is a constant proportion of the current expected price. Sale rates can then be modeled as being explained by price shocks and a constant, or “natural sale rate”. This natural sale rate (which may vary across different types of auctions) depends only on the variance of log prices and the seller’s discount rate. They estimate that the reserve price is generally set to be about 70 to 80 percent of the auctioneer’s low estimate. Although reserve prices are generally secret, the available evidence suggests that this prediction is reasonably accurate.

Table 5  
Average sale rates by department

Department	Average sold lot value		No. of auctions in sample	Sale rate (% of lots sold)		% Sold by value	
	1996	1995		Mean	Std. dev.	Mean	Std. dev.
Impressionist	£122,820	£135,430	8	71%	(0.11)	80%	(0.10)
Old Masters Drawings	£50,670	£29,210	4	77%	(0.09)	89%	(0.08)
Contemporary	£36,820	£36,840	7	79%	(0.04)	87%	(0.06)
British Pictures	£29,710	£23,560	7	78%	(0.14)	83%	(0.17)
Old Master Pictures	£29,180	£6560	11	73%	(0.15)	82%	(0.15)
Continental Pictures	£21,810	£10,450	7	72%	(0.11)	79%	(0.10)
Clocks	£14,340	£5130	4	88%	(0.03)	89%	(0.07)
Jewellery	£12,190	£6750	8	86%	(0.05)	89%	(0.04)
Furniture	£11,670	£8220	25	85%	(0.09)	92%	(0.06)
Silver	£11,080	£5910	10	87%	(0.11)	92%	(0.07)
Sculpture	£11,070	£6340	5	78%	(0.21)	81%	(0.20)
Modern British Pictures	£10,340	£7190	9	70%	(0.05)	81%	(0.05)
Victorian Pictures	£9460	£8400	6	66%	(0.13)	75%	(0.11)
British Drawings & Watercolours	£9160	£3400	14	72%	(0.14)	87%	(0.10)
Rugs & Carpets	£9160	£3700	8	80%	(0.17)	85%	(0.14)
Topographical Pictures	£8640	£8010	2	68%	(0.13)	81%	(0.00)
Islamic	£6670	£6950	5	68%	(0.22)	82%	(0.12)
Cars	£5750	£7610	6	71%	(0.16)	65%	(0.22)
Chinese Works of Art	£5640	£6400	8	70%	(0.19)	79%	(0.16)
Books & Manuscripts	£5220	£4270	15	81%	(0.12)	86%	(0.09)
Russian Works of Art	£4490	£5480	4	64%	(0.14)	69%	(0.15)
Japanese	£4410	£2840	5	72%	(0.04)	76%	(0.05)
Musical Instruments	£3960	£4110	5	77%	(0.05)	76%	(0.16)
Watches	£3870	£2190	6	71%	(0.09)	81%	(0.11)
Prints-Old Modern and Contemporary	£3850	£4230	8	81%	(0.12)	92%	(0.09)
Miniatures	£3350	£3260	2	82%	(0.05)	92%	(0.07)
Antiquities	£3260	£3640	3	57%	(0.08)	66%	(0.13)
Porcelain and Glass	£2700	£2600	14	76%	(0.12)	85%	(0.10)
Tribal Art	£2650	£2090	3	67%	(0.08)	75%	(0.19)
Photographica	£2580	£1660	3	61%	(0.27)	79%	(0.08)
Modern Guns	£2510	£3620	5	93%	(0.06)	94%	(0.04)
Garden Statuary	£2120	£1540	4	91%	(0.10)	91%	(0.11)
Arms & Armour	£1890	£2400	4	96%	(0.03)	99%	(0.01)
Frames	£1800	£2260	4	81%	(0.15)	85%	(0.14)
Stamps	£830	£650	22	78%	(0.13)	82%	(0.12)
Wine	£690	£580	37	89%	(0.09)	91%	(0.08)

Genesove (1995) tests a related theory in the context of wholesale automobile auctions. He finds that on average sale rates in used auto auctions are actually quite low; between about 58 percent and 68 percent of automobiles go unsold. In his paper, he



finds that an increase in the variance of price is associated with a lower probability of sale, and hence the “natural sale rate” is again dependent on the variance of prices.

A related question is whether the reserve prices that are set in art auctions are optimal. This question has not been looked at in the context of art auctions, but McAfee, Quan and Vincent (2000) derive a lower bound on the optimal reserve price for a general auction model with affiliated signals, common components to valuations, and endogenous entry (all characteristics which can be applied to art auctions or other auctions of cultural objects). They apply their computations to FDIC real estate auctions and find that the lower bound on the optimal reserve price for real estate to be about 75 percent of the appraised value.

Overall, there has been little research into why sales rates differ between items and whether reserve prices are optimal. Given the persistent differences in sale rates between items (which suggests differing reserve prices), more research in these areas would be useful.

### 3.5. *Why secret reserve prices?*

In almost all auctions of cultural items, not only are there reserve prices, but these reserve prices are secret. Auctioneers generally do not reveal the reserve price and they make it as difficult as they can for bidders to infer it. A reserve price clearly contains information about the seller’s valuation of an item; intuitively, revealing information matters if the items contain a common value component among buyers.

While people buy art for enjoyment, there is an investment component to many buyers’ motives; that investment component leads one to classify art as having common-value components. Thus, the fact that auctioneers tend to keep reserve prices secret has remained a puzzle since the publication of Milgrom and Weber’s (1982a) paper, where it was shown that it is optimal for a seller of a good at a common-value auction to reveal their valuation.

One reason that has been suggested for secret reserve prices is that these may be used to deter collusion. As Ashenfelter (1989) suggests, when the turnout is low, some sellers may prefer that their goods be bought in and offered for sale at a later date rather than risk a collusive ring bidding to depress the item’s price. If there is a ring operating, a secret reserve price might encourage bidders to bid higher than they would have otherwise.

Vincent (1995) has cleverly built upon (and overturned) the intuition from Milgrom and Weber’s (1982a) original result. His explanation is based upon the inhibiting effect that the announcement of a reserve price may have on the participation of bidders in a given auction. This announcement could discourage some bidders from participating. As revelation of information is very important for increasing revenues in a common value auction, the fact that these bidders are not participating prevents their information from playing a part in the auction and may lower overall bids. Hence, there is a trade-off between the reserve price revealing the seller’s information, and a reserve price discouraging participating which lowers total aggregation of information.

Table 6  
Secret reserve prices

Author	Result
Milgrom and Weber (1982a, 1982b)	Optimal for a seller of a good at a common-value auction to reveal valuation
Ashenfelter (1989)	Secret reserve prices deter collusion
Vincent (1995)	Announced reserve prices deter participants This deterrence could lower overall bids in a common-value auction
Horstmann and LaCasse (1997)	A secret reserve price could delay sale, allowing truthful information to be revealed over time

Horstmann and LaCasse (1997) provide yet another reason for secret reserve prices. If the seller in a common-value auction possesses information that cannot be directly transmitted to the buyers, then a seller can either attempt to signal his information via a reserve price announcement, or choose a secret reserve price. A secret reserve price could result in delay in sale. If the true information about an item is revealed over time, the delay in sale could be profitable for high value items, but costly for low value items, and hence sellers of high-value items will not be tempted to mimic sellers of low-value items. While this explanation may appear reasonable for oil leases, it appears less applicable to art auctions, though sellers of art are given access to professional valuation services provided by the auction house, which may provide an asymmetry of information. A summary of papers relating to secret reserve prices is presented in Table 6.

### 3.6. *The declining price anomaly*

Since Ashenfelter (1989) showed that prices are twice as likely to decrease as to increase for identical bottles of wine sold in same lot sizes at auction, there has been a tremendous amount of study of the declining price anomaly in many types of auctions. The literature on the declining price anomaly is extensive. Pesando and Shum (1996) and Beggs and Graddy (1997) address declining prices in art auctions, and much of the other literature is applicable to art. We review this literature below. We begin with a brief review of the theoretical literature on why declining prices may occur, and then review the types of auctions where declining prices have been found.

Soon after publication of Ashenfelter's (1989) article, there were many theoretical papers written to explain declining prices. Black and de Meza (1992) claimed it was no anomaly; declining prices in wine auctions exist primarily because the winner of the first auction in a sequence has the option to buy the remaining objects at the winning price. However, this theory is unable to explain why the anomaly continues to exist even where this option is not permitted. McAfee and Vincent (1993) showed that risk

aversion could create declining prices. One unappealing feature of their explanation is that a pure-strategy equilibrium exists only when there is nondecreasing absolute risk aversion, which is usually thought implausible. Mixed strategy equilibria are ex-post inefficient, which is sometimes also thought to be weakness of this theory, but which may nevertheless be a correct characterization of the actual market. [Beggs and Graddy \(1997\)](#) attribute a declining price to pre-sale estimate ratio throughout an auction to the fact that the value of art auctioned (as measured by the pre-sale estimate), on average, declines throughout the auction.

Other theoretical papers are as follows. [Von der Fehr \(1994\)](#) shows that participation costs could create declining prices through strategic bidding. [Engelbrecht-Wiggans \(1994\)](#), [Bernhardt and Scoones \(1994\)](#), [Gale and Hausch \(1994\)](#), and [Pezanis-Christou \(2001\)](#) relate the price decline to heterogeneity among buyers, and [Ginsburgh \(1998\)](#) shows that the presence of absentee bidders can generate declining prices.

The declining price anomaly has been documented in a number of different types of auctions with different auction structures. [Buccola \(1982\)](#) found it occurring in livestock auctions, [Milgrom and Weber \(1982b\)](#) for transponder leases, [McAfee and Vincent \(1993\)](#) and [di Vittorio and Ginsburgh \(1994\)](#) confirmed [Ashenfelter's \(1989\)](#) wine findings, [Thiel and Petry \(1995\)](#) in stamp auctions, [Ashenfelter and Genesove \(1992\)](#) and [Vanderporten \(1992a, 1992b\)](#) for condominiums, [Engelbrecht-Wiggans and Kahn \(1992\)](#) for dairy cattle, [Lusht \(1994\)](#) for commercial real estate, [Chanel, Gérard-Varet and Vincent \(1996\)](#) for gold jewellery; [Pesando and Shum \(1996\)](#) for Picasso prints; [Thurston \(1997\)](#) for mink pelts, [Pezanis-Christou \(2001\)](#) for fish auctions, [van den Berg, van Ours and Pradhan \(2001\)](#) for Dutch flower auctions, and [Ginsburgh and van Ours \(2007\)](#) for Chinese porcelain recovered from shipwrecks. [Burns \(1985\)](#) and [Keser and Olson \(1996\)](#) have set up experiments that have reached the same conclusions.

Several authors have also found increasing prices. Among them are [Gandal \(1997\)](#) for Israeli cable television licenses, and [Donald, Paarsch and Robert \(1997\)](#) for Siberian timber-export permits. [Jones, Menezes and Vella \(1996\)](#) found that prices could increase or decrease in sequential auctions of wool, as did [Chanel, Gérard-Varet and Vincent \(1996\)](#) for watches; [Milgrom and Weber \(1982b\)](#) show theoretically that if bidders' valuations are affiliated, then prices will tend to rise over time. [Deltas and Kosmopoulou \(2001\)](#) find in a sale of library books that expected prices increase over the auction, but that probability of sale decreases. They attribute their findings to "catalogue" effects: it is important how and where and item appears in the pre-sale catalogue. [Natzkoff \(2001\)](#) provides an excellent survey of papers on the declining price anomaly.

It is an interesting result that in a variety of different types of auctions, price direction throughout an auction can be predicted. As reported in [Table 7](#), declining prices (on average) have been documented in more types of auctions than have rising prices. A variety of economic theories have been developed to explain price direction, and in all likelihood, the price direction results from a combination of these effects. Declining prices do not occur in every auction or every art auction, but they appear to be an important effect that the auction mechanism has on price.

Table 7  
Declining price anomaly

---

<i>Empirical Work (Declining Prices)</i>	
Buccola (1982)	Livestock
Burns (1985)	Experimental results
Ashenfelter (1989)	Wine
Milgrom and Weber (1982b)	Transponder leases
Thiel and Petry (1995)	Stamps
Ashenfelter and Genesove (1992)	Condominiums
Vanderporten (1992a, 1992b)	Condominiums
Engelbrecht-Wiggans and Kahn (1992)	Dairy cattle
McAfee and Vincent (1993)	Wine
Di Vittorio and Ginsburgh (1994)	Wine
Lusht (1994)	Commercial real estate
Chanel, Gérard-Varet and Vincent (1996)	Gold jewelry
Pesando and Shum (1996)	Picasso prints
Keser and Olson (1996)	Experimental results
Beggs and Graddy (1997)	Art
Thurston (1997)	Mink pelts
Pezanis-Christou (2001)	Fish
Van den Berg, van Ours and Pradhan (2001)	Flowers
Ginsburgh and van Ours (2007)	Chinese porcelain from shipwrecks
<i>Empirical Work (Increasing Prices)</i>	
Jones, Menezes and Vella (1996)	Wool auctions
Chanel, Gérard-Varet and Vincent (1996)	Watches
Gandal (1997)	Israeli cable television auctions
Donald, Paarsch and Robert (1997)	Siberian timber auctions
Deltas and Kosmopoulou (2001)	Library books
<i>Theoretical Work</i>	
Black and de Meza (1992)	Declining prices in wine auctions are due to buyers' options
McAfee and Vincent (1993)	Risk aversion could create declining prices
von der Fehr (1994)	Participation costs could create declining prices
Engelbrecht-Wiggans (1994)	Relate price decline to heterogeneity of objects
Bernhardt and Scoones (1994)	Relate price decline to heterogeneity of objects
Gale and Hausch (1994)	Relate price decline to heterogeneity of objects
Beggs and Graddy (1997)	Ordering by value can generate price/estimate declines
Ginsburgh (1998)	Absentee bidders can generate declining prices

---

An important corollary to the declining price debate is the extent to which auction houses, given the choice, should organize their sales differently. [Ginsburgh and van Ours \(2007\)](#) look at the results of three auctions of shipwreck findings where different setups were used. They find that parcels containing lots with the same number of items (rather than increasing or decreasing the number of items) achieve greater prices relative to the pre-sale estimate than parcels that were organized differently. They conjecture that organizing an auction with the same number of items in a parcel is revenue maximizing.

#### 4. Conclusion

The empirical study of art auctions really has two purposes. On the one hand, the auction mechanism provides a very public report on the prices of art objects. As we have shown, because of the unique nature of many art objects, the interpretation of market prices requires great care. Nevertheless, this information is the primary way that high-quality art objects are valued and it provides us with our primary objective information on preferences regarding art. Although the market is surely not all that is important in the judgment of art and artists, it is certainly one of the key components of our understanding of what is good and bad.

The empirical study of art auctions also has another purpose. Art auctions provide data that may be used to test and refine strategic models of behavior. Here the object of study is the economic mechanism and it makes very little difference what object is for sale. It appears that a great deal of what we know about the operation of auction mechanisms may also lead to the rather happy study of objects of considerable interest in their own right.

The empirical study of art auctions and the price of art assets has been a growth field in the last decade and has resulted in an increasing sophistication in the questions being asked and in the empirical methods being used. It seems likely that this trend will continue into the future.

#### Appendix A

The dataset on Impressionist and Modern Art auctions was constructed by Orley Ashenfelter and Andrew Richardson. This dataset is restricted to 58 selected Impressionist and Modern Artists and includes only paintings, not sculptures. These artists were chosen primarily because their work is well represented at auction. The period covered is 1980 to 1990, and the dataset includes over 16,000 items in 150 auctions that were held in London and New York at both Christie's and Sotheby's. The auction prices were collected from public price lists, and the estimated prices and observable painting characteristics were collected from the pre-sale catalogues. This dataset does not include all items sold in each auction, only a sample of the 58 selected artists. Furthermore, we only have prices for items that were sold at auction.

To construct a repeat-sales price index, we identified 230 paintings that sold at least twice during the period 1980–1990 (for a total of 474 observations). To make a positive identification, we required that paintings have an identical title, medium, artist and painting date. As many paintings have identical titles, title and artist are not sufficient identifiers. We regress the log of the sale price of the painting on a dummy variable for each painting, and the time period in which the painting was sold. We include a dummy variable for each year. Using the antilogs of the coefficients on the time dummies, we construct our repeat sales price index for Impressionist Art as reported in [Figure 1](#).

For the hedonic price index in Figure 1, the log of the sale price is regressed on the hedonic painting characteristics in addition to time dummies for each period. The hedonic characteristics used for Impressionist and Modern Art are painting date, length, width, signed, monogrammed, stamped, medium in which it was painted, and artist. We also include dummy variables indicating whether the painting was sold at Sotheby's or Christie's and New York or London.

## References

- Abowd, J., Ashenfelter, O. (1988). "Art auctions: Price indices and sale rates for impressionist and contemporary pictures". Mimeo, Department of Economics, Princeton University.
- Anderson, R.C. (1974). "Paintings as an investment". *Economic Inquiry* 12, 13–25.
- Ashenfelter, O. (1989). "How auctions work for wine and art". *Journal of Economic Perspectives* 3, 23–36.
- Ashenfelter, O. (2000). "The demand for expert opinion: Bordeaux wine". *Les Cahiers de L'OCVE, Cahier Scientifique* 3, Mars.
- Ashenfelter, O., Genesove, D. (1992). "Testing for price anomalies in real-estate auctions". *American Economic Review* 82, 501–505.
- Ashenfelter, O., Graddy, K. (2003). "Auctions and the price of art". *Journal of Economic Literature* 41, 763–787.
- Ashenfelter, O., Graddy, K. (2005). "Anatomy of the rise and fall of a price-fixing conspiracy: Auctions at Sotheby's and Christie's". *Journal of Competition Law and Economics* 1, 3–20.
- Ashenfelter, O., Graddy, K., Stevens, M. (2002). "A study of sale rates and prices in impressionist and contemporary art auctions". Unpublished manuscript.
- Ashenfelter, O., Graddy, K., Ginsburgh, V., Legros, P., Sahuguet, N. (2003). "The class actions against Christie's and Sotheby's, or a buyer's guide on how to win twice at auction". *Art Newspaper*, November, 141, 42.
- Bailey, M.J., Muth, R.F., Nourse, H.O. (1963). "A regression method for real estate price index construction". *Journal of the American Statistical Association* 58, 933–942.
- Baumol, W.J. (1986). "Unnatural value: Or art investment as a floating crap game". *American Economic Review, Papers and Proceedings* 76, 10–14.
- Bauwens, L., Ginsburgh, V. (2000). "Art experts and auctions: Are pre-sale estimates unbiased and fully informative?" *Recherches Economiques de Louvain* 66, 131–144.
- Beggs, A., Graddy, K. (1997). "Declining values and the afternoon effect: Evidence from art auctions". *Rand Journal of Economics* 28, 544–565.
- Beggs, A., Graddy, K. (2004). "Testing for reference dependence: An application to the art market". Unpublished manuscript.
- Beggs, A., Graddy, K. (2006). "Failure to meet the reserve price: The impact on returns to art". Unpublished manuscript.
- Bernhardt, D., Scoones, D. (1994). "A note on sequential auctions". *American Economic Review* 84, 653–657.
- Biey, M.L., Zanola, R. (2004). "The market for Picasso prints: An hybrid model approach". Unpublished manuscript.
- Black, J., de Meza, D. (1992). "Systematic price differences between successive auctions are no anomaly". *Journal of Economics and Management Strategy* 1, 607–628.
- Buccola, S. (1982). "Price trends at livestock auctions". *American Journal of Agricultural Economics* 64, 63–69.
- Buelens, N., Ginsburgh, V. (1993). "Revisiting Baumol's art as floating crap game". *European Economic Review* 37, 1351–1371.

- Burns, P. (1985). "Experience and decision making: A comparison of students and businessmen in a simulated progressive auction". In: Smith, V. (Ed.), *Research in Experimental Economics: A Research Annual*, vol. 3. JAI Press, Greenwich, CT, pp. 139–157.
- Campbell, J. (1987). "Stock returns and the term structure". *Journal of Financial Economics* 18, 373–399.
- Carter Hill, R., Knight, J.R., Sirmans, C.F. (1997). "Estimating capital asset price indexes". *The Review of Economics and Statistics* 79, 226–233.
- Case, K.E., Quigley, T.M. (1991). "The dynamics of real estate prices". *Review of Economics and Statistics* 73, 50–58.
- Case, K.E., Shiller, R.J. (1987). "Prices of single-family homes since 1970: New indexes for four cities". *New England Economic Review*, September–October, 45–56.
- Chanel, O., Gérard-Varet, L., Ginsburgh, V. (1996). "The relevance of hedonic price indices". *Journal of Cultural Economics* 20, 1–24.
- Chanel, O., Gérard-Varet, L., Vincent, S. (1996). "Auction theory and practice: Evidence from the market for jewellery". In: Ginsburgh, V., Menger, P. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier, Amsterdam, pp. 135–149.
- Czujack, C. (1997). "Picasso paintings at auction, 1963–1994". *Journal of Cultural Economics* 21, 229–247.
- de la Barre, M., Docclo, S., Ginsburgh, V. (1996). "Returns of impressionist, modern and contemporary European paintings 1962–1991". *Annales d'économie et de Statistique* 35, 143–181.
- Deltas, G., Kosmopoulou, G. (2001). "Bidding in sequential auctions: 'Catalogue' vs. 'order of sale' effects (in a sale of library books)". Mimeo, Department of Economics, University of Illinois.
- de Marchi, N. (2004). "Auctioning paintings in late seventeenth-century London: Rules, segmentation and prices in an emergent market". In: Ginsburgh, V. (Ed.), *Economics of Art and Culture*. Elsevier, Amsterdam.
- di Vittorio, A., Ginsburgh, V. (1994). "Pricing red wines of Médoc vintages from 1949 to 1989 at Christie's auctions". Mimeo, Université Libre de Bruxelles.
- Donald, S., Paarsch, H., Robert, J. (1997). "Identification, estimation, and testing in empirical models of sequential, ascending-price auctions with multi-unit demand: An application to Siberian timber-export permits". Mimeo, Boston University, University of Iowa and Université de Montréal.
- Engelbrecht-Wiggans, R. (1994). "Sequential auctions of stochastically equivalent objects". *Economics Letters* 44, 87–90.
- Engelbrecht-Wiggans, R., Kahn, C. (1992). "An empirical analysis of dairy cattle auctions". Mimeo, University of Illinois.
- Frey, B.S., Eichenberger, R. (1995). "On the rate of return in the art market: Survey and evaluation". *European Economic Review* 39, 528–537.
- Frey, B.S., Pommerehne, W.W. (1989). *Muses and Markets, Explorations in the Economics of the Arts*. Blackwell, Oxford.
- Gale, I., Hausch, D. (1994). "Bottom-fishing and declining prices in sequential auctions". *Games and Economic Behavior* 7, 318–331.
- Galenson, D.W. (2000). "The careers of modern artists: Evidence from auctions of contemporary paintings". *Journal of Cultural Economics* 24, 87–112.
- Galenson, D.W. (2002). "Quantifying artistic success: Ranking French painters – and paintings – from Impressionism to Cubism". *Historical Methods* 35, 5–20.
- Galenson, D.W. (2004). "The life cycles of modern artists". *Historical Methods* 37, 123–136.
- Galenson, D.W., Weinberg, B.A. (2000). "Age and the quality of work: The case of modern American painters". *Journal of Political Economy* 108, 761–777.
- Galenson, D.W., Weinberg, B.A. (2001). "Creating modern art: The changing careers of painters in France from Impressionism to Cubism". *American Economic Review* 91, 1063–1071.
- Gandal, N. (1997). "Sequential auctions of interdependent objects: Israeli cable television licenses". *Journal of Industrial Economics* 45, 227–244.
- Genesove, D. (1995). "Search at wholesale auto auctions". *Quarterly Journal of Economics* 110, 23–49.
- Genesove, D., Mayer, C. (2001). "Loss aversion and seller behavior: Evidence from the housing market". *Quarterly Journal of Economics* 116, 1233–1260.

- Ginsburgh, V. (1998). "Absentee bidders and the declining price anomaly in wine auctions". *Journal of Political Economy* 106, 1302–1319.
- Ginsburgh, V., Jeanfils, P. (1995). "Long-term co-movements in international markets for paintings". *European Economic Review* 39, 538–548.
- Ginsburgh, V., Legros, P., Sahuguet, N. (2004). "How to win twice at an auction; On the incidence of commissions in auction markets". Unpublished manuscript.
- Ginsburgh, V., Mei, J., Moses, M. (2006). "The computation of prices indices". In: Ginsburgh, V.A., Throsby, D. (Eds.), *Handbook of the Economics of Art and Culture*, vol. 1, Elsevier, Amsterdam. Chapter 27 in this handbook.
- Ginsburgh, V., van Ours, J. (2007). "How to organize a sequential auction: Results of a natural experiment by Christie's". *Oxford Economic Papers*. In press.
- Goetzmann, W.N. (1992). "The accuracy of real estate indices: Repeat sale estimators". *Journal of Real Estate Finance and Economics* 5, 5–53.
- Goetzmann, W.N. (1993). "Accounting for taste: Art and financial markets over three centuries". *American Economic Review* 83, 1370–1376.
- Goetzmann, W.N. (1996). "How costly is the fall from fashion? Survivorship bias in the painting market". In: Ginsburgh, V., Menger, P. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier, Amsterdam, pp. 71–83.
- Goetzmann, W.N., Spiegel, M. (1997). "A spatial model of housing returns and neighborhood substitutability". *Journal of Real Estate Finance and Economics* 14, 11–31.
- Goetzmann, W.N., Spiegel, M. (2003). "Art market repeat sales indices based upon Gabrius S.P.A. data". *Gabrius White Paper*.
- Graeser, P. (1993). "Rate of return to investment in American antique furniture". *Southern Economic Journal* 59, 817–821.
- Horstmann, I., LaCasse, C. (1997). "Secret reserve prices in a bidding model with a resale option". *American Economic Review* 87, 663–684.
- Hosios, A., Pesando, J. (1991). "Measuring prices in resale housing markets in Canada: Evidence and implications". *Journal of Housing Economics* 1, 303–317.
- In Re Auction Houses Antitrust Litigation (2001). US Dist. LEXIS 1713 (S.D.N.Y.).
- Jones, C., Menezes, F., Vella, F. (1996). "Auction price anomalies: Evidence from wool auctions in Australia". Working Paper 303, Australian National University.
- Kahneman, D., Tversky, A. (1979). "Prospect theory: An analysis of decisions under risk". *Econometrica* 47, 263–291.
- Keser, C., Olson, M. (1996). "Experimental examination of the declining price anomaly". In: Ginsburgh, V., Menger, P. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier, Amsterdam, pp. 151–175.
- Klemperer, P. (2004). *Auctions: Theory and Practice*. Princeton University Press, Princeton, NJ.
- Landes, W. (2000). "Winning the art lottery: The economic returns to the Ganz collection". *Recherches Economiques de Louvain* 66, 111–130.
- Lusht, K. (1994). "Order and price in a sequential auction". *Journal of Real Estate Finance and Economics* 8, 259–266.
- McAfee, R.P., Quan, D., Vincent, D. (2000). "How to set minimum acceptable bids, with an application to real estate auctions". Unpublished manuscript.
- McAfee, R.P., Vincent, D. (1993). "The declining price anomaly". *Journal of Economic Theory* 60, 191–212.
- McAndrew, C., Thompson, R. (2004a). "Are presale art auction estimates unbiased? Some affirming evidence". Unpublished manuscript.
- McAndrew, C., Thompson, R. (2004b). "Presale estimates, risk analysis, and the loan quality of fine art". Unpublished manuscript.
- Mei, J., Moses, M. (2002). "Art as an investment and the underperformance of masterpieces". *American Economic Review* 92, 1656–1668.
- Mei, J., Moses, M. (2005). "Vested interest and biased price estimates: Evidence from an auction market". *Journal of Finance* 60, 2409–2435.



- Milgrom, P., Weber, R. (1982a). "A theory of auctions and competitive bidding". *Econometrica* 50, 1089–1122.
- Milgrom, P., Weber, R. (1982b). "A theory of auctions and competitive bidding, II". Mimeo, Stanford University and Northwestern University. Published with new foreword in: Klemperer, P. (Ed.), *The Economic Theory of Auctions*. Edward Elgar, Cheltenham, pp. 179–194.
- Natzkoff, P. (2001). "Sequential auctions: Survey of the literature and empirical analysis of auctions of raw wool in Australia". University of Oxford, M. Phil. thesis.
- Pesando, J.E. (1993). "Art as an investment. The market for modern prints". *American Economic Review* 83, 1075–1089.
- Pesando, J.E., Shum, P. (1996). "Price anomalies at auction: Evidence from the market for modern prints". In: Ginsburgh, V., Menger, P. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier, Amsterdam, pp. 113–136.
- Pezanis-Christou, P. (2001). "Sequential auctions with asymmetric buyers: Evidence from a fish market". Mimeo, Bonn University, Department of Economics.
- Pommerehne, W.W., Feld, L.P. (1997). "The impact of museum purchases on the auction prices of paintings". *Journal of Cultural Economics* 21, 249–271.
- Ross, M.H., Zondervan, S. (1989). "Capital gains and the rate of return on a Stradivarius". *Economic Inquiry* 27, 529–540.
- Schoenfeld, S. (2004). "The market for contemporary art". University of Oxford, MSc. thesis.
- Stein, J.P. (1977). "The monetary appreciation of paintings". *Journal of Political Economy* 85, 1021–1035.
- Stewart, J.B. (2001). "Bidding war: How an antitrust investigation into Christie's and Sotheby's became a race to see who could betray whom". *The New Yorker*, October 15, 2001, 158.
- Thiel, S., Petry, G. (1995). "Bidding behavior in second-price auctions: Rare stamp sales, 1923–1937". *Applied Economics* 27, 11–16.
- Thurston, N. (1997). "Explaining the declining price anomaly: Empirical evidence on buyer heterogeneity". Working Paper, Brigham Young University.
- van den Berg, G., van Ours, J., Pradhan, M. (2001). "The declining price anomaly in Dutch rose auctions". *American Economic Review* 91, 1055–1062.
- Vanderporten, B. (1992a). "Strategic behavior in pooled condominium auctions". *Journal of Urban Economics* 31, 123–137.
- Vanderporten, B. (1992b). "Timing of bids at pooled real estate auctions". *Journal of Real Estate Finance and Economics* 5, 255–267.
- Vincent, D. (1995). "Bidding off the wall: Why reserve prices may be kept secret". *Journal of Economic Theory* 65, 575–584.
- von der Fehr, N. (1994). "Predatory bidding in sequential auctions". *Oxford Economic Papers* 46, 345–356.
- Watson, P. (1992). *From Monet to Manhattan: The Rise of the Modern Art Market*. Random House, New York.

## THE COMPUTATION OF PRICES INDICES\*

VICTOR GINSBURGH

*ECARES, Université Libre de Bruxelles and CORE, Université catholique de Louvain, Belgium*

JIANPING MEI and MICHAEL MOSES

*Stern School of Business, New York University, USA*

### Contents

Abstract	948
Keywords	948
1. Introduction	949
2. Hedonic and repeat-sales estimators: A first approach	951
2.1. The hedonic estimator	952
2.2. The repeat-sales estimator	954
3. Hedonic, repeat-sales and other estimators: Further issues	957
3.1. Rosen's interpretation of hedonic models	957
3.2. Issues with hedonic indexing	958
3.3. Functional form of the hedonic function	959
3.4. Hedonic characteristics for artworks	960
3.5. Other repeat-sales estimators	963
3.6. Combining repeat-sales and hedonic estimators	963
3.7. Other estimators	964
4. Developing hedonic and repeat-sales models: Data acquisition	965
5. Comparing hedonic and repeat-sales estimation	966
5.1. The pros and the cons of both types of indices	966
5.2. The literature on comparing indices	967
5.3. Comparing hedonic and repeat-sales estimation: Some Monte-Carlo experimentation	968
5.4. Comparing three GLS estimators for repeat-sales estimation	973
Appendix A: Studies on returns	973
References	974

\* We are very grateful to Avi Giloni who worked out the computations discussed in Section 5, and to Orley Ashenfelter for his comments on the version presented at the Princeton Conference.

## Abstract

While there are no significant investment characteristics that inhibit art from being considered as an asset, a major hurdle has long been the lack of a systematic measure of its financial performance. Due to its heterogeneity (each piece is different) and its infrequency of trading (the exact same piece does not come to the market very often), the determination of changes in market value is difficult to ascertain. Two estimation methods are commonly used to construct indices. Repeat-sales regression (RSR) uses prices of individual objects traded at two distinct moments in time. If the characteristics of an object do not change (which is usually so for collectibles), the heterogeneity issue is bypassed. The basic idea of the hedonic regression (HR) method is to regress prices on various attributes of objects (dimensions, artist, subject matter, etc.) and to use the residuals of the regression which can be considered as “characteristic-free prices” to compute the price index. The chapter deals with the basics of hedonic and repeat-sales estimators, and tries to interpret in economic terms what both are trying to achieve. It also goes into some more technical details which may be useful for researchers who want to construct such indices, and gives some guidelines on how to go about collecting data, and the choice between RSR and HR that this induces. Both methods are compared using simulated returns, pointing to which method should be used given the data at hand.

## Keywords

prices indices, repeat sales, hedonic pricing, auctions

*JEL classification:* C13, C22, D44, Z11

## 1. Introduction

While there are no significant investment characteristics that inhibit art from being considered as an asset, a major hurdle has long been the lack of a systematic measure of its financial performance. Due to its heterogeneity (each piece is different) and its infrequency of trading (the exact same piece does not come to the market very often), the determination of changes in market value is difficult to ascertain.

Anderson (1974) and Stein (1977) were among the very first to study returns in art markets. But it is undoubtedly Baumol's (1986) surprising result about the very small long-term return of art (0.55 percent between 1600 and 1950), coupled with the price boom in the late 1980s, shortly after his paper was published, that triggered most of the literature. It became so large, that it is hard to find, quote and do justice to all studies.

Many different art or antique markets have been looked at: African art, books, vintage cars, ancient coins, Old Master drawings, furniture, crafted ivory objects, jewelry, photographs, prints, sculpture, silverware, stamps, and probably others.<sup>1</sup> Of course, the largest effort was devoted to portfolios of paintings, often Impressionists and their followers, but also American, Belgian, Canadian, Italian, Latin-American, Pre-Raphaelite, and other groups of painters. Some papers look at individual painters, some study individual collections.<sup>2</sup>

There are several important uses for a well-constructed art market index:

- (a) Outline general market trends, much like the Dow Jones Industrial Average describing the general direction of the US stock market. This would help us measure art market returns and compare it to other asset classes so that we can ascertain whether art is a viable investment asset class.
- (b) Provide a concise measure of art market volatility as well as its correlation with other financial instruments, such as stocks and bonds, making it possible to address the question whether investing in art would diversify risk in a long-term investment portfolio.
- (c) Allow to examine what are the major social and economic factors affecting art market price movements. It would be easy to examine, for example, how inflation affects art market prices.
- (d) Provide a simple way of appraising the value of artworks. By assuming artworks appreciate at the art market rate, just as assuming a US stock price moves at the rate of the S&P 500, one can mark-to-market and derive a simple valuation for artworks. Recent studies, such as Mei and Moses (2002b) have discovered that this simple valuation estimate can provide some anchor to art prices, much like book-value is a useful valuation tool for stocks.

There are several important properties that art indices must have in order to serve as investment benchmarks:

<sup>1</sup> As well as wines, which can also in some sense be considered as art.

<sup>2</sup> See Appendix A for details and references.

- (a) They must be based on reliable and publicly available price information (catalogues or computer databases) so that they can be independently replicated.
- (b) Methods used to construct indices must address the heterogeneity issue raised above and should be based on well-established economic models and estimation procedures that make the results transparent and easy to verify.
- (c) The data should avoid sample selection biases.
- (d) Indices should distinguish many different collecting categories, since returns may vary dramatically.
- (e) The data must be regularly updated to provide a real time investment guide; it may also allow an art portfolio to be “marked to market” on a timely basis.

We now briefly turn to each of these aspects. Most papers use auction prices, since these are easily available; very few are able to get hold of other data. Gallery prices are analyzed by [Candela and Scorcu \(2001\)](#), while the analysis of returns for stamps [[Feuillolay \(1996\)](#)] or coins [[Verbert \(1991\)](#)] is sometimes carried out on catalogues.

Two estimation methods are commonly used to construct indices. Repeat-sales regression (RSR) uses prices of individual objects traded at two distinct moments in time. If the characteristics of an object do not change (which is usually so for collectibles), the heterogeneity issue is bypassed. This approach, extensively used in real estate studies, is applied to art markets by [Anderson \(1974\)](#), [Baumol \(1986\)](#),<sup>3</sup> [Goetzmann \(1990, 1993, 1996\)](#), [Goetzmann and Spiegel \(2003\)](#), [Locatelli Biey and Zanola \(1999\)](#), [Mei and Moses \(2002a, 2002b\)](#), [Pesando \(1993\)](#), and [Pesando and Shum \(1996, 1999\)](#). The basic idea of the hedonic regression (HR) method is to regress prices on various attributes of objects (dimensions, artist, subject matter, etc.) and to use the residuals of the regression which can be considered as “characteristic-free prices” to compute the price index. The method was used by [Frey and Pommerehne \(1989a, 1989b\)](#), [Buelens and Ginsburgh \(1993\)](#), [Chanel et al. \(1993\)](#) and became very popular in the mid-1990s. The benefit of HR is that the index is constructed from all sales, not from an (often small) subset of the available transactions. One of its drawbacks is that it depends on the characteristics used to describe the objects, and on the functional form of the equation.

Auction prices may give a biased view, since auctions seem to represent 33 to 50 percent of the art market. Other biases are problematic with auction prices. The first is especially relevant for studies that use repeat-sales sales, since some transactions do not go through salesrooms. This point was raised by [Guerzoni \(1994\)](#), who shows that [Reitlinger's \(1961, 1971\)](#) data,<sup>4</sup> on which [Baumol \(1986\)](#), [Frey and Pommerehne \(1989a, 1989b\)](#), [Buelens and Ginsburgh \(1993\)](#), and others base their calculations miss some important transactions, and this may seriously alter rates of return. The other problem is that works that are bought-in at auction are generally not included, since there is no hammer price. [Ashenfelter, Graddy and Stevens \(2001\)](#), [Ekelund, Ressler and Watson \(1998\)](#), [Goetzmann and Peng \(2003\)](#), [Goetzmann and Spiegel \(2003\)](#), and [McAndrew](#)

<sup>3</sup> Note that Baumol does not construct an index. See below.

<sup>4</sup> This may of course affect all resale data, not only those collected by Reitlinger.

and Thompson (2003, 2004) analyze the consequences, and suggest some solutions, such as estimating the secret reserve prices, or using some percentage of the lower bound of the pre-sale estimate published nowadays by most salesrooms.<sup>5</sup>

As mentioned above, many categories of collectibles have been studied: African art, books, vintage cars, coins, Old Master drawings, furniture, ivory objects, jewelry, photographs, prints, sculpture, musical instruments, silver, stamps, etc. Some are reasonably easy to describe (drawings) some are less so (furniture, sculpture). Some are sold frequently, since there are many “copies” available (prints), some are seen very rarely at auction (paintings by Vermeer). This points to which type of method (RSR vs. HR) can be used with some success, given the shortcomings of each one, and in particular, their dependence on the number of observations, and the possibility to describe objects by characteristics.

Timeliness is another important issue. There are fortunately more and more databases that are updated on a regularly, if not daily, basis, such as Artprice.com (<http://www.artprice.com>), Hislop’s Art Sales Index (<http://www.art-sales-index.com/system/index.html>), Gabrius (<http://gabrius.com/>), Gordon’s On Line Art Mall (<http://www.onlineartmall.com/limited/lgordon/>) and probably some other. Still, for many types of collectibles (furniture, violins) this is not so. In some cases, prices are made available in annual catalogues (such as Mayer for paintings and drawings, Gordon’s for prints), but often it is necessary to go to individual sales catalogues of auctioneers to get hold of the data. This again has an influence on the estimation method used to construct price indices.

The paper is organized as follows. Section 2 deals with the basics of hedonic and repeat-sales estimators, and tries to interpret in economic terms what both are trying to achieve. Section 3 goes into some more technical details which may be useful for researchers who want to construct such indices. Section 4 briefly discusses on how to go about collecting data, and the choice between RSR and HR that this induces. Section 5 compares both methods using simulated returns, and tries to point to which method should be used given the data at hand.

## 2. Hedonic and repeat-sales estimators: A first approach

In this section, we describe possible estimators of price indices, obtained from observing a set of  $2N$  transactions related to  $i = 1, 2, \dots, N$  different objects, that are also described in terms of some attributes or characteristics. To simplify exposition, we assume that each object is transacted twice. The set of dates (say, years) is  $t = 0, 1, \dots, T$  and defines possible periods (period  $t$  goes from date  $t - 1$  to date  $t$ ) or market runs. There exist data on prices for each object during some (here, 2) periods, but not for all

<sup>5</sup> Mei and Moses (2005) show that the publication of pre-sale estimates since the mid-1970s has an influence on hammer prices.

objects in every period. A transaction of object  $i$  in period  $t$  is indexed by subscripts  $(i, t)$ .

The estimators considered are the hedonic and the repeat-sales estimators, and include some others, that can be obtained as special cases. In particular, we show that the repeat-sales estimator is a special case of the hedonic estimator.

We illustrate our discussion using an example in which there are  $2N = 12$  sales of  $N = 6$  objects at three possible dates ( $T = 2$ ). We denote by  $p_{it}$  log of the price  $P_{it}$  of object  $i$ , sold at date  $t$ , and assume objects 1 and 4 were sold in  $t = 0$  and 1, objects 3 and 5 were sold in  $t = 1$  and 2 and finally, objects 2 and 6 were sold in  $t = 0$  and 2. The logged price (column) vector is denoted by  $p = (p_{10}, p_{20}, p_{40}, p_{60}, p_{11}, p_{31}, p_{41}, p_{51}, p_{22}, p_{32}, p_{52}, p_{62})$ . For convenience, we rank the observations for  $t = 0$  first, then those for  $t = 1$ , etc., without taking into account that some of the prices concern resales of the same object. We also define a vector  $y = (p_{11} - p_{10}, p_{22} - p_{20}, p_{32} - p_{31}, p_{41} - p_{40}, p_{52} - p_{51}, p_{62} - p_{60})$  with elements  $y_i$  (the logged differences of prices obtained at two dates for object  $i$ ).

### 2.1. The hedonic estimator

Hedonic regression is typically used to control for the changing quality of goods transacted. It seems to have been introduced by Court (1939), and subsequently used by Griliches (1961), Triplett (1969), and Griliches (1971a, 1971b) to construct price indices for automobiles. It is widely used by statistical offices to estimate consumer price indices,<sup>6</sup> and in academic work to establish price indices for real estate, computers, automobiles, dishwashers, the arts, and more generally for cases in which quality varies over space and/or time.

We start, however, with the simple case that ignores characteristics. Let  $C$  be a matrix consisting of  $2N$  rows and  $T + 1$  columns, denoted  $c_0, c_1, \dots, c_T$ . Element  $c_{it}$  is equal to *one* if a transaction on commodity  $i$  occurs in year  $t$ , and is *zero* otherwise. For the example at hand in which the (column) vector of prices is  $(p_{10}, p_{20}, p_{40}, p_{60}, p_{11}, p_{31}, p_{41}, p_{51}, p_{22}, p_{32}, p_{52}, p_{62})$ , the first column of  $C$  contains four ones, followed by 8 zeros, the second contains four zeros, followed by four ones and four zeros, etc.

We estimate the parameters  $\delta_t$  of the linear model:

$$p_{it} = \sum_{\tau=0}^T \delta_{\tau} c_{i\tau} + \varepsilon_{it}, \quad (1)$$

where  $\varepsilon_{it}$  is a random disturbance. The OLS estimator is:

$$\hat{\delta} = (C' C)^{-1} C' p. \quad (2)$$

<sup>6</sup> See, for example, Abraham, Greenless and Moulton (1998), Boskin (1996), Boskin et al. (1998) and of course the recent book by Triplett (2004) that is entirely devoted to this issue.

It is straightforward to check that the estimator for the price in year  $\tau$  is simply the average of the (log of) prices of the  $n_\tau$  objects sold during that year, that is a geometric average of prices:<sup>7</sup>

$$\hat{\delta}_\tau = \frac{1}{n_\tau} \sum_i p_{i\tau}, \quad \tau = 0, 1, \dots, T. \quad (3)$$

The index, normalized to 1 in  $\tau = 0$ , is obtained as the sequence of  $\exp \hat{\delta}_\tau / \exp \hat{\delta}_0$ .<sup>8,9</sup> Obviously, this is a sound approach as long as the mix of objects sold in each year has the same characteristics, or is of the same quality. This is often not the case, and then the hedonic approach is useful since it homogenizes sales mixes over time.

Consider therefore the set of objects sold in a specific year  $t$  and assume that the price of an object  $i$  sold in  $t$  can be considered as a function of  $m$  time-invariant characteristics  $v_{ik}$ ,  $k = 1, 2, \dots, m$  (e.g., the dimensions of a painting) and of  $n$  time-varying characteristics  $w_{ij\tau}$ ,  $\tau = 0, 1, \dots, t$  (e.g., the changing owners of a painting),  $j = 1, 2, \dots, n$ . If this assumption holds, we can write that  $p_{it} = f(v_{i1}, \dots, v_{im}, w_{i10}, \dots, w_{i1t}, w_{i20}, \dots)$ . We specialize the functional form to:<sup>10</sup>

$$p_{it} = \sum_{k=1}^m \alpha_k v_{ik} + \sum_{\tau=0}^t \sum_{j=1}^n \theta_{j\tau} w_{ij\tau} + \varepsilon_{it}. \quad (4)$$

The parameters  $\alpha_k$  and  $\theta_{j\tau}$  appearing in (4) are, often abusively,<sup>11</sup> interpreted as implicit prices of the various characteristics describing the commodity, and  $\varepsilon_{it}$  is a random error term. These implicit prices are thus obtained by a regression of prices on observable characteristics (also often in logarithmic form); once they are known, it is possible to compute, like in (3), the average price  $\hat{\delta}_\tau$  of a characteristic-free, or quality-adjusted commodity in year  $\tau$  as:

$$\hat{\delta}_\tau = \frac{1}{n_\tau} \sum_i \left( p_{i\tau} - \sum_{k=1}^m \alpha_k v_{ik} - \sum_{t=0}^T \sum_{j=1}^n \theta_{jt} w_{ijt} \right). \quad (5)$$

The sequence of  $\hat{\delta}_\tau$ ,  $\tau = 0, \dots, T$ , would then describe the price of an (artificial) characteristic-free commodity over time, and can obviously be obtained by a hedonic

<sup>7</sup> It is obvious that one can also derive an estimator for arithmetic averages of prices, using  $P_{it}$  instead of  $p_{it}$ .

<sup>8</sup> Alternatively, one can set  $c_0$  as a vector of ones, and estimate Equation (1) with an intercept. Then the sequence of  $\exp \hat{\delta}_\tau$  gives directly the index.

<sup>9</sup> Note that the antilog of the OLS estimates of the  $\delta$  are not unbiased, and that a correction has to be made, by adding one half of the coefficient's squared standard error to the estimated coefficient. Since such standard errors are usually small, this makes little difference. See Triplett (2004, Chapter 3, footnote 12).

<sup>10</sup> See more on the choice of functional forms, e.g., in Halvorsen and Pollakowski (1981).

<sup>11</sup> See the discussion below.



regression pooling the sales over time, by combining (1) and (4):

$$p_{it} = \sum_{k=1}^m \alpha_k v_{ik} + \sum_{\tau=0}^T \sum_{j=1}^n \theta_{j\tau} w_{ij\tau} + \sum_{\tau=0}^T \delta_{\tau} c_{i\tau} + \varepsilon_{it}.^{12} \quad (6)$$

The method allows for interactions between time and characteristics, if one believes that the implicit prices of some characteristics vary over time. For this, one merely has to introduce new variables  $\omega_{kt} = v_k c_t$ , which pick regression coefficients that describe the time path of the implicit price of characteristic  $k$ . The two previous estimators can also provide such information, by computing the parameters on sub-samples (e.g., a specific painter). However, given that the number of observations will be small compared with the total number of sales, the estimated coefficients will be estimated with little precision.

Obviously, there are many other ways to specify how prices depend on time. The  $\sum_{\tau} \delta_{\tau} c_{\tau}$  formulation makes it possible to construct a price index, in a reasonably flexible way. Alternatively, one can use a variable  $\tau$  which takes the values  $0, 1, 2, \dots, T$  and specify (6) with a term  $\phi\tau$ , where  $\phi$  would be an estimate of the price trend. One can also estimate time trends for sub-periods.

## 2.2. The repeat-sales estimator

The repeat-sales estimator was developed to derive price indexes for real estate. The method was formalized by Bailey, Muth and Nourse (1963), though they refer to previous suggestions by Wenzlick (1952) and Wyngarden (1927). The Bailey, Muth and Nourse paper was followed by a large number of theoretical approaches as well as applications, in most of the field journals.<sup>13</sup> This estimator is now also used for other markets, including artworks.

The usual approach to derive the underpinnings of the estimator is as follows. Assume that  $r_{it}$ , the continuously compounded return for a certain art asset  $i$  in period  $t$  may be represented by  $\delta_t$ , the continuously compounded return of “art”, and an error term  $\eta_{it}$ :

$$r_{it} = \delta_t + \eta_{it},$$

where  $\delta_t$  may be thought of as the average return in period  $t$  of artworks in the portfolio. We will use sales data about individual works to estimate the index  $\delta$  (a  $T$ -dimensional vector whose individual elements are  $\delta_t$ ) over some interval  $t = 1, \dots, T$ . The observed data consist of purchase and sales (logged) price pairs  $(p_{ib}, p_{is})$  of individual objects  $i$ ,

<sup>12</sup> Note that, following the usual specification, one can introduce an intercept  $\alpha_0$ , and estimate only  $T - 1$  parameters  $\delta_{\tau}$ , and normalize  $\delta_0$  to be equal to 1.

<sup>13</sup> AREUEA Journal, FRBSF Review, Journal of Environmental Planning and Management, Journal of Housing Economics, Journal of Real Estate Finance and Economics, Journal of Urban Economics, Land Economics, as well as in the Review of Economics and Statistics.

as well as the dates of purchase and sale, designated by  $b_i$  and  $s_i$ . Then, the logged relative price for object  $i$ , held between its purchase date  $b_i$  and its sales date  $s_i$ , may be expressed as:

$$r_i = p_{is} - p_{ib} = \sum_{t=b_i+1}^{s_i} r_{it}.$$

There are many ways to derive the repeat-sales estimator to understand how the regression should be run.<sup>14</sup> An easy way, which also happens to provide a link between the RSR and the HR estimators is to start with the hedonic equation, in which characteristics are constant over time:

$$p_{it} = \sum_{k=1}^m \alpha_k v_{ik} + \sum_{\tau=0}^T \delta_\tau c_{i\tau} + \varepsilon_{it}. \quad (7)$$

For notational purposes, it is convenient to define, as above,  $p_{ib}$  as the first sale, and  $p_{is}$  as the second one, and to redefine accordingly the columns of the matrix  $C$  and the  $\delta_\tau$  parameters. Then,

$$p_{is} - p_{ib} = \delta_s c_{is} - \delta_b c_{ib} + \eta_i, \quad (8)$$

where the  $\sum_k \alpha_k v_{ik}$  terms vanish, since the characteristics are identical over time, and  $\eta_i = \varepsilon_{is} - \varepsilon_{ib}$ . It is easy to check that for our example with six repeat-sales, this leads to the following system of equations:

$$\begin{bmatrix} p_{11} - p_{10} \\ p_{22} - p_{20} \\ p_{32} - p_{31} \\ p_{41} - p_{40} \\ p_{52} - p_{51} \\ p_{62} - p_{60} \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \delta_0 \\ \delta_1 \\ \delta_2 \end{bmatrix} + \eta,$$

where  $\eta$  is a vector of the six error terms  $\eta_i$ .

The three columns of the matrix that appears in the right-hand side are linearly dependent, so that one column, say the first, should be discarded. This will set the normalization  $\delta_0 = 0$ , and leave us with an  $N \times T$  ( $= 6 \times 2$ ) matrix  $Z$  that will be interpreted later on. The  $T$  (here 2) regression coefficients can be estimated by OLS:

$$\hat{\delta} = (Z'Z)^{-1} Z'y. \quad (9)$$

An alternative way to derive the estimator – which leads to an easy interpretation of the coefficients – is to construct an  $N \times T$  matrix  $X$ , the columns of which represent periods (not dates); observation  $i$  is in row  $i$ , which contains *ones* for periods during

<sup>14</sup> We derive the so-called geometric repeat-sales estimator. See Shiller (1991) and Goetzmann and Peng (2001) who discuss an arithmetic estimator.

which the object was held and *zeros* otherwise. For our example, this matrix is:

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

The OLS estimator of the two coefficients, say  $\beta_1$  and  $\beta_2$ , is given by

$$\hat{\beta} = (X'X)^{-1}X'y. \quad (10)$$

The two normal equations are

$$\begin{aligned} 4\hat{\beta}_1 + 2\hat{\beta}_2 &= (p_{11} - p_{10}) + (p_{22} - p_{20}) + (p_{41} - p_{40}) + (p_{62} - p_{60}), \\ 2\hat{\beta}_1 + 4\hat{\beta}_2 &= (p_{22} - p_{20}) + (p_{32} - p_{31}) + (p_{52} - p_{51}) + (p_{62} - p_{60}). \end{aligned}$$

They can also be written as

$$\begin{aligned} \hat{\beta}_1 &= \frac{1}{4}[(p_{11} - p_{10}) + ((p_{22} - \hat{\beta}_2) - p_{20}) + (p_{41} - p_{40}) \\ &\quad + ((p_{62} - \hat{\beta}_2) - p_{60})], \\ \hat{\beta}_2 &= \frac{1}{4}[(p_{22} - (p_{20} + \hat{\beta}_1)) + (p_{32} - p_{31}) + (p_{52} - p_{51}) \\ &\quad + (p_{62} - (p_{60} + \hat{\beta}_1))]. \end{aligned}$$

If we now interpret  $\hat{\beta}_1$  and  $\hat{\beta}_2$  as being estimates of the mean rates of return in periods 1 and 2, respectively,  $(p_{22} - \hat{\beta}_2)$  and  $(p_{62} - \hat{\beta}_2)$  are estimates of the prices of objects 2 and 6, had they been resold in year 1 instead of year 2, while  $(p_{20} + \hat{\beta}_1)$  and  $(p_{60} + \hat{\beta}_1)$  are estimates of the prices of the same objects, had they been sold for the first time in year 1 instead of year 0. Once this interpretation is accepted, one immediately sees that  $\hat{\beta}_1$  is the average return of the objects sold in  $t = 0$  and in  $t = 1$ , while  $\hat{\beta}_2$  is the average return of the objects sold in  $t = 1$  and in  $t = 2$ .

It is straightforward to link formulations (9) and (10). Let  $B$  be a  $T \times T$  matrix constructed as follows: row  $t$  starts with  $t$  ones, while the other elements of the row are zeros. For our example, this matrix is:

$$B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

We then construct a matrix  $XB^{-1}$  of explanatory variables, which is exactly the matrix  $Z$  used in (9). Some straightforward matrix algebra shows that (9) can also be written

$$\hat{\delta} = B\hat{\beta}, \quad (11)$$

which relates estimators (9) and (10), implying that

$$\hat{\delta}_t = \sum_{\tau=1}^t \hat{\beta}_\tau, \quad t = 1, 2, \dots, T.$$

For our example, this means that  $\hat{\delta}_1 = \hat{\beta}_1$  and  $\hat{\delta}_2 = \hat{\beta}_1 + \hat{\beta}_2$ . Since we can set  $\hat{\delta}_0 = \hat{\beta}_0 = 0$ , the sequence  $\exp(\hat{\delta}_0)$ ,  $\exp(\hat{\delta}_1)$ ,  $\exp(\hat{\delta}_2)$  is the price index over the three years.

A special case of the geometric repeat-sales estimator is the geometric mean estimator. Start with the following linear model:

$$y_i/\tau_i = \gamma + \varepsilon_i,$$

where  $\gamma$  is a parameter to be estimated and  $\tau_i$  is a variable which takes as value the number of periods during which an object was held by an owner (i.e. not sold); in the example,  $\tau_i$  is thus equal to 1 for  $i = 1, 3, 4$  and 5 and equal to 2 for  $i = 2$  and 6;  $\varepsilon_i$  is a random disturbance with the usual properties. The variable  $y_i/\tau_i$  is the annualized rate of return on commodity  $i$ . The parameter  $\gamma$  can be estimated by running a regression of  $y/\tau$  on a variable which takes the value one for each observation. It is trivial to check that the OLS estimate for  $\gamma$  is the average of annualized returns:

$$\hat{\gamma} = \frac{1}{N} \sum_i y_i/\tau_i. \quad (12)$$

For our example, (12) leads to:

$$\hat{\beta} = \frac{1}{6} [(p_{11} - p_{10}) + (p_{22} - p_{20})/2 + (p_{32} - p_{31}) + (p_{41} - p_{40}) + (p_{52} - p_{51}) + (p_{62} - p_{60})/2].$$

This is the estimator used by Baumol (1986) and Frey and Pommerehne (1989b).<sup>15</sup> It is obviously very easy to compute, but does not provide an index over time. Moreover, it puts equal weights on all annualized rates, irrespective of the length of time during which the object was held.

### 3. Hedonic, repeat-sales and other estimators: Further issues

#### 3.1. Rosen's interpretation of hedonic models

Lancaster (1966) was at the root of giving to the purely econometric technique used in hedonic regression, its theoretical foundations, based on the idea that commodities

<sup>15</sup> Actually, Baumol and Frey and Pommerehne use the exact formula  $(P_{it}/P_{it'})^{1/(t-t')}$  to compute the annual return of a work sold in  $t'$  and subsequently in  $t$ . We use the approximation instead. The two lead to comparable results if  $P_{it}$  is close to  $P_{it'}$ .

were not consumed per se, but for their combination of characteristics say, in the case of automobiles, speed, mileage, length, engine capacity, number of doors, etc., each of which carrying an implicit price. Rosen (1974) suggested that the simple estimation of such implicit prices by regressing observed prices on characteristics, was flawed, in the sense that, in general, the function does not allow “to recover the underlying utility and cost functions from such data alone” [Griliches (1990, p. 189)]. Implicit prices emerge from the equilibrium between demand and supply on markets for characteristics. Rosen outlines a two step procedure in which the first step is to estimate a hedonic function  $p = f(v)$ , to evaluate its derivatives at points corresponding to the observed values of  $v$ , and use these derivatives as prices in a system containing supply and demand functions for characteristics, paying attention to the usual identification problems in estimating simultaneous equations. Brown and Rosen (1982), Epple (1987), Bartik (1987) showed that the problem was even more complicated than what Rosen had thought, and suggest alternatives to Rosen’s estimation procedure. Empirical estimation was carried out by Wittke, Sumka and Ereksion (1979), Brown and Rosen (1982), Palmquist (1984), and Bartik (1987) for houses, Nerlove (1995) for wines, Pakes (2003) for personal computers.

Rosen’s (1974) approach deals with perfect competition. The analysis was carried to imperfect competition (or at least to cases in which the number of producers is small) by Goldberg (1995), Berry, Levinsohn and Pakes (1995), and Pakes (2003).

In many cases, and art markets are one of these, the purpose of hedonics is merely to determine a price index over time, and not, as in the case of hedonics in consumer price analysis, to derive consumer welfare, which is possible only if utility parameters can be inferred from the estimated model.<sup>16</sup> Therefore, in some sense, it is superfluous to appeal to Rosen’s theory in the case of art price indices.

### 3.2. *Issues with hedonic indexing*<sup>17</sup>

Still, reliable and unbiased estimates have to be obtained, and the problems that are present in single-equation model estimation, such as heteroskedasticity, multicollinearity,<sup>18</sup> choice of functional form, choice of characteristics (independent variables), etc. have also to be dealt with.

In order for the index to be as little contaminated as possible by the heterogeneity of the sales mix over time, and account correctly for quality adjusted price changes, the choice of hedonic characteristics is important. This is dealt with in some detail below.

A further issue is whether one can assume that the coefficients of the hedonic variables (other than time dummies) are constant over time. Though it is doubtful that the

<sup>16</sup> On this issue, see also Palmquist (1992), Feenstra (1995) and Diewert (2001).

<sup>17</sup> One feels somewhat uncomfortable to discuss in one double-spaced page, a topic for which Triplett (2004) needs over 250 single-spaced pages.

<sup>18</sup> Pakes (2003) observes that characteristics are often highly correlated for industrial products (in his case, computers). This is unlikely to be the case for artworks.

hedonic equation really captures changes in consumers' tastes, these can vary over time. Characteristics whose implicit prices are assumed to vary can be made variable, by introducing interaction terms between characteristics and time, or by running adjacent-period regressions, and testing whether the null hypothesis of constancy has to be rejected. Triplett (2004, Chapter 3, Section 2c) argues that constraining the hedonic coefficients to be equal over time does make a difference on the time index, and that, as long as the number of observations is sufficient, running a sequence of adjacent-period regressions produces better results.

A problem that is not encountered in art indexes, but in many other cases, such as computers, cars, and most other consumer durables, is that the ordinary least squares estimator does not take into account the size of the market for each brand or make. The same weight is given to every observation. The issue is dealt with in Silver and Heravi (2004) and Triplett (2004, Chapter 6, Section D).

A related problem is that brands often disappear from the market, and new brands with other characteristics are introduced. This is what makes for most of the difficulty in the construction of consumer price indices. See Boskin (1996) and Boskin et al. (1998), as well as Triplett (2004, Chapter 2).

Finally, we mention a problem analyzed by Melser (2004), who shows that the time-dummy method suggested above may fail to satisfy the monotonicity axiom, which requires the following properties to hold: a price index which compares two periods must increase (decline) if the second (first) period's prices rise, holding other factors fixed. Melser implements a suggestion made by Diewert (2001) which satisfies the axiom. Diewert (2003) also discusses many other unresolved issues in hedonic price indexing.

### 3.3. *Functional form of the hedonic function*

The most widely used functional form is the double-log function in which both prices and characteristics (as long as these are captured by continuous variables<sup>19</sup>) are transformed into natural logarithms, though, Rosen (1974) shows that, in general, theory has nothing to say on the functional form.<sup>20</sup> Therefore, as mentioned by Triplett (2004, Chapter 6, p. 10), "the form of the hedonic function is entirely an empirical matter . . . accordingly, one should choose the functional form that best fits the data, empirically". The empirical question is analyzed (for houses) by Cropper, Deck and McConnell (1988) who run some Monte-Carlo experiments to determine which form should be used to determine implicit prices as correctly as possible. They show that the linear

<sup>19</sup> Note that the space of characteristics is usually not dense. One can hardly find a car whose engine capacity is 1527.3 cc, or a computer whose memory cannot be expressed by an exponential of 2. Therefore, there is no need for the functional form to be smooth.

<sup>20</sup> See however Bartik (1987) who shows that some forms cannot be used in the two-step estimation procedure suggested by Rosen (1974).

Box–Cox functional form may be a reasonable compromise.<sup>21</sup> Wallace (1996) suggests non-parametric estimation which avoids the problem of choosing or imposing a functional form.

### 3.4. Hedonic characteristics for artworks

A good starting point is de Piles *balance des peintres*, added as an appendix to his *Cours de peinture par principes* [de Piles (1708)]. The balance is a table in which de Piles decomposes painting into four fundamental characteristics: composition, drawing, color, expression,<sup>22</sup> and rates each of these on a scale between zero and twenty for 56 painters from his and previous times. Rembrandt, for example, is very low on drawing and obtains 15, 6, 17 and 12 on the characteristics just mentioned, while Michelangelo is very high on drawing, with scores of 8, 17, 4 and 8, respectively. De Piles himself looked at this as a game, but his contemporaries considered it as a “clever way to characterize genius” [Thuillier (1989, p. xxvii)]. Later on, this view changed, and many art historians hate this idea, describing it as a “notorious aberration” [Gombrich (1966, p. 76)] or thinking that de Piles was “at his worst when he tried to be most systematic” [Puttfarcken (1985, p. 42)]. The originality of the *balance* is that it introduces a view of aesthetics that breaks up beauty into its parts. This is of course the very same idea as the one expressed by Lancaster (1966), and used in the early econometric work on hedonics.

Short of having been pursued by many other art critics,<sup>23</sup> such characterizations do hardly exist, and we must rely on alternative descriptions, that are surrogates for aesthetic characteristics. Mandeville, a physician, “attracted attention mainly as the author of provocative essays on economic and social subjects” [de Marchi and van Miegroet (1994)]. In one of his essays [Mandeville (1728)], he lists four factors which explain value: the name of the master, his age (in Mandeville’s words “the time of his age”), the scarcity of his work, and the rank of those owning them.<sup>24</sup>

Characteristics clearly depend on the type of collectible that is studied.<sup>25</sup> For paintings, which have been studied most, Sagot-Duvaurox (2003) singles out as characteristics signature (therefore, name of the artist), provenance (former owners, including exhibitions and literature describing the work), technique (oil, mixed media, etc.), subject matter (landscape, portrait, etc.), support (canvas, paper, etc.), size, place of sale. Nothing very new with respect to Mandeville. Note that, in the past, prices also depended on colors used, subject, number of figures represented.

Since most empirical studies are based on auction data, they use the characteristics made available in sales catalogues, which include name of the artist, size, technique,

<sup>21</sup> See also Halvorsen and Pollakowski (1981) and the discussion in Triplett (2004, Chapter 6, Section C).

<sup>22</sup> Note that the decomposition itself goes back to Plinius.

<sup>23</sup> See Ginsburgh and Weyers (2006) who used de Piles’ characteristics to price out their values in terms of today’s prices and aesthetic values.

<sup>24</sup> See de Marchi and van Miegroet (1994, p. 454).

<sup>25</sup> For details, see the papers cited in Appendix A.

elements of provenance (though names of previous owners rarely appear). Photographs of the paintings are also often included, so that subject matter, colors, etc. can be coded as characteristics as well.<sup>26</sup> Art prices may also be influenced by rarity, as suggested by Mandeville. It is however not clear whether rarity (for example, based on the number of works of an artist that are still in private hands) can be introduced as a hedonic characteristic since it is not a characteristic of the work (or of the artist), but of the market.

Table 1 shows the results for some of the studies that included characteristics in a reasonably uniform manner. Height and width always carry a positive sign, but the negative sign of the coefficient picked by the area variable shows that dimensions cannot become too large: price is a concave function of dimensions. Oil paintings and canvasses are generally more expensive than other media or supports. There is a positive influence of signature and date, when these appear on a painting.

The most controversial characteristic is probably the name of the master (dummy variables that are introduced when portfolios of art are studied), since this is no explanation for why his works are valued. Sociologists, art philosophers and economists seem however to agree that names do change values. Becker (1982), Moulin (1992) and Bonus and Ronte (1997) document in great detail how artists' reputations are constructed, and are responsible for the sometimes high prices that some works command. Art philosophers also hold the opinion that names make for aesthetic value. It is interesting to bring in what Danto, one of the great American analytic art philosophers, writes in one of his essays on contemporary art [Danto (1986, pp. 13–14 and 45)]: “Duchamp’s *Fountain* is, as everyone knows, to all outward appearances a urinal – it *was* a urinal until it became a work of art and acquired such further properties as works of art possess in excess of those possessed by mere real things like urinals . . .” or “The interpretation is not something outside the work: work and interpretation arise together in aesthetic consciousness. As interpretation is inseparable from work, it is inseparable from the artist if it is the artist’s work.” As economist, Grampp (1989, p. 131) purports that the name of the artist belongs to his work: the aesthetic object consists of the painting, the artist and the title. “Imagine”, he writes, “how a dealer would fare if he alone in the market and none of his competitors did not provide information about the painting he offered for sale: no name, no title, no provenance . . . nothing but the price.” Attribution matters. There are hundreds of such examples, one of the latest being Raphael’s *Madonna of the Pinks*, worth some £8000 until its attribution to Raphael in 1991, and paid £22 million by the National Gallery in 2004. All these are good reasons for not being afraid of introducing artist dummies as hedonic characteristics.

The main purpose of the work is to estimate a price index – though this may be biased if the choice of characteristics is poor, or if some important characteristics are omitted. Other hedonic coefficients should be interpreted with some care, given Rosen’s (1974)

<sup>26</sup> Czujack (1997) and Lazzaro (2003) who study respectively paintings by Picasso paintings and prints by Rembrandt are probably the two papers that make use of the largest set of hedonic characteristics.



Table 1  
Examples of results for hedonic equations for painters

	Impressionist & Modern Europ. 1962–1991	Other Europeans 1962–1991	American painters 1962–1990	Belgian painters 1962–1992	Mei–Moses database 1950–2001
<b>Dimensions</b>					
Height (in cm)	0.0111 (0.0002)	0.0080 (0.0010)	0.0090 (0.0002)	0.0102 (0.0004)	0.0572 (0.0040)
Width (in cm)	0.0077 (0.0002)	0.0050 (0.0010)	0.0081 (0.0003)	0.0132 (0.0004)	0.0804 (0.0051)
Area (in 1000 sq. cm)	−0.1898 (0.0053)	−0.2400 (0.0080)	−0.0269 (0.0013)	−0.0622 (0.0036)	−1.3594 (0.1000)
<b>Medium</b>					
Oil	0.0000 –	0.0000 –	0.0000 –	0.0000 –	0.7705 (0.0783)
Collage	−0.5306 (0.0928)	−0.1830 (0.0840)	−0.2690 (0.0721)	–	–
Pastel	–	–	–	–	0.0605 (0.1207)
Drawing	–	–	–	–	−0.6817 (0.1003)
Mixed media	–	–	0.1132 (0.0460)	–	–
Other	–	–	–	–	0.0000
<b>Support</b>					
Canvas	0.0000 –	0.0000 –	0.0000 –	0.0000 –	–
Wood panel	–	−0.0790 (0.0550)	–	−0.0009 (0.0201)	–
Cardboard	–	−0.1300 (0.0710)	–	−0.2229 (0.0438)	–
Paper	–	–	–	−0.2163 (0.0569)	–
Not signed	–	–	–	–	−0.3829 (0.0652)
Not dated	–	–	–	–	−0.1440 (0.0481)
No. of observations	24,540	6410	6224	12,118	3342

*Sources.* Impressionists and Modern, and Other Europeans: de la Barre, Docclo and Ginsburgh (1994); American painters: Demortier (1992); Belgian painters: Ginsburgh and Mertens (1994); Mei and Moses database. All the regressions contain dummies for painters, years and auction houses, but the detailed results are not reported in the table.

warnings. In concluding this section, however, it is hard resisting to quote Triplett (2004, Chapter 5, p. 3) who suggests that “the first principle for conducting a hedonic study is [to] know your product”.

### 3.5. *Other repeat-sales estimators*

The OLS approach in (9) and (10) simply sets the weighting matrix as the identity matrix. While it is the easiest to use, it nonetheless makes the frequently unrealistic assumption that the error term in (8) is homoskedastic. Generalized least squares (GLS) set the weighting matrix as a diagonal matrix where the elements are the ex post squares estimation errors of a first step OLS estimation.

Palmquist (1982) shows that the use of OLS is problematic once there are more than two sales for a specific object and that GLS are needed to derive minimum variance estimates. Goetzmann (1992) shows that GLS provide maximum likelihood estimates of the parameters.

Case and Shiller (1987, 1989) propose an alternative weighting matrix. They assume that the variance of the error term in (8) is a simple linear function of the holding period plus a constant, and suggest a three-stage least square approach. In the first stage, OLS are used. Then one computes the error terms for (8). In the second stage, one regresses the squared error terms against a constant and the holding period. Then, based on the regression, one computes an estimated variance for each object using the holding period. Finally, one sets the diagonal term using the estimated variance and then runs a third stage regression using GLS.

The repeat-sales regressions discussed above are known to introduce certain biases. One of the most serious is spurious negative autocorrelation. This bias is potentially severe at the beginning of the estimated series. Goetzmann (1992) proposes a two-stage Bayesian regression to mitigate the problem over the early periods. The Bayesian formulation imposes an additional restriction that the errors are normally, independently and identically-distributed. The effect on the estimate is dramatic for the early period when data are scarce, and minimal for the period during which data are plentiful. Goetzmann (1992) also suggests a couple of other Bayes estimators.

The repeat-sales estimator is usually presented under its geometric form, since this gives returns in a straightforward way. Shiller (1991), and Goetzmann and Peng (2001) also propose arithmetic repeat-sales estimators that are unbiased and based on arithmetic averages of returns.

### 3.6. *Combining repeat-sales and hedonic estimators*

Case and Quigley (1991) try to use all the information and combine sales and resales (of houses) in a system of equations.<sup>27</sup> They use a hedonic equation for sales and a repeat-sales equation for resales. They also distinguish resales for which characteris-

<sup>27</sup> See also Shiller (1991) for more references and a discussion of combining both estimators.

tics have changed from other resales. Quigley (1995) complements Case and Quigley (1991) by introducing a procedure that is based on an explicit structure of the error term.

Though the results are extremely interesting – in particular, they provide estimates with smaller standard deviations –, the suggestion is hard to apply to paintings, since characteristics are mainly described by qualitative variables, while Case and Quigley deal with (a small number of) continuous variables only. Since in most cases, time is represented by annual dummies, one would need to introduce a very large number of variables.

### 3.7. Other estimators

Many alternative estimators have been suggested and used. A very simple one is discussed in the beginning of the section concerning hedonic regression, the geometric mean in each period. Clearly, one can also think of simple estimators such as arithmetic means or median prices in each period,<sup>28</sup> trimmed or not for outliers. During the 1980s, Sotheby's was carrying an index based on representative objects, the prices of which were reassessed by experts at regular time intervals.<sup>29</sup>

Mark and Goldberg (1984) add a long list of alternative methods, often based on hedonics:<sup>30</sup>

- the “transactions weighted by base period weights” index, in which base period implicit prices (obtained from a hedonic regression run on sales in the first year) are used to compute predicted prices for all units sold in each other year, and the unweighted arithmetic mean of the predicted values is used as the basic price series. This reflects changes in the distribution of the characteristics of units actually sold through time;
- the “transactions weighted by final period weights” index; the idea is the same as in the previous method, except that final period implicit prices are used;
- the “changing annual weights in every period” index; in this method, the weights change every year; predicted prices are calculated and serve as basis for the index;
- a “Laspeyre's index analog”; the Laspeyre index uses base period quantities as weights; the analog suggested uses as weights the average bundle of characteristics from the base year sample, while the changing prices are the hedonic coefficients obtained for each year;

<sup>28</sup> Median prices are used by the US National Association of Realtors. An index based on average prices is computed by Art Market Research in the UK and is regularly published by the *Daily Telegraph*.

<sup>29</sup> See also Clapp and Giacotto (1992) who use price assessments by tax assessors for residential properties, and show that these are strongly correlated with observed prices, so that assessments can play the role of hedonic characteristics. This is very close to using pre-sale estimates by salesrooms as describing artworks, and thus avoiding hedonic characteristics. See, e.g., Beggs and Graddy (1997).

<sup>30</sup> See also Triplett (2004, Chapter 3) for such methods, that he calls “characteristic price index methods”, in which implicit prices are used to compute an index, without any need for time dummies.

- a “Paasche’s index analog”; the Paasche index uses final period quantities as weights; otherwise, calculations are identical to the Laspeyre’s index analog.

Engle, Lilien and Watson (1985) estimate housing returns using a dynamic multiple-indicator multiple-cause system of equations, estimated by maximum likelihood. Candela, Figini and Scorcu (2002), and Candela and Scorcu (2001) discuss other possibilities.

#### 4. Developing hedonic and repeat-sales models: Data acquisition

We will develop this section assuming a researcher is starting at a moment in time and deciding on how to go about collecting data so that he can develop an index that portrays the movement of art prices or returns over some historical period ending with the current period. We assume that the only valid data to use for this type of analysis is auction data either from the direct source catalogs and/or an online computer database.

The RSR-researcher is only interested in the information contained in the provenance of the piece. This gives all the prior owners and any prior auction transaction, to the best of the auction houses ability. Prior auction prices are normally not provided and to find these requires finding the prior auction catalog and price lists. The prior transaction price can sometimes be found on computer databases such as the Hislop Art Sales Index or Art Net but care must be taken to insure that one has found the correct pairing (*Still Life* by Cezanne hardly gives a clue). Moreover titles can change and sizes can be miss-recorded. The shortcoming of the RSR technique is that the number of objects with prior sale results is usually much smaller than the total number of lots offered.

The HR-researcher can use the information given in sales catalogs to serve as a set of independent variables in a hedonic regression. This would allow him to forgo the search for the prior price data. With the help of on line databases, which provide a subset of the data available in catalogs, the task can be made even simpler. The major shortcoming of this technique is that two works can have very similar characteristics for the same year but widely different prices leaving the year dummy alone to try to explain the difference. But since the data collection is easier all the works that sell in each period can be used.

The following variables are usually available: artist, title, size, lot number, date, auction house, medium and support, whether the object is signed or dated and the date, and since 1973 the high and low price estimates. For important works, catalogs also include illustrations, ownership and exhibition history as well as the number of literature citations, and their dates. The ownership history is essential for those hoping to use repeat-sales techniques since it will mention any prior auction sale but rarely the hammer price. This is found by using a good library with an extensive auction catalog collection.

## 5. Comparing hedonic and repeat-sales estimation

### 5.1. The pros and the cons of both types of indices

*Number of observations.* The number of cases in which one can retrieve repeat-sales is usually small, and certainly much smaller than the total number of sales. Ashenfelter and Graddy (2006) compare a hedonic index using 8792 observations with a repeat-sales index with 474 observations. Mei and Moses (2002a, 2002b) have 4900 pairs (9800 observations) for the period 1875–2000 (of which 2300 are for Old Masters). By comparison, the database used by de la Barre, Docclo and Ginsburgh (1994) on Impressionists and followers for paintings sold between 1962 and 1991 contains 24,500 observations. In the Mei–Moses All Art Index that will be used in Section 5.3, the proportion of repeat-sales with respect to all sales is roughly 7% within a ten year time span (1971–1980), 13% within a 20 year span (1971–1990) and 15% within a 30 year span (1971–2000).

In some cases, such as prints which exist in several copies, the number of repeat-sales is much larger. Pesando (1993) collected almost 28,000 repeat-sales, though prints with the same title are not necessarily interchangeable: some may be in better condition than others, some have a good provenance, etc.

Repeat-sales estimation also needs much longer periods on which the index is estimated, since the one before the last sale may have happened long before the last one. The method hardly allows fine disaggregation into submarkets, not to mention constructing indices for individual artists, because the number of observations is usually too small. Goetzmann (1992) explores the issue of sparse data, and suggests some solutions. Sparse data are prone to generate spurious negative autocorrelation and too much volatility. Finally, there is the problem that most data (those for which no resales can be retrieved) are not used, and information is wasted.<sup>31</sup>

*Sample biases in repeat-sales estimators.* Gatzlaff and Haurin (1997) show that because only a small percentage of objects (in their case, houses) sell each year, the sample of those that sell may have non-random statistical properties, due for example to changing economic conditions which influence reserve prices of sellers. They suggest using Heckman's (1979) sample selection correction model. More generally, repeat-sales estimation may suffer from several other sample biases, such as transactions which did not go through auction and whose prices are missing, and outliers, which have more importance than for hedonic indexes, given that data are more sparse.

*Specification biases.* Choosing the functional form and the variables that represent quality are pervasive in hedonic indexing, and can lead to all the problems linked to mis-specification. This is of course avoided in repeat-sales estimation. However, the repeat-sales estimator will not capture the effects of changes in the characteristics of an object between two sales. This happens very often in the case of houses, less so in the

<sup>31</sup> See however the discussion on hybrid methods.

case of art objects, though these may get damaged, restored or even reattributed over time, and this may change their hedonic characteristics.

*Revision volatility.* Wang and Zorn (1997) among others note that “revision volatility” affects both types of indices. As new data become available, previous estimates of the index change, though this may be thought of as good since they result from increased efficiency in the estimators.

*Price inflation and exchange rates.* Inflation can be corrected easily, either by deflating prices themselves, or by deflating the index obtained using either method, since deflating is a log-linear operation. Exchange rates are hardly more problematic, except that here individual prices have to be corrected before going to the estimation work. Think however of a Japanese collector who buys in London in pounds, pays in yens, and sells a few years later in New York in dollars that he changes into yens. Obviously, the only thing that can be done is to compute the rate of return based on the first sale in pounds, translate the pounds into dollars at the going exchange rate, and compute the return in dollars with respect to the second sale. This computed return may be positive, though it may be negative for the collector. At least part of this will not happen with stocks, since most of these are bought and sold in the same place.

## 5.2. *The literature on comparing indices*

Many papers deal with the comparison of indices, but there is very little done in a systematic way. In most cases, real data are used, different indices are constructed, but since the “true” underlying return is not known, the indices are simply compared with each other. Most comparisons deal with real estate prices.

Mark and Goldberg (1984) find that among the indices that are enumerated under “Other estimators” in Section 3 (to which they add a hedonic index constructed on the basis of time dummies for every year, a hedonic index run on adjacent years, a repeat-sales index, an arithmetic (unweighted) mean of sale prices in each period index, a median sale price in each period index), the arithmetic mean, surprisingly, performs well as do several of the hedonic price indices. Case (1986) and Case and Shiller (1987) compare RSR with median prices, and find that the latter are not a good measure of appreciation.

Goetzmann (1992) runs simulations on seven different repeat-sales estimators, including those that are discussed above, using NYSE data from which he draws random samples. He shows that all estimators perform well when the number of repeat-sales is large relative to the number of time intervals. Otherwise, Bayesian-type estimators, such as those discussed in Goetzmann (1990, 1992) are the only ones to perform well.

Chanel, Gérard-Varet and Ginsburgh (1996) run simulations by bootstrapping from the real set of Reitlinger’s (1961, 1971) art data, but compute only a mean return and not an index. They compare the hedonic, the repeat-sales and the geometric mean estimators and find that none of them is biased, but the hedonic method provides much smaller standard errors for the return coefficient (which is obviously the consequence of a larger number of observations).

Meese and Wallace (1997) compare hedonic, repeat-sales and hybrid methods. They find that hedonic techniques are better suited. Repeat-sales techniques are subject to sample selection bias, they violate the assumption of constant implicit prices over time, they are too sensitive to small samples and to influential observations, and the usual method to correct for heteroskedasticity is inappropriate.

The results are thus far from clearcut, since most studies [with exception of Goetzmann (1992) and Chanel, Gérard-Varet and Ginsburgh (1996)] cannot really tell which method is better, since they do not control for the underlying “true” returns. Goetzmann (1992) only compares various repeat-sales estimators (and does not compare these with hedonic estimators), and the experiments run by Chanel, Gérard-Varet and Ginsburgh (1996) are not comprehensive enough to draw solid conclusions.<sup>32</sup>

This is the reason for which we decided to run some Monte-Carlo experiments, which are described and discussed now.

### 5.3. *Comparing hedonic and repeat-sales estimation: Some Monte-Carlo experimentation*

The major concern that faces researchers on finding index values for art have to do with the amount of information available and the time frame of reference. If we knew with certainty and transparency the purchase and the sale price of every object sold for a long period of time there would be no question that RSR would give the best measure of the financial return to holding the set of objects included. When the time frame is short or the ratio of repeat-sales data to total sales data is small, questions arise as to whether a hedonic regression with large amounts of data would give a better reading of market returns.

To study this question we utilized the data collected on the Impressionist and Modern market from 1950 to 2002 and used in the Mei–Moses All Art Index. For any sale that originated in New York during that period we used the existing catalogue description of ownership to search out any prior auction sale anywhere in the world in any time period, as well as for the prior hammer price.<sup>33</sup> If prices were available for both sales they were included in the database. This research produced over 2000 repeat-sale pairs, of which 1671 were usable (at least three observations per artist). In addition to collecting the prior prices we also collected hedonic variables for these pairs. These include dummies for artist, salesroom, shape, medium, aspect (height/width), as well as height, width, and area. Other characteristics (signature, date, number of previous owners, number of times auctioned, number of exhibitions, number of citations and whether pictured in the sales catalogue, shape, and subject) were also collected but not used in the Monte-Carlo experiments.

<sup>32</sup> See also Case, Pollakowski and Wachter (1991), and Crone and Voith (1992) for more comparisons.

<sup>33</sup> Using the resources of the Watson Library at the Metropolitan Museum of Art and the New York Public library, 42nd Street branch.

The steps taken next are as follows, and ensure that both the repeat-sales and the hedonic regressions should generate returns that can be compared with some “true” reference returns.

*Step 1.* Generate a “reference series” of returns, that will be used to compare the results obtained by repeat-sales and hedonic regression over varying time frames. This reference series is obtained from a repeat-sales regression using the 1671 pairs (3342 observations) over the 31 year period. Store the estimated returns, as well as the residual variance of the regression.

*Step 2.* Construct artificial repeat-sales prices using the estimated returns in Step 1.

*Step 3.* For each time frame consisting in years  $t = 1, 2, \dots, T$ , to be discussed later, construct 100 samples  $i = 1, 2, \dots, 100$  of prices obtained by adding to the artificial repeat-sales prices generated in Step 2, a normal random disturbance with the usual properties and whose variance is equal to the residual variance of the regression in Step 1. Run a repeat-sales regression on each of the 100 samples, compute the mean return, and the mean tracking error  $TE_R$  obtained as the square root of the sum of squares of the differences between the “true” and the mean returns over the whole period, that is,

$$TE_R = \left[ \frac{1}{T} \sum_t (\gamma_{R,t} - \hat{\gamma}_t)^2 \right]^{1/2},$$

where  $\gamma_{R,t}$  is the mean over the 100 samples of the RSR return in year  $t$ ,  $\hat{\gamma}_t$  is the “true” return in year  $t$ . Compute also the mean standard error of the estimated returns, obtained as

$$SD_R = \frac{1}{T} \sum_t \left[ \frac{1}{100} \sum_i (\gamma_{R,i,t} - \hat{\gamma}_t)^2 \right]^{1/2},$$

where  $\gamma_{R,i,t}$  is the RSR return in sample  $i$ , for year  $t$ .

*Step 4.* Compute a hedonic regression including the characteristics described above, as well as time dummies, using the 3342 observations. Store the estimated implicit prices for characteristics, as well as the residual variance of the regression.

*Step 5.* Construct artificial hedonic prices using the estimated implicit prices obtained in Step 2, and the “true” returns obtained in Step 1.

*Step 6.* For different time frames to be discussed later, pick 14, 8 and 6 times each observation<sup>34</sup> and construct 100 samples of prices obtained by adding to the artificial hedonic prices generated in Step 5, a normal random disturbance with the usual properties and whose variance is equal to the residual variance of the regression in Step 4.

*Step 7.* Run a hedonic regression on each of the 100 samples, compute the mean return, and the mean tracking error  $TE_H$ , obtained as the square root of the sum of squares

<sup>34</sup> Each sale (or resale) is taken 14, 8 and 6 times in order to “simulate” the fact that there are 14, 8 and 6 times more sales than resales in the 10, 20 and 30 year database.



of the differences between the “true” and the mean returns over the whole period, that is,

$$TE_H = \left[ \frac{1}{T} \sum_t (\gamma_{H,t} - \hat{\gamma}_t)^2 \right]^{1/2},$$

where  $\gamma_{H,t}$  is the mean over the 100 samples of the HR return in year  $t$ ,  $\hat{\gamma}_t$  is the “true” return in year  $t$ . Compute also the mean standard error of the estimated returns, obtained as

$$SD_H = \frac{1}{T} \sum_t \left[ \frac{1}{100} \sum_i (\gamma_{H,i,t} - \hat{\gamma}_t)^2 \right]^{1/2},$$

where  $\gamma_{H,i,t}$  is the HR return in sample  $i$ , for year  $t$ .

Observations cover the period 1950–2001. The time frames chosen are 10 years (1972–1981), 20 years (1972–1991) and 30 years (1972–2001), so that one can compare the performance of both estimators over ten, twenty and thirty year periods. There is one problem due to repeat-sales. In order to estimate a repeat-sales regression over the period 1972–1981 say, one may need observations for earlier years during which the first sale occurred. For this reason, all the regressions (repeat-sales and hedonic) are started in 1950, but the tracking errors and the mean standard errors are computed with 1972 as starting year.

Three estimators are compared: the OLS repeat-sales estimators,<sup>35</sup> the OLS hedonic estimator using all of the characteristics described above, and OLS hedonics that use only a subset of characteristics, to check how much ignoring some of the characteristics (in this case, dimensions: height, width, area and shape, that is width/height) matters. The three estimators are compared in Table 2 in terms of tracking errors and mean square errors.

The tracking errors  $TE_R$  and  $TE_H$  are much lower for hedonics than for repeat-sales for small samples (156 and 612 pairs), but, as expected, both methods tend to produce comparable results once the sample gets larger (1671 pairs). Ignoring characteristics does not seem to matter: both tracking errors and mean standard errors are of the same order of magnitude than before.<sup>36</sup> Figures 1–3 compare both the RSR and the HR returns and indices with the “true” ones. Simulation results also show that the mean standard errors are lower for hedonics than for repeat-sales.

As expected, HR performs much better than RSR, even if some of the characteristics are ignored, when the sample size is small. It seems clear that RSR methods should not be used for time frames that include less than 20 years, unless the number of pairs is large.

<sup>35</sup> Since in our artificial samples, errors are normal IID, a GLS estimator would lead to the same results as an OLS estimator. Therefore, the OLS estimator has all the required properties.

<sup>36</sup> For some reason that we do not understand, they are even smaller in some cases.

Table 2  
Comparing OLS hedonic and repeat-sales estimators (tracking errors and standard deviations over various time frames)

	1972–1981	1972–1991	1972–2001
No. of RSR observations	312	1224	3342
No. of HR observations	4368	9792	20,052
Tracking errors			
RSR	0.2723	0.1566	0.0306
HR full	0.0333	0.0110	0.0070
HR without dimensions	0.0307	0.0012	0.0026
Standard errors			
RSR	0.6580	0.4903	0.3775
HR full	0.3485	0.2983	0.2925
HR without dimensions	0.3914	0.3201	0.2665

HR without dimensions contains the same regressors as HR full, except Height, Width and Area of the work.

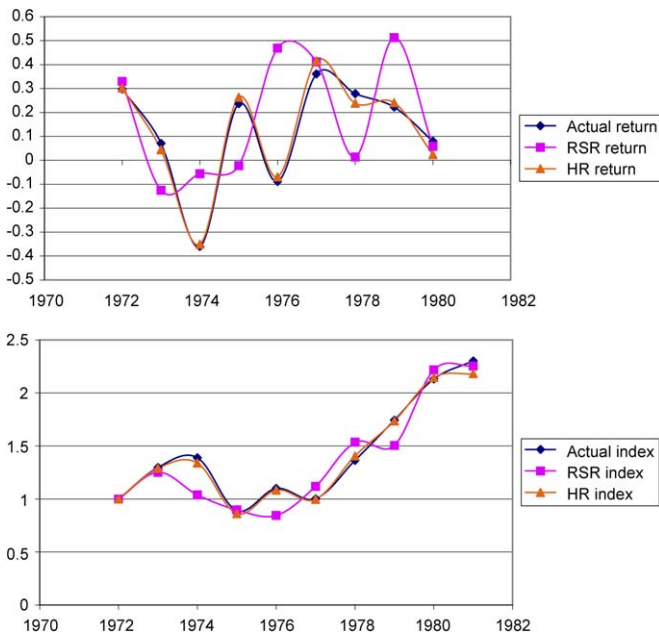


Figure 1. Repeat-sales and hedonic returns and indices 1972–1981.

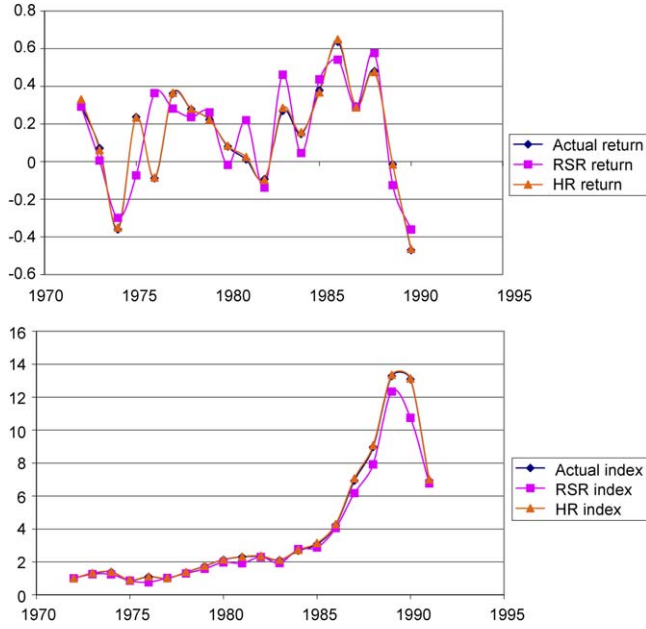


Figure 2. Repeat-sales and hedonic returns and indices 1972–1991.

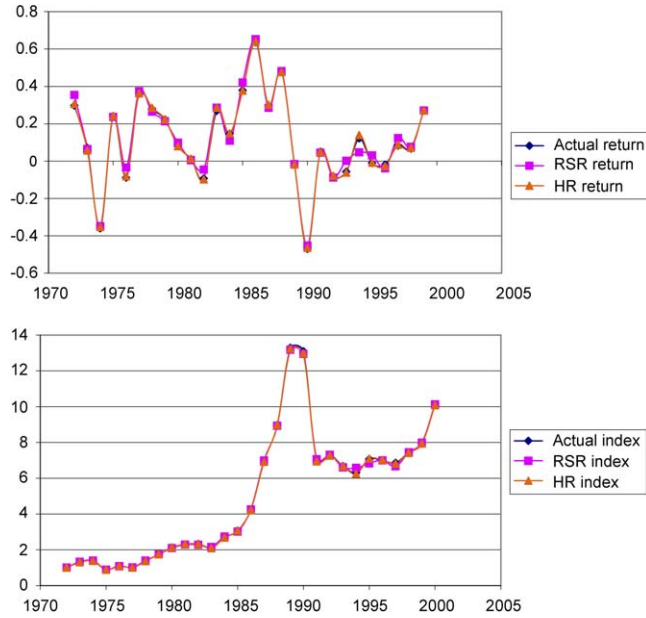


Figure 3. Repeat-sales and hedonic returns and indices 1972–2000.

Table 3  
Comparison of OLS and GLS repeat-sales estimators

	OLS	GLS	Case–Shiller	Bayes
Mean return	0.118	0.119	0.118	0.119
Standard deviation of return	0.223	0.203	0.194	0.202
Correlations between annual returns				
OLS	1.00	–	–	–
GLS	0.74	1.00	–	–
Case and Shiller	0.87	0.96	1.00	–
Bayes	0.74	0.99	0.96	1.00

#### 5.4. Comparing three GLS estimators for repeat-sales estimation

Using the same dataset (1950–2001), we now compare the OLS estimator and three generalized least-squares estimators to compute repeat-sales regressions: the standard generalized least-squares estimator, with a diagonal weighting matrix in which the weights are equal to the squared estimation errors of a first step OLS estimation, the [Case and Shiller \(1987\)](#) three-stage least squares estimator and [Goetzmann’s \(1992\)](#) two-stage Bayesian estimator. [Table 3](#) reproduces the main results in terms of correlation between indices. Results show that though the mean return resulting from all four methods are almost identical, the correlation of the returns between the OLS and GLS estimators is not as high as the correlation of one GLS estimator with another. It also shows that all three GLS estimators give almost identical results.

### Appendix A: Studies on returns

A very large number of different art or antique markets have been looked at: African art [[Degand \(1994\)](#)], books [[Blogie \(1965–1967\)](#); [Ginsburgh and Orban de Xivry \(1994\)](#)], vintage cars [[Van Haverbeke \(1991\)](#)], ancient coins [[Verbert \(1991\)](#)], Old Master drawings [[Ginsburgh and Schwed \(1992\)](#)], furniture [[Graeser \(1993\)](#)], crafted ivory objects [[Lebrun \(1993\)](#)], jewellery [[Chanel, Gérard-Varet and Vincent \(1996\)](#)], photographs [[Pflieger and Sagot-Duvaurox \(1994\)](#)], prints [[Candela and Scorcu \(2001\)](#); [Czujack, Flores and Ginsburgh \(1996\)](#); [Lavand’homme \(1992\)](#); [Pesando \(1993\)](#); [Pesando and Schum \(1996, 1999\)](#)], sculpture [[Locatelli Biey and Zanola \(2002\)](#)], violins [[Heinen \(1993\)](#); [Ross and Zonderman \(1989\)](#)], silverware [[Dorchy and Ginsburgh \(1993\)](#); [Bauwens and Ginsburgh \(2000\)](#)], stamps [[Cardell, Kling and Petry \(1995\)](#); [Feuillolay \(1996\)](#)].

Paintings, often Impressionists and their followers, but also American, Belgian, Canadian, Italian, Latin-American, Pre-Raphaelite, and other painters, raised most of the effort. See [Agnello and Pierce \(1996\)](#), [Anderson \(1974\)](#), [Ashenfelter \(1989\)](#), [Ashenfelter, Graddy and Stevens \(2001\)](#), [Baumol \(1986\)](#), [Beggs and Graddy \(1997\)](#), [Buelens and](#)

Ginsburgh (1993), Candela and Scorcu (1997), Candela, Figini and Scorcu (2003), Chanel (1995), Chanel, Gérard-Varet and Ginsburgh (1994, 1996), de la Barre, Docclo and Ginsburgh (1994), de Limburg-Stirum and Ginsburgh (1995), Demortier (1992), Ekelund, Ressler and Watson (1998), Flores, Ginsburgh and Jeanfils (1999), Frey and Pommerehne (1989a, 1989b), Ginsburgh and Jeanfils (1995), Ginsburgh and Mertens (1994), Ginsburgh and Penders (1997), Goetzmann (1990, 1993, 1996), Goetzmann and Spiegel (2003), Locatelli Biey and Zanola (1999), Mei and Moses (2001, 2002a, 2002b, 2005), Pommerehne and Feld (1997), Schneider and Pommerehne (1983), Stein (1977), Valsan (2002). Some papers look at individuals painters: a couple of Impressionists and Post-Impressionists painters [de la Barre, Docclo and Ginsburgh (1994)] and Picasso [Czujack (1997); Pesando and Schum (1999)]. Landes (2000) looks at the Ganz collection, and some studies have dealt with the return on the collection built up in the 1980s by the British railway pension fund.<sup>37</sup>

Many more studies are compiled and surveyed in Ashenfelter and Graddy (2003, 2006), Burton and Jacobsen (1999), and Frey and Eichenberger (1995a, 1995b), who provide convenient tables with comparative results.

## References

- Abraham, K., Greenless, J., Moulton, B. (1998). "Working to improve the consumer price index". *Journal of Economic Perspectives* 12, 23–36.
- Agnello, R.J., Pierce, R.K. (1996). "Financial returns, price determinants, and genre effects". *Journal of Cultural Economics* 20, 359–383.
- Anderson, R.C. (1974). "Paintings as an investment". *Economic Inquiry* 12, 13–25.
- Ashenfelter, O. (1989). "How auctions work for wine and art". *Journal of Economic Perspectives* 3, 23–36.
- Ashenfelter, O., Graddy, K. (2003). "Auctions and the price of art". *Journal of Economic Literature* XLI, 763–787.
- Ashenfelter, O., Graddy, K. (2006). "Art auctions". In: Ginsburgh, V.A., Throsby, D. (Eds.), *Handbook of the Economics of Art and Culture*, vol. 1. Elsevier, Amsterdam. Chapter 26 in this volume.
- Ashenfelter, O., Graddy, K., Stevens, M. (2001). "A study of sale rates and prices in Impressionist and Contemporary art auctions". Working paper.
- Bailey, M., Muth, J.F., Nourse, H.O. (1963). "A regression method for real estate price index construction". *Journal of the American Statistical Association* 58, 933–942.
- Baumol, W.J. (1986). "Unnatural value: Or art investment as floating crap game". *American Economic Review, Papers and Proceedings* 76, 10–14.
- Bartik, T. (1987). "The estimation of demand parameters in hedonic price models". *Journal of Political Economy* 95, 81–88.
- Bauwens, L., Ginsburgh, V. (2000). "Art experts and art auctions. Are presale estimates fully informative?". *Louvain Economic Review* 66, 131–144.
- Becker, H. (1982). *Art Worlds*. The University of California Press, Berkeley CA.
- Beggs, A., Graddy, K. (1997). "Declining value and the afternoon effect: Evidence from art auctions". *RAND Journal of Economics* 28, 544–565.

<sup>37</sup> See, e.g., *The Toronto Globe and Mail*, July 5, 1996 [quoted by Pesando and Schum (1999)], or *Forbes*, December 5, 1994 [quoted by Burton and Jacobsen (1999)].

- Berry, S., Levinsohn, J., Pakes, A. (1995). "Automobile prices in market equilibrium". *Econometrica* 63, 841–890.
- Blogie, J. (1965–1967). "Essai sur l'évolution des prix de quelques livres marquants de 1850 à nos jours". *Le Livre et l'Estampe* 41–42, 3–18; 43, 3–24; 44, 3–21; 45–46, 3–48; 47–48, 3–27; 49–50, 3–9; 51–52, 3–8.
- Bonus, H., Ronthe, D. (1997). "Credibility and economic value in the visual arts". *Journal of Cultural Economics* 21, 103–118.
- Boskin, M. (1996). "Towards a more accurate measure of the cost of living". Interim Report to the Senate Financial Committee, Advisory Commission to the Study of the Consumer Price Index, Washington, DC.
- Boskin, M., Delberger, E., Gordon, R., Griliches, Z., Jorgenson, D. (1998). "Consumer prices in the consumer price index and the cost of living". *Journal of Economic Perspectives* 12, 3–26.
- Brown, J., Rosen, H. (1982). "On the estimation of structural hedonic models". *Econometrica* 50, 765–768.
- Buelens, N., Ginsburgh, V. (1993). "Revisiting Baumol's 'Unnatural value: Or art as a floating crap game'". *European Economic Review* 37, 1351–1371.
- Burton, B.J., Jacobsen, J.P. (1999). "Measuring returns on investments in collectibles". *Journal of Economic Perspectives* 13, 219–231.
- Candela, G., Figini, P., Scorcu, A. (2002). "Hedonic prices in the art market: A reassessment". Paper presented at the conference of the Association of Cultural Economics International, Rotterdam.
- Candela, G., Figini, P., Scorcu, A. (2003). "Price indices for artists – A proposal". Working Paper.
- Candela, G., Scorcu, A.E. (1997). "A price index for art market auctions". *Journal of Cultural Economics* 21, 175–196.
- Candela, G., Scorcu, A.E. (2001). "In search of stylized facts on art markets for prints and drawings in Italy". *Journal of Cultural Economics* 25, 219–231.
- Cardell, N., Kling, J., Petry, G. (1995). "Stamps returns and economic factors". *Southern Economic Journal* 62, 411–427.
- Case, K.E. (1986). "The market for simple family homes in the Boston area". *New England Economic Review*, May–June, 38–48.
- Case, K.E., Pollakowski, H., Wachter, S. (1991). "On choosing among house price index methodologies". *AREUEA Journal* 19, 286–307.
- Case, K.E., Quigley, T.M. (1991). "The dynamics of real estate prices". *Review of Economics and Statistics* 73, 50–58.
- Case, K.E., Shiller, R.T. (1987). "Prices of single family homes since 1970: New indices for four cities". *New England Economic Review*, September–October, 44–56.
- Case, K.E., Shiller, R.T. (1989). "The efficiency of the market for single family homes". *American Economic Review* 79, 125–137.
- Chanel, O. (1995). "Is art market behaviour predictable". *European Economic Review* 39, 519–527.
- Chanel, O., Gérard-Varet, L.-A., Ginsburgh, V. (1994). "Prices and returns on paintings: An exercise on how to price the priceless". *The Geneva Papers on Risk and Insurance Theory* 19, 7–21.
- Chanel, O., Gérard-Varet, L.-A., Ginsburgh, V. (1996). "The relevance of hedonic price indexes: The case of paintings". *Journal of Cultural Economics* 20, 1–24.
- Chanel, O., Gérard-Varet, L.-A., Vincent, S. (1996). "Auction theory and practice: Evidence from the market for jewelry". In: Ginsburgh, V., Menger, P.-M. (Eds.), *Economics of the Arts. Selected Essays*. North-Holland, Amsterdam.
- Chanel, O., Docclo, S., Gérard-Varet, L.-A., Ginsburgh, V. (1993). "Rentabilité des placements sur le marché de l'art de 1957 à 1988". *Risques* 13, 133–158.
- Clapp, J., Giacotto, C. (1992). "Estimating price indices for residential property: A comparison of repeat-sales and assessed value methods". *Journal of the American Statistical Association* 87, 300–306.
- Court, A.T. (1939). "Hedonic price indexes with automotive examples". In: *The Dynamics of Automobile Demand*. The General Motors Corporation, New York, pp. 99–117.
- Crone, T., Voith, R. (1992). "Estimating house price appreciation: A comparison of methods". *Journal of Housing Economics* 2, 324–338.
- Cropper, M., Deck, L., McConnell, K. (1988). "On the choice of functional forms for hedonic price functions". *Review of Economics and Statistics* 70, 668–675.

- Czujack, C. (1997). "Picasso paintings at auction, 1963–1994". *Journal of Cultural Economics* 21, 229–247.
- Czujack, C., Flores, R., Ginsburgh, V. (1996). "On long-run price comovements between paintings and prints on international markets". In: Ginsburgh, V., Menger, P.-M. (Eds.), *Economics of the Arts. Selected Essays*. North-Holland, Amsterdam.
- Danto, A. (1986). *The Philosophical Disenfranchisement of Art*. Columbia University Press, New York.
- Degand, J.C. (1994). "Evolution du marché de l'art africain de 1970 à 1992". Université Libre de Bruxelles.
- de la Barre, M., Docclo, S., Ginsburgh, V. (1994). "Returns of Impressionist, Modern and Contemporary European paintings, 1962–1991". *Annales d'Economie et de Statistique* 35, 143–181.
- de Limburg-Stirum, M., Ginsburgh, V. (1995). "The fall and rise of symbolist and pre-Raphaelite painters". *The Art Newspaper*, September.
- de Marchi, N., van Miegroet, H. (1994). "Art, value, and market practices in the Netherlands in the seventeenth century". *Art Bulletin* LXXVI, 451–464.
- de Piles, R. (1708). *Cours de peinture par principes*. Gallimard, Paris. Edited by J. Thuillier, 1989.
- Demortier, G. (1992). "Rentabilité et évolution des prix des peintures américaines modernes et contemporaines". Université Libre de Bruxelles.
- Diewert, E. (2001). "Hedonic regressions: A consumer theory approach". Department of Economics, University of British Columbia.
- Diewert, E. (2003). "Hedonic regressions: A review of some unresolved issues". Department of Economics, University of British Columbia.
- Dorchy, L., Ginsburgh, V. (1993). "Price trends for English and French silver 1976–1991". *Arts, Antiques, Auctions* 240, 152–155.
- Ekelund, R., Ressler, R., Watson, J. (1998). "Estimates, bias and no-sales in Latin-American art auctions". *Journal of Cultural Economics* 22, 33–42.
- Engle, R., Lilien, D., Watson, M. (1985). "A dymimic model of housing price determination". *Journal of Econometrics* 28, 307–326.
- Epple (1987). "Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products". *Journal of Political Economy* 95, 59–80.
- Feenstra, R. (1995). "Exact hedonic price indexes". *Review of Economics and Statistics* 77, 634–653.
- Feuillolay, M. (1996). "Spécificités des objets d'art et de collection comme actifs patrimoniaux. Le timbre-poste valeur universelle". Doctoral Thesis, Université des Sciences et Technologies, Lille.
- Flores, R., Ginsburgh, V., Jeanfils, Ph. (1999). "Long and short-term portfolio choices of paintings". *Journal of Cultural Economics* 23 (3), 193–210.
- Frey, B., Eichenberger, R. (1995a). "On the rate of return in the art market: Survey and evaluation". *European Economic Review* 39, 529–537.
- Frey, B., Eichenberger, R. (1995b). "On the return of art investment: Return analyses". *Journal of Cultural Economics* 19 (3), 207–220.
- Frey, B., Pommerehne, W. (1989a). *Muses and Markets*. Basil Blackwell, London.
- Frey, B., Pommerehne, W. (1989b). "Art investment: An empirical inquiry". *Southern Economic Journal* 56, 538–548.
- Gatzlaff, D., Haurin, D. (1997). "Sample selection bias and repeated-sales index estimates". *Journal of Real Estate Finance and Economics* 14, 33–50.
- Ginsburgh, V., Jeanfils, P. (1995). "Long-term comovements in international markets for paintings". *European Economic Review* 38, 538–548.
- Ginsburgh, V., Mertens, S. (1994). "La peinture belge moderne et contemporaine. Evolution des prix dans les ventes internationales, 1962–1992". *Revue du Crédit Communal* 190, 57–80.
- Ginsburgh, V., Orban de Xivry, B. (1994). "La formation du prix du livre ancien. Quelques exemples". *Le Livre et l'Estampe* 40, 65–91.
- Ginsburgh, V., Penders, F. (1997). "Land artists and art markets". *Journal of Cultural Economics* 21, 219–228.
- Ginsburgh, V., Schwed, N. (1992). "Can the art market be subject to econometrics, the case of Old Masters drawings". *The Art Newspaper*, July–September.
- Ginsburgh, V., Weyers, S. (2006). "On the contemporaneousness of Roger de Piles' Balance des Peintres". In: Amariglio, J., Cullenberg, S., Childers, J. (Eds.), *The Aesthetics of Value*. Routledge, London. In Press.

- Goetzmann, W.N. (1990). "Accounting for taste: An analysis of art returns over three centuries". First Boston Working Paper Series FB-90-11, November.
- Goetzmann, W.N. (1992). "The accuracy of real estate indices: Repeat-sales estimators". *Journal of Real Estate Finance and Economics* 5, 71–84.
- Goetzmann, W.N. (1993). "Accounting for taste: Art and financial markets over three centuries". *American Economic Review* 83, 1370–1376.
- Goetzmann, W.N. (1996). "How costly is the fall from fashion. Survivorship bias in the painting market". In: Ginsburgh, V., Menger, P.-M. (Eds.), *Economics of the Arts. Selected Essays*. North-Holland, Amsterdam.
- Goetzmann, W.N., Peng, L. (2001). "The bias of the RSR estimator and the accuracy of some alternatives". Yale School of Management Working Paper ICF-00-27.
- Goetzmann, W.N., Peng, L. (2003). "Estimating indices in the presence of seller reservation prices". Yale School of Management Working Paper ysm352.
- Goetzmann, W., Spiegel, M. (2003). "Art market repeat-sales indices based upon Gabruius S.P.A. data".
- Goldberg, P. (1995). "Product differentiation and oligopoly in international markets: The case of the U.S. automobile industry". *Econometrica* 63, 891–951.
- Gombrich, E. (1966). *Norm and Form. Studies in the Art of the Renaissance*. Phaidon, London.
- Graesser, P. (1993). "Rate of return to investment in American antique furniture". *Southern Economic Journal* 59, 817–821.
- Grapp, W.D. (1989). *Pricing the Priceless, Art, Artists and Economics*. Basic Books Inc., New York.
- Griliches, Z. (1961). "Hedonic price indexes for automobiles: An econometric analysis of quality change". U.S. Congress, Joint Economic Committee, Government Price Statistics. Government Printing Office, Washington, DC.
- Griliches, Z. (Ed.) (1971a). *Price Indexes and Quality Change: Studies in New Methods of Measurement*. Harvard University Press, Cambridge, MA.
- Griliches, Z. (1971b). "Hedonic prices revisited". In: Griliches, Z. (Ed.), *Price Indexes and Quality Change: Studies in New Methods of Measurement*. Harvard University Press, Cambridge, MA, pp. 3–15.
- Griliches, Z. (1990). "Hedonic price indexes and the measurement of capital and productivity". In: Berndt, E., Triplett, J. (Eds.), *Fifty Years of Economic Measurement*. University of Chicago Press, Chicago, IL.
- Guerzoni, G. (1994). "Reflections on historical series of art prices: Reitlinger's data revisited". *Journal of Cultural Economics* 19, 251–260.
- Halvorsen, R., Pollakowski, H. (1981). "Choice of functional forms for hedonic price equations". *Journal of Urban Economics* 10, 37–49.
- Heckman, J. (1979). "Sample selection bias as a specification error". *Econometrica* 47, 153–161.
- Heinen, M.L. (1993). "L'évaluation monétaire des violons". Université Libre de Bruxelles.
- Lancaster, K. (1966). "A new approach to consumer theory". *Journal of Political Economy* 74, 132–157.
- Landes, W.M. (2000). "Winning the art lottery: The economic returns to the Ganz collection". *Louvain Economic Review* 66, 111–130.
- Lavand'homme, P. (1992). "Analyse du marché des estampes et de l'évolution de leurs prix". Université Libre de Bruxelles.
- Lazzaro, E. (2003). "Assessing the quality of cultural goods: The value of originality in Rembrandt's prints". Université Libre de Bruxelles.
- Lebrun, D. (1993). "De l'ivoire brut à l'ivoire travaillé et ancien: Contraintes, évolution, perspectives et vecteurs des marchés". Université Libre de Bruxelles.
- Locatelli Biey, M., Zanola, R. (1999). "Investment in paintings: A short-run price index". *Journal of Cultural Economics* 23, 209–219.
- Locatelli Biey, M., Zanola, R. (2002). "The sculpture market: An adjacent year regression index". *Journal of Cultural Economics* 26, 65–78.
- Mandeville, B. (1728). *The Fable of the Bees or Private Vices, Publick Benefits. With a critical, historical, and explanatory comment by F.B. Kaye*. Oxford, 1924.
- Mark, J.H., Goldberg, M.A. (1984). "Alternative housing price indices: An evaluation". *AREUEA Journal* 12, 31–49.



- McAndrew, C., Thompson, R. (2003). "Pre-sale estimates, risk analysis, and the investment quality of fine art". Unpublished manuscript.
- McAndrew, C., Thompson, R. (2004). "Are presale art auction estimates unbiased? Some affirming evidence". Unpublished manuscript.
- Meese, R., Wallace, N. (1997). "The construction of residential housing price indices. A comparison of repeat-sales, hedonic regression and hybrid approaches". *Journal of Real Estate Finance and Economics* 14, 51–73.
- Mei, J., Moses, M. (2001). "Art as an investment and the origin of the 'masterpiece effect'. Evidence from 1875–2000". Mimeo, Stern School of Business, New York.
- Mei, J., Moses, M. (2002a). "Art as an investment and the underperformance of masterpieces". *American Economic Review* 92, 1656–1668.
- Mei, J., Moses, M. (2002b). "Are investors credulous? Some preliminary evidence from art auctions". Working Paper FIN 02-039, NYU Stern School of Business.
- Mei, J., Moses, M. (2005). "Vested interest and biased price estimates. Evidence from an auction market". *Journal of Finance* 60, 2409–2435.
- Melser, D. (2004). "The hedonic regression time-dummy method and the monotonicity axioms". Department of Economics, University of New South Wales.
- Moulin, R. (1992). *L'artiste, l'institution et le marché*. Flammarion, Paris.
- Nerlove, M. (1995). "Hedonic price functions and the measurement of preferences: The case of Swedish wine consumers". *European Economic Review* 39, 1697–1716.
- Pakes, A. (2003). "A reconsideration of hedonic price indexes with an application to PC's". *American Economic Review* 93, 1578–1596.
- Palmquist, R.B. (1982). "Measuring environmental effects on property values without hedonic regressions". *Journal of Urban Economics* 11, 333–347.
- Palmquist, R. (1984). "Estimating the demand for characteristics of housing". *The Review of Economics and Statistics* 66, 394–404.
- Palmquist, R.B. (1992). "Hedonic methods". In: Braden, X., Kolstad, X. (Eds.), *Measuring the Demand for Environmental Quality*. North-Holland, Amsterdam.
- Pesando, J.E. (1993). "Art as an investment. The market for modern prints". *American Economic Review* 83, 1075–1089.
- Pesando, J., Shum, P. (1996). "Price anomalies at auction: Evidence from the market for modern prints". In: Ginsburgh, V., Menger, P.-M. (Eds.), *Economics of the Arts. Selected Essays*. North-Holland, Amsterdam.
- Pesando, J., Schum, P. (1999). "The return to Picasso's prints and to traditional financial assets 1997 to 1996". *Journal of Cultural Economics* 23, 183–192.
- Pflieger, S., Sagot-Duvaurox, D. (1994). *Le marché des tirages photographiques*. La Documentation Française, Paris.
- Pommerehne, W.W., Feld, L. (1997). "The impact of museum purchase on the auction price of paintings". *Journal of Cultural Economics* 21, 249–271.
- Puttfarcken, T. (1985). *Roger de Piles' Theory of Art*. Yale University Press, New Haven, CT.
- Quigley, J. (1995). "A simple hybrid method for estimating real estate price indexes". *Journal of Housing Economics* 4, 1–12.
- Reitlinger, G. (1961). *The Economics of Taste: The Rise and Fall of the Picture Market, 1760–1960*. Barrie and Rockliff Ltd., London.
- Reitlinger, G. (1971). *The Economics of Taste: The Rise and Fall of the Picture Market, 1760–1960*. Barrie and Jenkins Ltd., London.
- Rosen, S. (1974). "Hedonic prices and implicit markets: Product differentiation in pure competition". *Journal of Political Economy* 82, 34–55.
- Ross, M.H., Zonderman, S. (1989). "Capital gains and the rate of return on a Stradivarius". *Economic Inquiry* 27, 529–540.
- Sagot-Duvaurox, D. (2003). "Art prices". In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham.

- Schneider, F., Pommerehne, W. (1983). "Analyzing the market of works of contemporary fine arts: An exploratory study". *Journal of Cultural Economics* 7, 41–68.
- Shiller, R.T. (1991). "Arithmetic repeat-sales price estimators". *Journal of Housing Economics* 1, 110–126.
- Silver, M., Heravi, S. (2004). "Hedonic price indices and the matched models approach". *The Manchester School* 72, 24–49.
- Stein, J.P. (1977). "The monetary appreciation of paintings". *Journal of Political Economy* 85, 1021–1035.
- Thuillier, J. (1989). "Préface". In: de Piles' *Cours de peinture par principes*. Gallimard, Paris.
- Triplett, J. (1969). "Automobiles and hedonic quality measurement". *Journal of Political Economy* 77, 408–417.
- Triplett, J. (2004). *Handbook on Hedonic Indexes and Quality Adjustment in Price Indexes: Special Application to Information Technology Products*. Brookings Institution. Downloadable from [http://www.brookings.edu/views/papers/triplett/200407\\_book.htm](http://www.brookings.edu/views/papers/triplett/200407_book.htm), August 2004.
- Valsan, C. (2002). "Canadian versus American art: What pays off and why". *Journal of Cultural Economics* 26, 203–216.
- Van Haverbeke, P. (1991). "L'automobile de collection en tant que placement, mémoire de licence en sciences économiques". Université Libre de Bruxelles.
- Verbert, N. (1991). "L'évolution du prix des monnaies grecques archaïques entre 1980 et 1990". Université Libre de Bruxelles.
- Wallace, N. (1996). "Hedonic-based price indexes for housing. Theory, estimation and index construction". *FRBSF Economic Review* 3, 34–48.
- Wang, F., Zorn, P. (1997). "Estimating house price growth with repeat-sales data. What's the aim of the game?" *Journal of Housing Economics* 6, 93–118.
- Wenzlick, R. (1952). "As I see the fluctuations in the selling prices of single-family residences". *The Real Estate Analyst* XXI, 47–48.
- Wittke, A., Sumka, H., Erekson, H. (1979). "An estimation of a structural hedonic price model of the housing market: An application of Rosen's theory of implicit markets". *Econometrica* 47, 1151–1173.
- Wyngarden, H. (1927). *An Index of Local Real Estate Prices*. University of Michigan, Ann Arbor, MI.

## CULTURAL HERITAGE: ECONOMIC ANALYSIS AND PUBLIC POLICY\*

ILDE RIZZO

*University of Catania, Italy*

DAVID THROSBY

*Macquarie University, Sydney, Australia*

### Contents

Abstract	984
Keywords	984
1. Introduction	985
2. Heritage as capital asset	986
3. Cultural capital, natural capital and sustainability	988
3.1. Natural and cultural capital: Some similarities	988
3.2. Sustainability paradigms	989
3.3. Sustainability of cultural capital	991
4. Investment appraisal procedures applied to cultural heritage projects	995
4.1. Cost–benefit analysis applied to heritage decisions	995
4.2. Estimation of the non-market benefits of heritage	996
4.3. Cost–benefit analysis and cultural value	998
5. Public intervention in heritage markets	999
6. Public-choice approach to heritage decision-making	1002
7. Devolution in the formation and implementation of heritage policy	1005
7.1. Advantages and disadvantages of devolution	1005
7.2. Practical issues	1006
7.3. Summary	1008
8. The role of the private sector	1008
8.1. Effects of regulation	1008
8.2. Sustainability in heritage financing	1009
8.3. Philanthropy and heritage	1011
9. Conclusions	1012
References	1013

\* With the usual caveat, we express our gratitude to Françoise Benhamou and Victor Ginsburgh for constructive comments on an earlier version of this chapter.

**Abstract**

This chapter shows how economic theory and public policy analysis can illuminate decision-making relating to cultural heritage. We argue that from an economic viewpoint the appropriate conceptualisation of heritage is as a capital asset. Regarding heritage as cultural capital invites consideration of sustainability aspects, in parallel with the treatment of natural capital in economic theory, allowing us to derive a sustainability rule for cultural capital accumulation. The application of cost–benefit analysis to heritage investment appraisal is discussed, with particular reference to the assessment of non-market benefits. Turning to policy issues, we examine ways in which governments intervene in heritage markets, with particular attention to listing and other forms of regulation. Questions of institutional design, financing and policy delivery in a multi-jurisdictional framework are discussed, and finally the role of the private sector is considered, with emphasis on the possibility of crowding out and the incentive effects of public policy on private behaviour in the heritage field.

**Keywords**

cultural capital, sustainable development, non-market benefits, devolution, regulation, fiscal federalism

*JEL classification:* H59, H77, Q01, Q57, Z11

## 1. Introduction

There are a number of definitions of cultural heritage, though there is no agreement on a precise specification of how restricted or extensive the concept should be. Indeed, [Benhamou \(2003, p. 255\)](#) suggests that “heritage is a social construction where boundaries are unstable and blurred”. Nevertheless, common to most accounts is the fact that, as its name indicates, heritage is something that is received from the past; although we are accustomed to thinking of heritage as being old, bequeathed to us from generations long since dead, there is no reason why it may not have been inherited from the recent past, indeed from within the present generation. Attaching the term “cultural” to “heritage” signifies that the items of concern have some cultural meaning or significance that is regarded as important. So, for example, [Guerzoni \(1997, p. 107\)](#) defines cultural heritage as “a heterogenous set of goods that, in the course of time and in a process of historicization, comes to be recognised as the conveyor of specific cultural traditions”. Examples of cultural heritage that spring readily to mind include tangible items such as buildings, monuments, sites, artworks, artefacts, etc. and also intangible phenomena such as traditions, customs, memories, ideas, languages, beliefs, etc. The present chapter deals in principle with all of these, though most of our discussion focuses on heritage in its tangible form.<sup>1</sup>

Decisions about cultural heritage – what it is, how it should be conserved – are traditionally the province of art historians, conservators, archaeologists, museum directors and similar professionals. When economists raise questions about the criteria on which these decisions are based, their intrusion is often resented.<sup>2</sup> Yet at its simplest, the maintenance or restoration of heritage requires resources, and resources have opportunity costs. Since resources are limited in supply, choices are necessary: What can be preserved and what cannot? How much renovation or restoration is warranted? Whose preferences should guide conservation decisions? Economics may not be capable of providing a complete answer to these sorts of questions, but it can certainly illustrate the issues and point out some of the consequences of alternative courses of action.

In this chapter we examine first the theoretical underpinnings of an economic analysis of cultural heritage, building on the concept of cultural capital as a distinct form of capital in economics (Section 2). In this analysis heritage is interpreted as a set of assets with particular characteristics that affect the ways in which consumption and investment decisions are made. In particular, regarding heritage as cultural capital naturally invites consideration of sustainability aspects, in parallel with the treatment of natural capital

<sup>1</sup> For a comprehensive account of the evolution of ideas of heritage over time and space, see [Chastel \(1986\)](#), who notes that the term is also applied to natural and genetic heritage as well as to cultural heritage. He discusses how items of built heritage and their interpretation change as a result of human interventions and with the passage of time.

<sup>2</sup> Recent forays by economists into the heritage field include [Peacock \(1995\)](#), and the collections of essays edited by [Peacock and Rizzo \(1994\)](#), [Schuster, de Monchaux and Riley \(1997\)](#), [Hutter and Rizzo \(1997\)](#) and [Peacock \(1998\)](#); for an oppositional view from a cultural professional, see [Cannon-Brookes \(1996\)](#).

in economic theory (Section 3). Carrying further the idea of heritage as asset enables us in Section 4 to examine the application of investment appraisal methods to heritage projects. Not surprisingly, the assessment of benefits becomes a critical component of such an exercise. In Sections 5 and 6 we turn to public policy issues, looking at the ways in which governments intervene in heritage markets, and then in Section 7 at the particular financing issues raised in a multiple-jurisdiction framework. Finally, in Section 8, the role of the private sector is considered, with emphasis on the possibility of crowding-out and the incentive effects of public policy on private behaviour in the heritage field.

## 2. Heritage as capital asset

Heritage items such as a painting by Monet or a historic building can be seen as capital assets: both required investment of physical and human resources in their original manufacture and construction; both will deteriorate over time unless resources are devoted to their maintenance and upkeep; and both give rise to a flow of services over time which may enter the final consumption of individuals directly (e.g. when people view the painting in a museum or visit the historic building), or which may contribute to the production of further goods and services (e.g. when the painting inspires the creation of new artworks or when the historic building is used as a commercial office space). In other words, heritage items can be interpreted as capital assets with the standard characteristics of ordinary physical capital in economics. Recently suggestions have been made that heritage items are members of a class of capital which is distinct from other forms of capital; this class has been called *cultural capital*<sup>3</sup> [Throsby (1999); Ulibarri (2000); Shockley (2004)]. The distinction lies in the type of value that is embodied in these assets and is yielded by the goods and services they produce. A historic building such as Notre Dame Cathedral or the Taj Mahal is not just any building; certainly it has the characteristics of an “ordinary” building as an item of physical capital, but in addition it has historical and other attributes which an “ordinary” building does not have. These attributes can be described as the building’s “cultural value”, and the same type of cultural value can be attributed to the flow of services it provides.

This notion of the cultural value of certain goods and services such as heritage can be set alongside the more familiar concept of their economic value as measured by variables such as price or as assessed by the tools of economic analysis. Let us assume that cultural value can be measured according to a unit of account that plays a role comparable to that of a monetary scale in measuring economic value. Placing these two indicators of value, economic and cultural, side by side for a range of heritage items we

<sup>3</sup> The use of this term in economics differs from the concept widely used in sociology following Bourdieu (1986), where “cultural capital” refers to an individual’s competence in high-status culture within a holistic conception of society. Insofar as Bourdieu’s usage relates to characteristics of human beings, it is very close to the economic concept of human capital [Becker (1964)].

would expect some relationship between them – assets of greater cultural significance would generally be expected to be more highly valued in monetary terms than less significant ones. However, the relationship would not be perfect since contrary cases can be imagined – a remote religious building of little market value but with strong cultural or historical associations, for example. We return to the distinction between economic and cultural value below. For now we can proceed on the basis of these interpretations to define cultural capital formally as an asset that embodies a store of cultural value, separable from whatever economic value it might possess; in combination with other inputs the asset gives rise to a flow of goods and services over time which may also have cultural value (i.e. which are themselves cultural goods and services).

As noted above, heritage items can be *tangible* or *intangible*. Whilst the definition of tangible heritage in the above terms might be simple enough, the identification of intangible cultural heritage is more elusive. The recently formulated UN *Convention for Safeguarding of the Intangible Cultural Heritage* (2003) defines it as

“the practices, representations, expressions, knowledge, skills – as well as the instruments, objects, artefacts and cultural spaces associated therewith – that communities, groups and, in some cases, individuals recognise as part of their cultural heritage. This intangible cultural heritage, transmitted from generation to generation is constantly recreated by communities and groups in response to their environment, their interaction with nature and their history . . .” [UNESCO (2003, p. 2)].

Like any capital item, cultural capital (whether tangible or intangible) can be represented both as a *stock* of assets and as a *flow* of capital services over time. The value of the capital may be assessed in terms of its asset value at a given point in time or as the value of the flow of services to which it gives rise. Either way, the particular characteristic of cultural capital is that it embodies or gives rise to two types of value, economic and cultural.

Despite the apparent precision of these definitions of cultural heritage, two problems arise in applying them to the real world. First, if cultural heritage as a capital asset is distinguished from other forms of asset by its possession of a specific type of value, whose assessment of that value is to count and how is it to be measured? Formal statements of what should be regarded as heritage that are contained in conventions such as the UNESCO *Convention for the Protection of the World Cultural and Natural Heritage* [UNESCO (1972)], the Burra Charter [Australia ICOMOS (1999)], and the Intangible Heritage Convention referred to above do indeed rely on an appeal to notions of cultural significance, and go so far as to suggest criteria against which significance might be judged; nevertheless they remain imprecise as to the metrics that could be used and how they could be applied. As a result, the definitional problem is generally resolved in practice by looking for consensus judgements amongst a relevant group of experts, stakeholders or policy-makers, whose assessment of cultural value, however imprecisely expressed, is taken as decisive. Whether it is representative or not of society’s cultural perceptions is an open question, as will be examined further in a later section.

The second problem is that in some situations what is regarded as cultural heritage is not so much a well-defined item but rather an assorted collection of goods that were never thought of as having any particular importance at the time they were produced [Peacock (1995)]. This relates to the notion of “material culture” or “material heritage” as identified, for example, by anthropologists. What constitutes heritage in this context is no longer an objective fact but rather a social and cultural construct that is likely to change over time.

### 3. Cultural capital, natural capital and sustainability

#### 3.1. *Natural and cultural capital: Some similarities*

The concept of cultural capital bears some similarities to that of natural capital as it has developed within ecological economics over the last decade or so [El Sarafy (1991); Costanza and Daly (1992); Jansson et al. (1994)]. That is, cultural capital which has been inherited from the past can be seen to have something in common with natural resources, which have also been provided to us as an endowment; natural resources have come from the beneficence of nature, cultural capital has arisen from the creative activities of human kind. Both can be interpreted as imposing a duty of care on the present generation, the essence of the sustainability problem to be discussed further below. Further, a similarity can be seen between the function of natural ecosystems in supporting and maintaining the “natural balance” and the function of what might be referred to as “cultural ecosystems” in supporting and maintaining the cultural life and vitality of human civilisation. Finally, the notion of diversity, so important in the natural world, has an equally significant role to play within cultural systems. It is a characteristic of most cultural goods that they are unique, and this applies particularly to cultural heritage, both tangible and intangible. It can be suggested that cultural diversity is at least as far-reaching as is diversity in nature, and perhaps more so. Hence much of the analysis of biodiversity might be applicable to a consideration of cultural heritage.<sup>4</sup>

A parallel between natural and cultural capital of particular significance in the heritage context is the way in which the long-term management of both types of capital can be cast in terms of the principles of sustainable development. When applied to natural capital, sustainable development implies management of natural resources in a way that provides for the needs of the present generation without compromising the capacity of future generations to meet their own needs,<sup>5</sup> i.e. the principle of intergenerational equity. Another key element of sustainability in natural capital management

<sup>4</sup> On cultural diversity and cultural pluralism, see the report of the World Commission on Environment and Development (1995), contributions to the *World Culture Reports* of 1998 and 2000 published by UNESCO, and the *Universal Declaration on Cultural Diversity* adopted by UNESCO in 2001; see also Acheson and Maule (2004), Throsby (2004).

<sup>5</sup> This “definition” of sustainable development is a paraphrase of that proposed in the Brundtland Report [WCED (1987 p. 43)].



is the precautionary principle which argues for a risk-averse stance in decision-making when irreversible consequences such as species loss are in prospect. Both of these principles are relevant to cultural sustainability. Because the stock of cultural capital, both tangible and intangible, embodies the culture we have inherited from our forebears and which we hand on to future generations, it is inevitable that questions of intergenerational equity are raised; heritage decision-making is constantly faced with the long-term implications of strategies for conservation, preservation, restoration and re-use of buildings and sites. Similarly the precautionary principle can be invoked when demolition of a historic building is threatened; once gone, such unique cultural heritage cannot be replaced.<sup>6</sup>

### 3.2. Sustainability paradigms<sup>7</sup>

A key aspect of sustainability is the maintenance of capital stocks. In terms of physical capital, if a country's aggregate consumption is less than or equal to its net domestic product, it must be at least maintaining its total capital stock, and can therefore be defined as following a sustainable development path. If the capital stock referred to is extended to include natural, human and cultural capital as well as physical capital, the question arises as to whether different types of capital can simply be aggregated, such that a decline in the level of one type of capital can be compensated for by an increase in another. In other words, this raises the issue of substitutability between forms of capital. The contrasting positions in regard to the substitutability between natural and other forms of capital can be seen in two standard texts, Dasgupta and Heal (1979) representing a strictly neoclassical (substitutable) approach, and Pearce and Turner (1990) who put the non-substitutability case.

In the long-running debate about the issue of capital substitutability, two essential paradigms for sustainable development have emerged [Neumayer (2003)]. The first, which can be called "weak sustainability", derives from the original work of Robert Solow and John Hartwick. In a series of papers [Solow (1974a, 1974b); Hartwick (1977, 1978)] they investigated the question of investing the rents from exhaustible resources in the presence of concern for intergenerational equity.<sup>8</sup> In its simplest form this model portrays an economy in which the competitive rents from current use of the exhaustible resource are reinvested in human-made capital goods, enabling society to maintain a constant consumption stream; the accumulation of physical capital exactly offsets the decline in natural capital. As is apparent, this model assumes that natural and human-made capital are perfect substitutes in the production of consumption goods and in the direct provision of utility for both present and future generations. Hence it is the aggregate capital stock that matters and not how it is comprised; in other words, it doesn't

<sup>6</sup> Similar concerns may be expressed about intangible heritage such as languages or customs that may disappear if not maintained.

<sup>7</sup> The following discussion draws on material treated more fully in Throsby (2005).

<sup>8</sup> See also Solow (1986; 1994) and Hartwick (1995).

matter if the present generation uses up exhaustible resources as long as sufficient new physical capital can be provided to future generations by way of compensation.

The other paradigm is that of “strong sustainability” which regards natural capital as being strictly non-substitutable for human-made capital, a view deriving in part from the unique life-supporting properties of global air, land and water systems. Proponents of strong sustainability argue that no other form of capital is capable of providing the basic functions that make human, animal and plant life possible [Barbier et al. (1994)]. Moreover, some forms of natural capital cannot be reconstructed once they are destroyed; for example, the destruction of biodiversity is a loss of natural capital that cannot be reversed, and even climate change could result in ecosystem damage that is irreversible. In other words, the strong sustainability paradigm assumes that the functions of natural capital cannot be replicated no matter how spectacular future technological advances might be.

Particular interest has been focussed on the specification of optimal development paths for an economy under the different paradigms. This requires some means of measuring whether or not a given path is sustainable according to the assumed criteria [Victor (1991)]. An effort in this direction is provided by Pearce and Atkinson (1993), who propose a sustainability indicator  $Z$  for a weakly sustainable development path where the economy is defined as sustainable if it saves more than the combined depreciation on man-made and natural capital, i.e.  $Z > 0$  if and only if

$$S > (\delta_M + \delta_N), \quad (1)$$

where  $S$  is savings,  $\delta$  is depreciation, with the subscripts  $M$  and  $N$  indicating man-made and natural capital, respectively. Pearce and Atkinson proceed to estimate the indicator empirically for a range of countries, concluding that “even on a weak sustainability rule many countries are not likely to pass a sustainability test” (p. 105).<sup>9</sup>

How do these sustainability paradigms apply to cultural capital? Here it has to be recognised that cultural capital gives rise by definition to two sorts of value: economic and cultural. It is clear that provision of many of the *economic* functions of cultural capital is readily imaginable through substitution by physical capital; for example the services of shelter, amenity etc. provided by a historic building could as well be provided by another structure without cultural content. However, since by definition cultural capital is distinguished from physical capital by its embodiment and production of *cultural* value, there would be expected to be zero substitutability between cultural and physical capital in respect of its cultural output, since no other form of capital is capable of providing this sort of value; the new building cannot replicate the historical content of the old.<sup>10</sup>

<sup>9</sup> For further discussion of this model and of valuation problems more generally, see Pearce (1993), especially pp. 44–53.

<sup>10</sup> This is not to say that new buildings cannot themselves be regarded as cultural capital; see further the discussion below on investment in capital creation.

### 3.3. Sustainability of cultural capital

We turn now to the possibility of specifying sustainable development paths for cultural capital, taking account of its particular characteristics. For simplicity we assume a quasi-strong form of sustainability by defining it in terms of the cultural capital stock only, thus ruling out questions of substitutability with other types of capital. A possible model is summarised as follows.

Assume a closed economy which possess a stock of cultural capital  $K$  with an aggregate value to society of  $V = V(K^e, K^c)$  where  $K^e$  is the economic valuation of the capital stock at a given point in time, measured in financial terms, and  $K^c$  is the cultural valuation of the stock measured according to some agreed system of units reflecting the significance or worth to society of the cultural asset. During any time period,<sup>11</sup>  $K$  produces a flow of income  $y$  measured in both monetary-value and cultural-value terms. So, for example, if the capital assets in question are artworks in a museum, or historic buildings or sites, these income flows might be generated by displaying the artworks for people to look at, or by opening the buildings and sites to tourists. In each case a stream of monetary income is generated which accrues to the immediate owners of the asset in question; at the same time a stream of “cultural income” is also generated, some of which accrues to society at large as public-good benefits arising from the existence of these items of the cultural capital stock.

Assume that production takes place according to the following one-period production functions:

$$y^e = F^e(X, K^e), \quad (2a)$$

$$y^c = F^c(X, K^c), \quad (2b)$$

where  $X$  is a vector of other inputs (labour, operating capital, etc.) whose level is determined by the policy-maker, and where the functions  $F$ , for convenience, have the usual properties of constant returns to scale and diminishing marginal product.

We assume  $y^c$  is completely consumed in the period in which it is produced, whereas  $y^e$  can be allocated to current consumption ( $C$ ) or to investment ( $I$ ) in the maintenance of  $K$ , the existing capital stock.<sup>12</sup> Then

$$y^e = C + I. \quad (3)$$

For given  $X$  we can write

$$y^e = rK^e, \quad (4)$$

where  $r$  is interpreted as the economic rate of return on the capital stock. For simplicity, let us assume that all the income is re-invested, hence

$$I = rK^e. \quad (5)$$

<sup>11</sup> For simplicity, we omit time subscripts throughout.

<sup>12</sup> For the present we ignore the possibility of investment in new capital stock (see further below).

Similarly, for given  $X$  we can write

$$y^c = \alpha K^c, \quad (6)$$

where  $\alpha$  is defined as a “cultural appreciation parameter” measuring the extent to which members of society, in a given time period, understand and appreciate the significance or importance of  $K$ . (Thus for a society that cared nothing for its cultural assets,  $\alpha = 0$  and no cultural income would accrue.)

From the viewpoint of sustainability, we are interested in the rate of change in the capital stock from one period to the next, measured in both economic and cultural value terms. Changes over time in the level of  $K^e$  are caused by several different factors:

- exogenous influences affecting the price of the stock (e.g. for an art museum’s collection, these influences arise as a result of movements in the art market);
- depreciation caused by wear and tear (e.g. from damage to a cultural site caused by tourists) or by catastrophic events (e.g. war);
- conservation or restoration investment undertaken with the aim of maintaining the asset in good condition.

Let the first of these factors be measured by  $m$ , the rate of price appreciation (which could be negative) and let the second be measured by  $d$ , the rate of depreciation (always negative). The third factor is simply  $I$  as defined above. Without further ado we can write

$$\dot{K}^e = m - d + r \quad (7)$$

and note that economic sustainability of the capital stock, defined as  $\dot{K}^e \geq 0$ , will require

$$(m + r) \geq d. \quad (8)$$

Before turning to sustainability in the *cultural* value of the capital stock, we note the following assumption. We assume that the rate of change of the appreciation parameter  $\alpha$  is a function of the level of “cultural income” forthcoming in any period. In other words, the greater are the cultural benefits flowing from a given stock of cultural capital, the faster will the community’s cultural appreciation grow. In simpler words, the more people attend art museums, visit heritage sites, perform traditional cultural ceremonies, etc., the more “switched on” will the community become to their own cultural inheritance.<sup>13</sup> Correspondingly, the less these benefits are realised, the smaller will  $\alpha$  become. In fact it is quite possible for  $\alpha$  to be negative, that is, for cultural appreciation to decay if cultural participation falls below a given level. Specifically, let us assume there is a critical level of the flow of cultural value in a given time period,  $\hat{y}^c$  below which  $\alpha$  falls below zero, i.e. we assume

$$\dot{\alpha} > 0 \quad \text{if } y^c > \hat{y}^c, \quad \dot{\alpha} = 0 \quad \text{if } y^c = \hat{y}^c \quad \text{and} \quad \dot{\alpha} < 0 \quad \text{if } y^c < \hat{y}^c.$$

<sup>13</sup> Recall that we are assuming a closed economy, so the imposition of foreign cultural influences does not arise.

Now we can define the factors affecting the level of  $K^c$  in a given time period. We assume two factors are relevant:

- the rate of change of  $\alpha$  as discussed above; and
- the amount of maintenance investment  $I$ .

The reason for including the second of these factors should be readily apparent – at least some components of the cultural value of historic buildings, artworks, etc. will decline if they are neglected and allowed to fall into disrepair (i.e. if  $I = 0$ ).

Accordingly we can write

$$\dot{K}^c = f(\dot{\alpha}, I), \quad (9)$$

where  $f'(\dot{\alpha}) > 0$  and  $f'(I) > 0$ . We can now specify that cultural sustainability will require  $\dot{K}^c \geq 0$ . We note that implementation of this sustainability rule would require knowledge of the function  $f$  in (9) and in turn the critical value of  $\hat{y}^c$ .

It can be argued from a cultural viewpoint that it is *cultural* sustainability that matters in this economy. If this is so, the problem can be framed as follows. For a given initial stock of  $K$  with economic and cultural value of  $K^e$  and  $K^c$ , respectively, the decision problem is to choose  $X$  and  $I$  such that the cultural sustainability condition holds. In other words, in any given period society would need to allocate a sufficient level of resources to utilising its cultural capital stock rather than to other (non-cultural) purposes, and would need to re-invest a sufficient level of the financial income stream so generated in the conservation and maintenance of the stock, in order to ensure no deterioration in the cultural value of the stock in the next period.

Finally we turn to the matter of new investment in cultural capital. This refers to such actions as the creation of artworks, the construction of new buildings that may someday be regarded as “historic” and having particular cultural value, the cultivation of emerging cultural traditions that in due course will be handed on to the next generation, etc. Here it might be noted that the parallel with natural capital begins to break down. These forms of new cultural capital are not like renewable resources which have an inherent capacity for self-regeneration. They have to be created by deliberate production processes.

The incorporation of this new cultural investment into the above sustainability model raises the question of substitutability *within* forms of cultural capital. For example, is new cultural capital substitutable for old? If so, the loss of heritage items via destruction or neglect could be compensated for by the creation of new cultural assets. Alternatively, to what extent is new art capable of delivering the same level of cultural benefits as old art? And even if new art can produce the same level of benefits in some quantitative sense, there are likely to be qualitative differences in the types of value created. Furthermore cultures are not static, and some degree of decay and renewal may in any case occur, and indeed may be regarded as desirable as a means of maintaining cultural vitality over time.

Thus, whilst in principle there may be no formal difficulty in simply allowing for increments in the capital stock arising from new investment to be included in the above model, the specification of the cultural value yielded by the new capital goods presents

some problems. At one level these problems are no different from those of measuring cultural value elsewhere in the model; however, uncertainties surrounding evaluation criteria in contemporary culture may make measurement even more difficult than in the case of inherited cultural capital, when at least judgements have had time to mature and some consensus can be seen to have been reached. Moreover, there is a further problem with the introduction of new investment into the system – the difficulty of identifying how much of new artistic and cultural output will in fact add to the capital stock. In the case of built heritage, for example, a recognition of cultural significance may take some time to evolve – who is to know which modern building, large or small, will be regarded as culturally important in fifty or one-hundred years' time?<sup>14</sup> In regard to artworks, the transience of contemporary art presents a similar problem, since much currently produced art will sink without trace; only a very small proportion of works produced at any given time are likely to survive to become part of longer-term capital accumulation. In these circumstances, it is probably appropriate to regard a certain proportion – perhaps a majority – of contemporary art as consumption goods, or at least as investment goods that are fully amortised in the current period.

Overall, the approach adopted in specifying a sustainable development path in this model raises again the well-known debate about whether the intergenerational aspects of sustainable development are a matter of efficiency in intertemporal resource allocation, or whether they are matters of fairness or equity in the present generation's treatment of its successors. It might be observed that the admission of cultural value as an additional element in the picture does not change the basic propositions involved. The preservation of cultural capital for the benefit of future generations can be just as much a question of efficiency or equity in the allocation of resources producing *cultural* benefits as it is in the case of resources producing only economic return.

Finally, it can be noted that while the theoretical concept of a culturally sustainable development path defined according to explicit criteria may be an appealing one, it remains operationally constrained until robust value-assessment methods can be devised. A step in this direction might be to seek aggregate cultural indicators providing a first approximation to levels and changes in the cultural capital stock, along the same lines as Pearce and Atkinson (1993) did for natural capital. Of course this is more easily said than done; efforts to construct cultural indicators have some particular problems of their own [McKinley (1998), Pattanaik (1998)], and quantification is especially difficult because of the unavailability of suitable data on cultural resources for any country, let alone on an internationally comparable basis between countries. Nevertheless there are some hopeful signs that the documentation and measurement of cultural heritage stocks will benefit from current progress amongst official statistical agencies in improving the collection of cultural statistics more generally.

<sup>14</sup> Some contemporary architecture seems to acquire heritage status almost instantaneously, especially buildings devoted to some cultural purpose, such as the Bilbao Guggenheim Museum, the Getty Center in Los Angeles and the Sydney Opera House.

#### 4. Investment appraisal procedures applied to cultural heritage projects

We turn now to some consequences for economic analysis arising from the specification of heritage as cultural capital. In particular, we can ask how far we can go in applying conventional asset management techniques and investment theory to the assessment of cultural heritage, and how in particular the non-market benefits of heritage can be included in the analysis.

##### 4.1. Cost–benefit analysis applied to heritage decisions

As is well known, there are several methods for evaluating capital investment decisions, including the *payback method* (how long does it take for the asset's earnings to repay its initial capital cost?), the *benefit–cost ratio* and *net present value method* (do aggregate net benefits, suitably discounted, exceed the capital cost?), and the *internal rate of return method* (what discount rate just matches aggregate discounted net benefits with the initial capital cost?). There seems no reason why these methods could not in principle be applied to the appraisal of cultural capital. As we have already observed, heritage items have an existing asset value, require real resources in their maintenance, and yield flows of benefits into the future. Thus the evaluation of a heritage project involving, say, restoration of a site or of an artwork could aim to identify all the market and non-market benefits and costs involved, and then use one or other of these techniques to compare investment in this project with other competing heritage projects or with other (non-heritage) alternative investment opportunities. It is important to repeat, however, that since cultural capital is distinguished from ordinary physical capital by the cultural value it generates, evaluation methods applied to heritage projects, if they are to be comprehensive and account for the particularly cultural nature of the asset in question, should be focussed on *both* the economic *and* the cultural value of the projects under study.

How is this to be achieved? The economic calculus is straightforward: the future time stream of net benefits generated by the project, discounted to the present, can be readily compared with the project's initial investment cost, i.e. the project is economically viable if

$$\sum_t (b_{mt} + b_{nt} - c_t) / (1 + i)^t > I, \quad (10)$$

where

$b_{mt}$  are market (private-good) benefits yielded by the project in period  $t$  ( $t = 1, \dots, T$ );

$b_{nt}$  are non-market (public-good) benefits yielded by the project in period  $t$ ;

$c_t$  is operating cost of the project in period  $t$ ;

$I$  is project investment cost;

$i$  is rate of discount.

Both the market benefits and the operating costs, together with the initial investment cost, can be assessed as for any other investment project from known or assumed financial data.<sup>15</sup> The estimation of the non-market benefit component of Equation (10) is discussed in the following section.

#### 4.2. Estimation of the non-market benefits of heritage

Demand for those aspects of cultural heritage that are recognised as non-rival and non-excludable public goods can be ascribed to three sources. First, individuals may value the *existence* of a given item of cultural heritage, even though they may not consume the services of that item directly themselves. Second, they may have an *option* demand, i.e. a desire to keep open the possibility that they might consume the asset's services at some time in the future. Third, the asset may have *bequest* value, insofar as people wish to bequeath the asset to future generations. These values are sometimes described as *non-use* or *passive-use* values, in contrast to the *active-use* values enjoyed by those directly consuming the services of the asset themselves (reflected in  $b_{mt}$  in Equation (10)).

Measurement of the non-market demands for cultural heritage has made significant advances in recent years. Cultural economists have been able to take advantage of the extensive theoretical, methodological and empirical research that has gone into the evaluation of environmental amenities.<sup>16</sup> Methods used can be classified into revealed preference and stated preference approaches. Amongst the techniques relying on observation of market behaviour (revealed preference), *hedonic pricing* methods have some potential in the heritage field, but their use is limited to situations where a reasonably wide spread of market data can be found. So, for example, the influence of heritage values on property prices, including the effects of listing, might be able to be assessed by these means. An illustration is the study by Moorhouse and Smith (1994), who investigated the influence of architectural styles on the prices of nineteenth century terrace houses in Boston. Another application is that of Chanel, Gerard-Varet and Ginsburgh (1996) who used hedonic methods to analyse the auction prices of artworks. Despite the validity of such studies in terms of what they set out to do, they suffer from the fact that they essentially measure private individual benefit rather than wider public-good effects.

An alternative revealed preference approach is to use the *travel cost* method. Individuals are presumed to reveal how much they value the benefits provided by an environmental or cultural site by the amount they are prepared to pay in making the journey to visit it. Studies have been carried out on visitors to particular sites, asking

<sup>15</sup> Note that  $T$  might be unusually large in the case of heritage, since it may be expected to last indefinitely; however, of course, at any reasonable discount rate benefits in the remote future are effectively reduced to zero in present value terms.

<sup>16</sup> For an overview of environmental valuation see, for example, Braden and Kolstad (1991), Freeman (1993), Grafton et al. (2004, Chapters 8–10).



how much time they spent travelling to it, what financial outlays were involved, how many times they have visited, etc. From the results for many respondents, a demand function for the benefits of the site can be established from which, amongst other things, an estimate of consumers' surplus can be derived. However, there are a number of difficulties with this approach. Not all sites require a lot of travelling to visit them, and often multiple purposes are involved in the trip: for example, how does one allocate the cost of visiting Florence between the Duomo, the Uffizi, and Santa Croce? Furthermore, even if reasonable estimates can be made by these methods, they relate more to direct use value rather than to non-use values. Thus, the application of travel cost methods to the estimation of the non-market benefits of heritage would appear to be limited.

Because of the difficulties of finding appropriate market data, assessment of the non-market benefits of cultural heritage has concentrated overwhelmingly on *stated* preference methods using discrete choice modelling or contingent valuation methodology (CVM).<sup>17</sup> These approaches involve asking people their willingness to pay (WTP) for the benefits received, or their willingness to accept compensation for their loss (WTA). The asking may take place under quasi-experimental conditions, or more commonly may be administered through sample surveys of individuals drawn from the population of those experiencing the benefit in question. Thus, for instance, the non-use value of a local heritage site might be assessed using CVM by means of a survey of a sample of residents of the area. The survey might be conducted by telephone, mail or personal interview. The effects of various sources of bias such as the free-rider problem, the embedding effect, etc. can be reduced or eliminated by careful experimental design. Respondents might be asked hypothetically to indicate the maximum financial contribution they would make to a fund to support the site, or they might be asked whether or not they would contribute a fixed amount to such a fund. Either way an analyst could use the results to estimate a hypothetical demand function for the non-use benefits of the site in question.

The use of CVM was given some endorsement by the findings of an expert panel appointed by the US National Oceanic and Atmospheric Administration to review the applicability of the technique in determining liability claims following the *Exxon Valdez* oil spill in Alaska in 1989. The panel, co-chaired by Kenneth Arrow and Robert Solow, and including Edward Leamer, Roy Radner, Paul Portney and Howard Schuman, found that "CV studies can produce estimates reliable enough to be the starting point of a judicial process of damage assessment, including lost passive-use values" [Arrow et al. (1993)], provided that they are carefully carried out, with due attention paid to the biases and other problems affecting the technique.

Overall, it can be said that since the non-use values from cultural heritage are very similar in kind to those arising from environmental amenity, the fact that many successful applications of CVM in the environmental sphere have now been carried out augurs well for the further application of these same techniques in the heritage field.

<sup>17</sup> For overviews of the method and its application in the cultural heritage field, see Frey and Oberholzer-Gee (1998), Klammer and Zuidhof (1999), Navrud and Ready (2002) and Noonan (2003).

### 4.3. Cost–benefit analysis and cultural value

The above procedures can provide a comprehensive evaluation of the economic effects (including both market and non-market effects) of this investment project. But we noted above that a further step is required if a full account of the cultural benefits and costs of the project is to be provided, i.e. a form of cultural cost–benefit analysis in which the cultural value created by the project is the subject of appraisal. How is this to be done? In principle there is no reason why an approach similar to that shown in (10) above could not be applied also to the cultural value component of this investment. The project is expected to yield a flow of cultural benefits into the future; indeed it is the prospect of this flow, more than of the economic return, that is likely to have given rise to the project in the first place. Furthermore, cultural benefits can accrue both to individuals (as rival excludable benefits) and to society at large (as non-rival non-excludable benefits), suggesting that a division into private and public components is just as relevant to the assessment of cultural value as it is for economic effects. Moreover, it is reasonable to suggest that a unit of cultural benefit at some time in the future is worth less to the project's stakeholders than a unit now, making some form of discounting of future benefit streams appropriate.

But the problem remains of how these benefits are to be measured. Given the multi-dimensional character of cultural value, it can be suggested that the only sensible way to evaluate it will be via a disaggregation into its component elements. Throsby (2001) has suggested that, for cultural heritage, those elements might include

- *aesthetic value*: beauty, harmony;
- *spiritual value*: understanding, enlightenment, insight;
- *social value*: connection with others, a sense of identity;
- *historical value*: connection with the past;
- *symbolic value*: objects or sites as repositories or conveyors of meaning;
- *authenticity value*: integrity, uniqueness.

It may be possible to assign cardinal or ordinal scores to these components and aggregate them into a simple index according to given assumptions. If so, the flow of cultural benefits from alternative projects might be able to be compared. A limited step in this direction is provided by the example of Nijkamp (1995), who put together a composite index of different characteristics of historic urban districts evaluated according to a range of criteria. Alternatively, this set of characteristics can be placed in a broader context of public policy-making where the cultural value of the arts and culture is seen as a valid reason in its own right for government intervention [Holden (2004)].<sup>18</sup>

Nevertheless, it has to be recognised that comprehensive and replicable methods for assessing cultural value in these terms remain to be developed and applied. In the meantime, evaluation procedures derived from economic analysis have much to offer. Even if

<sup>18</sup> See also Ellis (2003) and the RAND Corporation report by McCarthy et al. (2004) which identifies the intrinsic benefits of the arts as a neglected area in contemporary policy-making with its stress on the instrumental (economic) benefits of culture; for a multi-disciplinary collection of essays on aspects of cultural value see contributions to Hutter and Throsby (2006).

they cannot capture those elements of cultural value that are not expressible in monetary terms, or that cannot be factored out to individual utilities [Throsby (2003)], they can at least provide important insights into the value of cultural heritage to individuals and to society, and make an important contribution to heritage policy-making, an area that we turn to in the following sections of this chapter.

## 5. Public intervention in heritage markets

Following the conventional normative approach of welfare economics, a long-sighted government adopting a public interest stance is assumed to provide efficient remedies for market failure through the use of the different tools of government intervention. In the case of heritage, market failure is presumed to arise through the existence of public goods, externalities, asymmetrical information, etc. and these arguments are usually put forward to justify public intervention aimed at avoiding under-provision of heritage services.

Government action to correct for market failure in the heritage area can follow different patterns involving direct or indirect intervention and using instruments with monetary or non-monetary content. Public spending, taxation and regulation are the tools usually adopted to promote heritage conservation, together with other wider instruments such as education, which can be used to spread information and improve citizens' awareness. These tools are summarised in the following table:

	Monetary	Non-monetary
Direct	Public expenditure	Hard regulation
Indirect	Tax expenditure	Soft regulation

*Public expenditure* involves both direct funding for the maintenance, operation and restoration of publicly-owned heritage and museums, and subsidies to private and not-for-profit entities. *Tax expenditure* refers to the provision of tax incentives to private individuals or firms who spend money on maintaining or restoring their own heritage properties or who donate funds to heritage causes.<sup>19</sup> There is a crucial difference between direct and indirect financial support; in the former case decisions regarding the size and the composition of heritage support (for instance, which monument should be restored or which form of art should be supported) are taken by the public decision-maker, while in the case of indirect financial support through the tax system, funding decisions are private and the policy outcome is likely to differ. Frey (2003), looking at the overall arts sector (including performing and visual arts) argues in this respect that private decisions might support a broader range of artistic activities than public

<sup>19</sup> See further Chapter 36 by Schuster in this volume.

ones, including those activities that are more controversial and experimental, since public decision-makers tend to shy away from scandalising public opinion.<sup>20</sup> However, this might not be the case when dealing with heritage; private decisions might be driven by the prestige of the monument or the institution gaining support, thus directing intervention toward well-known objects and institutions. Moreover, any comparison depends also on the functioning of the public decision-making process, whether it is supply-oriented or demand-oriented and how powerful is lobbying in influencing it, as will be discussed further below.

*Regulation* refers to restrictions or modifications imposed on the activities of economic agents in line with government policy objectives. In the heritage case, regulation is aimed at controlling the stock of heritage, its size and its quality. Following Throsby (1997, 2001) we can distinguish *hard* and *soft* regulation; the former involves enforceable acts, such as listing, permissions, authorisations, demolition orders, standards and penalties for non-compliance, while the latter consists of non-enforceable directives (charters, codes of practice, guidelines, etc.) implemented by agreement and not involving penalties. In the heritage field regulation has certain advantages compared to other government tools. Its adoption or removal takes less time than is required for other forms of public intervention such as subsidies or tax expenditures, and therefore it allows for a greater timeliness of public action. Flexibility, in fact, can be extremely useful in heritage to cope with the necessity for quick decisions and to avoid the infinite costs linked to irreversible decisions; such flexibility might be exercised, for example, in the rapid imposition of an order preventing the demolition of a listed building until such time as a full evaluation of the consequences has been undertaken.

However, some specific problems arise when regulation is applied to the heritage field. First, the identification of the scope and the range of intervention is a matter of discretion to a greater extent than in other fields of public policy because the definition of heritage is not straightforward. Moreover, such discretion varies according to the different types of heritage. For outstanding heritage with characteristics of uniqueness, such as items on the World Heritage List, no discretion problems arise and any decision about conservation comes of necessity under the scrutiny of public opinion. On the other hand, for minor heritage the extent of regulation as well as the type of intervention to be carried out might be by no means so clear-cut. In such a case, the identification of the scope of intervention is a matter for discussion based, for example, on evaluations made by hired experts who are well informed but may have contrasting views on orders of priority. As a consequence, the type of expert (archaeologist, art historian and so on) involved in this kind of decision, and the features of the decision-making process, are likely to be quite important in determining the size and the composition of the stock of cultural heritage, as well as the type of conservation that can take place.

Listing of heritage, whether based on local, national or international significance, is a regulatory instrument existing in most countries. Lists can be a useful source of

<sup>20</sup> On whether governments might prefer to avoid supporting arts which can give rise to controversies, see Zolberg (2000).

information concerning the social value of heritage and might offer the opportunity for voluntary collective action to preserve the listed items [Schuster (2004)]. However, whether the provision of information in this way is of itself an effective tool to stimulate individuals' and groups' concern for heritage is an open question. Indeed, there is little evidence in this respect, apart from observations of the stimulus exerted by the UNESCO World Heritage List in leading to improvements in visitor facilities and in cultivating regional and national pride. In practice listing, especially when coupled with other government actions, is not neutral. It affects prices as well as market decisions and may have relevant distributional implications because of the costs and benefits it generates.

The economic impact of regulation on the heritage market varies also according to the types of costs it imposes. In addition to the administrative and bureaucratic costs related to the drawing up of regulatory legislation and the monitoring of its effective implementation, regulation generates compliance costs imposed on all those private and public actors who have to comply with the prescriptions. Some of these costs can be foreseen in advance because they are closely connected to the conservation task (for instance, the requirement to use special materials and qualified operators to ensure quality of maintenance or restoration works); others are subject to a high degree of uncertainty as a consequence of the regulator's decisions. In the latter case, a modification to a building allowed by urban planning authorities might not be permitted by the heritage regulator [Rizzo (2003)]. Moreover, whenever heritage involves a combination of different styles and historical periods, as often happens in Italy, any discovery made while the work is in progress may reduce the scope for planned use. In such a situation the heritage authority may give priority to the preservation of the new discovery and as a consequence will interrupt the work and/or will impose changes compatible with the new discovery [Pignataro and Rizzo (1997)]. Finally, there are the indirect costs imposed on any activity that interferes with heritage; for instance, [Peacock (1995)] refers to the considerable hidden costs imposed by planning regulations, for example, those requiring the diversion of roads to protect archaeological sites. As will be pointed out below, these costs should not be disregarded and indeed some form of evaluation of the economic consequences of experts' decisions is highly desirable.

It is clear that no single policy tool is likely to be successful on its own in giving effect to heritage policy; rather what matters is how different instruments are combined and which level of government is entitled to use them. In many cases, a given tool can be used as a complement or a substitute for others. Complementarities arise between public subsidies and regulation when private cultural activities are publicly funded; for example, a private owner receiving public financial support for restoring a historic building might be compelled to carry out the project according to precise rules and may be obliged to allow for public visits. On the other hand, regulation and public funding can act as substitutes when a public activity related to heritage is privatised; for instance, if a publicly-owned building is sold to the private sector to be re-used, regulation can be applied to ensure that government objectives in regard to the building's heritage quality are fulfilled.

## 6. Public-choice approach to heritage decision-making<sup>21</sup>

In a principal-agent model of a representative democracy, politicians are construed as the agent of citizens and, in turn, the principal of the bureaucracy. Given that each actor pursues the maximisation of its own utility (politicians want to be re-elected and bureaucrats want to better their careers), the outcome of the public decision-making process in terms of society's well-being will depend on its institutional design, i.e. on the incentives imposed on the various agents. In the heritage case, uncertainty about the definition of cultural heritage and the lack of voters' information weaken the control on political representatives; also politicians may suffer information shortage because of the specificity of the knowledge and expertise required to understand heritage issues.

The institutional features of the decision-making process in the heritage field may vary. Delegation from politicians to bureaucrats can be *complete*, i.e. implemented with the institution of independent agencies, or *incomplete*, in the sense that politicians and bureaucrats interact, where the bureaucrat is assigned some goals by the politicians and chooses the tools to fulfil them [Alesina and Tabellini (2004)]. Moreover, different levels of government may be involved in various ways. Nevertheless, in most jurisdictions it seems realistic to describe the policy-making process as one in which politicians and bureaucrats interact, with the definition of goals and the decisions concerning the financing of cultural policies being determined at the political level, while the implementation of the policies is delegated to bureaucrats.

Typically decision-makers enjoy a high level of discretion, but face conflicting demands for conservation. In these circumstances there is room for bargaining and for rent-seeking activities. How the different interests will be represented and satisfied will depend on the institutional design, especially on the mechanisms used to prevent the opportunistic behaviour of agents, reducing their information advantage. It can be reasonably argued that heritage conservation is an avenue of government expenditure that enjoys general public support.<sup>22</sup> Moreover, an ageing society is likely to support an increase in heritage conservation, especially in a globalised context when cultural identities can be perceived to be at risk [Benhamou (2003)]. These various demands are likely to be stronger the less visible the opportunity cost of conservation. Public opinion can be considered as being subject to a form of "fiscal illusion" with respect to heritage conservation, especially when regulation such as listing is used. This proposition is based on the assumption that conservation generates mainly benefits for the community; costs, if any, are concentrated on the owners of listed buildings and/or those otherwise directly affected by the regulation.

Other issues are relevant in shaping the decision-making process. If heritage deteriorates or is misused, bureaucrats will suffer. Because monitoring of bureaucrats by the

<sup>21</sup> For an extensive analysis of the public choice approach see Mueller (2003); an overview of the implications of public choice literature for the heritage case is provided by Mazza (2003)

<sup>22</sup> This might be inferred, for example, from more general evidence for public approval of government support for the arts and culture; see Frey (2003).

politicians will be incomplete, they are subject to moral hazard and will therefore have an incentive to adopt a more cautious (i.e. conservationist) stance. Therefore they will seek to minimise the risk of irreversible losses, such as those caused by demolition or radical transformation of a building, and will oppose the use of historical or archaeological sites or buildings for mass events such as rock concerts, which can lead to the deterioration of heritage. The term “conservationist” refers also to a more general attitude leading to the unjustifiable enlargement of listing and to the implementation of very strict requirements for conservation, well beyond the correction of externalities; as a consequence the full enjoyment and utilisation of heritage can be endangered, for instance when restrictions prevent alterations needed to bring the interior of an old building up to modern standards of comfort. Bureaucrats will thus have an incentive to act in a very restrictive way. Such an incentive will be even stronger if the existing legal system makes them liable for any damage the heritage may suffer from the activities carried out on the site [Rizzo (2003)].

Thus in the heritage field the range and intensity of regulation appears to be not simply a policy instrument but rather an endogenous product of the public decision-making process [Rizzo (2003)]; from this perspective, it is not without significance that the large discretionary power enjoyed by regulators is coupled with the widening scope and extent of heritage observed in some countries.<sup>23</sup> It is worth noting also that a sustainability issue arises from the enlargement of the concept of heritage because “delisting” proves to be highly unlikely. These considerations may help to explain why economic advice tends to find a greater “barrier to entry” into the decision-making process in the heritage conservation field than in other areas of cultural policy [Peacock (2004)].

Similar arguments can be developed with respect to the international circulation of works of art [Giardina and Rizzo (1994)]. The alleged normative rationale for regulating the international movement of works of art is the preservation of the visual arts and the protection of national identity and prestige. However, doubts arise as to the effectiveness of such a policy tool, since excessive regulation might induce collectors and/or dealers to leave the official economy and to undertake their exchanges in the underground sector; the likelihood of such an event is greater the smaller the risk of punishment, which is in turn a function of the amount of resources government allocates to monitoring activities. A different argument can be made to question regulation preventing the sale of the works of art belonging to public museums. Many museums exhibit only part of their stock;<sup>24</sup> scarce resources are mainly used for conservation, and there is no room for an acquisition policy consistent with the specific artistic vocation of the museum. If museums (regardless of whether they are privately or publicly owned) were allowed

<sup>23</sup> Benhamou (2004) reports that the number of listed historic monuments in France more than doubled between 1962 and 1999 and that the same happened in the UK in the period 1990–2000.

<sup>24</sup> According to Grampp (1996, p. 225), many museums exhibit 10 percent or less of the objects they have in their collections, including the Art Institute of Chicago (3 percent), the Hermitage in St. Petersburg (7 percent), the Prado in Madrid (9 percent) and the Alte Pinakothek in Munich (10 percent). An exception is the Louvre in Paris which is reported to display well over 50 percent of its paintings.



to sell or to exchange works of art that are not strictly connected with the museum's artistic interests, they could use the proceeds to finance new acquisitions, thereby enriching the existing collections.<sup>25</sup> Mutatis mutandis a similar argument applies in the archaeological field; some museums which collect material from an archaeological site store thousands of identical items (small amphorae, pottery, etc.), none of which has peculiar features. Once an adequate number of them is kept, the others might conveniently be sold and the proceeds used to finance either the proper conservation of the museum exhibits or the acquisition of new items. A good example is provided by the Pompeii archaeological site, where thousands of identical amphorae have been found and are currently stored while new discoveries or the conservation of the existing ones are constrained because of budgetary stringency. In these circumstances protectionist arguments for regulation cannot be justified on efficiency grounds.

It is apparent from the above discussion that institutional design may have a perverse impact: if the decision-making process is mainly affected by bureaucrats/experts' interests, it is likely to bring about consequences that contrast with the claimed objectives of conservation policy. Even if the society's demands might *seem* to be satisfied, this depends on the low visibility of conservation costs and on the general issue of asymmetrical information which does not allow for a clear perception of sustainability effects. The extent of such an argument depends upon the degree of autonomy experts are granted and on the incentive schemes faced by the bureaucrats. Therefore, devices are needed to improve taxpayers' influence on public decisions, bearing in mind that these decisions will still need to rely on experts.<sup>26</sup> One way of achieving this would be to follow Peacock's (1994) suggestion that public participation could be enhanced if greater openness were to characterise the appointments of "lay" persons to decision-making bodies and if citizens who are active in heritage matters were allowed to vote for their own representatives within these bodies. Moreover, to fulfil the same purpose, compulsory consultation and review procedures might be introduced into the public decision-making process. However, the benefits of a greater political participation should be weighed against the likely increase in administrative and time costs which would derive from it.

From this perspective a relevant argument is that devolution, as Schuster (1997) defines it, increases the accountability of government and allows for a better control of the decision-making process. In the heritage case this implies that lower levels of government are entitled to use all the means – regulation, expenditure and taxation – available to pursue the objectives of heritage conservation and, as we shall show in the next section, this can have an impact on the policy outcome.

<sup>25</sup> See further Chapter 29 by Frey and Meier in this volume.

<sup>26</sup> The need for a governance structure to define a mechanism to restrain the discretionary scope of regulators is common to regulation in general [see Levy and Spiller (1996)].



## 7. Devolution in the formation and implementation of heritage policy

### 7.1. *Advantages and disadvantages of devolution*

The normative rationale for devolution is well known: the allocation of functions among the various layers of government should take account, among other things, of the geographical coincidence between taxpayers and beneficiaries of a given good or service. Such a conventional argument acquires specific relevance in the heritage field. The scope to be assigned to devolution in this field is likely to be larger than is usually claimed, because spillover effects are internalised [Rizzo (2004)]. Sub-central levels of government in fact will promote heritage conservation even when the geographical coincidence does not hold; the more important a particular item of heritage and the wider its reputation, the greater are the external benefits related to its existence and the beneficial impact it can exert on tourism and, as a consequence, on the local economy [Brau, Lanza and Pigliaru (2003), Cellini (2004)].

However, to rely mainly upon tourism motivation in defining heritage policies might create problems; apart from congestion, it might bias policy, leading to undervaluation of other long-term benefits related to cultural value but less visible in the political market. For instance, such undervaluation might affect policies directed at the education of younger generations in order to raise their appreciation for the arts; if these policies are successful, the need for public support to the art sector could eventually be reduced. At the same time, if the rationale of economic development promoted by tourism prevails, devolution might lead to a concentration of resources on heritage with higher economic return, penalising minor heritage that may be closely connected with local identity and history but of lesser economic value [Mazza and Rizzo (1998)], although such an effect might be mitigated by the presence of voluntary local associations devoted to preserving minor heritage because of its relevance to the local community. A further issue is that stressing the tourism motivation can give rise to other problems such as the imposition of values that conflict with local tradition and identity, a problem that can be significant in developing countries. Again, the relevance of decision-making process and of the motivations underlying it comes into the picture.

A positive aspect of devolution is that it may be helpful in overcoming or reducing the sorts of asymmetries in information discussed in the previous section. In general it can be argued that devolution improves citizens' information by allowing comparison among different alternatives, and also that the existence of many agents improves the principal's information.<sup>27</sup> The issue acquires specific relevance in the heritage case where public policy can offer a variety of solutions; hence comparison may turn out to

<sup>27</sup> The extreme consequence of this is described by the well-known phenomenon of people "voting with their feet" [Tiebout (1956)]; i.e. rational individuals exit communities which offer less attractive packages of taxes and local public goods to move to others offering more attractive packages.

be useful. Moreover, there is every reason to believe that sub-national tiers of government will be able to interpret the need for conservation better than the central government, given that the link between heritage and citizens is closest in the community in which the heritage is located.

This issue is relevant in multicultural societies where local governments can be considered more suitable for providing services for different ethnic groups and for promoting the participation of minorities [Fondazione Eni Enrico Mattei (2003)]. It would be interesting to investigate how such participation, coupled with the extension of the right to vote to immigrants and ethnic minorities, would affect the outcome of the political decision-making process as far as local heritage conservation is concerned. In multi-ethnic societies the heterogeneity of preferences weakens the assumptions behind fiscal federalism. Oates' theorem demonstrates that heterogeneity within a specific community does not necessarily ensure the superiority of a decentralised provision of public goods and services in terms of social welfare. Equally it has been demonstrated [Bridgman (2004), Alesina, Baquir and Easterly (1999)] that, when there is majority voting, ethnically heterogeneous jurisdictions tend to provide smaller amounts of public goods than homogeneous jurisdictions. These effects are more pronounced in the heritage case because in a multicultural society the concept of local identity is more controversial than in a homogeneous society; as a consequence, if resources are devoted to promote local heritage as a symbol of local identity, such a decision is costly for those who do not recognise themselves in such an identity and have other priorities in terms of local public goods. Moreover, stressing local identity through heritage might reduce rather than strengthen integration of multicultural minorities. Overall it is not clear whether the net effects of decentralisation in this case will be positive or negative.

A final advantage of devolution is that the monitoring of bureaucratic behaviour is likely to be simpler and easier at the local level (assuming the system favours political accountability), since the opportunity costs of bureaucratic decisions can be more readily observed. The fact that bureaucrats can be monitored might make it easier, for instance, to adopt at local level codes of practice or guidelines agreed between the regulator and those involved in conservation activities (architects, building firms, engineers, cultural associations, etc.); such procedures reduce the uncertainty related to investments in heritage conservation.

## *7.2. Practical issues*

We turn now to some practical issues relating to devolution. The implications of devolution for heritage policies can be quite different according to the kind of heritage existing in the region. Consider, for example, the case of outstanding ("superstar") heritage items. For such heritage the choice between devolution and centralisation is not crucial because, as noted above, the unique characteristics of the heritage affect the decision-making process in similar ways regardless of the level at which policy is made. In addition, financial constraints are likely to be less stringent in the case of outstand-

ing heritage even if it is located in a poor region, because private sponsorship will be more willing to provide financial support for the conservation of this heritage than for less well-known heritage. At the same time, in a devolved context the existence of outstanding heritage might lead to complementary policies for the conservation of minor heritage within the same area, using the attractiveness of the former to promote income-generating uses for the latter. Policies tend to be easier to promote if decisions are taken by one decision-maker, i.e. the sub-central level of government, while transaction costs will be higher if an agreement has to be reached between different layers of government, each with different responsibilities. An interesting example is the case of City Museums in Venice where network economies are generated: a visit to the Palazzo Ducale, which is an item of state heritage and whose management has been assigned to the city, is offered in a package with a visit to the lesser museums in Piazza San Marco, with the positive result that the local government “makes profits” out of them [Zan (2001)].

Devolution might also lead to enlargement of the range of what is perceived as “culture” in particular locations. For example, visitors to a city or region might be offered a package including visits to minor sites and museums, attendance at performances which take place in historic buildings or archaeological sites, visits to handicraft ateliers, and so on. Different types of itineraries can be designed around a leading theme. To illustrate, in Italy at the regional level “wine routes” are designed to present unfamiliar heritage associated with specific social and economic functions, illuminating local history and producing an impact on the related economic activities. Another example is the so-called “literary parks”, where an itinerary is created to illustrate the work of famous writers and to experience the places that gave birth to them and inspired their masterpieces. In these sorts of cases, horizontal cooperation among local governments is usually needed and the free-rider problem is likely to be overcome by the mutual interest in gaining the benefits deriving from a joint cultural supply. Devolution matters because it may contribute to the creation of a “virtuous circle”, developing a common strategy among different partners such as local governments, economic and cultural operators, crafts makers, associations and so on which can be helpful in supporting minor heritage, enhancing its capability of providing income-generating services.

A further argument in support of devolution relates to the potential conflict which may arise between different levels of government when a central or regional regulatory power impacts upon urban policy carried out at local level. For example, local government decisions on matters such as urban renovation may be constrained by regulation established and implemented at a regional level [Mazza and Rizzo (2000)]. If so, a conservation decision taken at the regional level will produce costs and benefits for a set of voters that are relevant for politicians at the local level but are not taken into account by the politicians at the higher level of government. As a likely consequence, self-interested regional policy-makers will find it convenient to adopt strict rules, since the social costs will be borne only by local owners within the local jurisdiction which is not relevant for them, and the benefits will be spread out over the whole population.

### 7.3. Summary

To sum up, we have argued that the issue of devolution acquires special relevance in the case of minor heritage. The argument rests on the assumption that at local level a better representation of citizens' preferences is likely to occur as well as better scrutiny of public decisions, which can be expected to induce decision-makers into taking account of the opportunity cost of their choices. Moreover, at a devolved level it is easier to identify those who gain and those who lose from regulation; members of the latter group may have more room to organise themselves, acting as watchdogs rather than being only passive adjusters to heritage authority decisions made higher up. This does not mean that devolution will necessarily provide optimal solutions, but only that the decision-making process will be less supply-oriented. On the other hand, the opposite argument, invoking central decisions to reduce the role of local pressure groups opposing conservationism, leaves unanswered the sustainability question and is less likely to satisfy the society's demand for conservation.

Nevertheless, the findings of recent research on heritage policies in Sicily [Rizzo and Towse (2002)] indicate that unless an adequate incentive system is introduced in the decision-making process, devolution as such is not enough to bring about improvements in the accountability and responsiveness of heritage authorities to public opinion or the introduction of better economic management of cultural institutions. Thus, despite devolution, museums in Sicily have not improved their performance, paying little attention to the needs of visitors, with limited use of multimedia information and a lack of strategic cooperation among different institutions at the local level. In addition, public opinion has not been mobilised, with few voluntary associations such as "museum friends" compared to other Italian regions. A possible explanation may be found in the institutional features of the Sicilian case; lack of real fiscal autonomy coupled with a proportional voting system have led to low levels of accountability of regional government in Sicily and high political instability, weakening the instruments of control over bureaucracy. Although political and administrative reforms have been recently introduced to improve political and bureaucratic accountability in this region, it is not clear whether they will bring about any change in the management of heritage and museums.

## 8. The role of the private sector

### 8.1. Effects of regulation

The role of the private sector in heritage conservation can be analysed from different perspectives: individuals and firms are private investors when they are owners of heritage and are directly involved in conservation; at the same time, other private agents contribute with donations and sponsorships to heritage projects of various sorts. The behaviour of these private participants in cultural heritage is based on a variety of motivations and is affected by whatever mix of policy instruments might be chosen.

Regulation affects private decisions to invest in heritage conservation and therefore shapes the public–private mix in the heritage market. If regulatory decisions are taken adopting a conservationist stance, private investors will face uncertainty and higher costs and are likely to be discouraged. The extent of such an effect, however, depends on the policy mix, i.e. whether direct or indirect public spending is coupled with regulation. If public financial support is indeed to be provided to the owners of listed buildings, the question arises as to how it should be designed. Direct spending leaves more room for discretion for the decision-maker to identify the eligible beneficiaries, while tax allowances are automatically enjoyed by all the owners of listed buildings. As far as effectiveness and equity are concerned, the effects will depend on how these tools are actually shaped. For instance, subsidies might be in the form of matching grants, tax allowances could be designed as tax exemptions or deductions, and so on. In regard to equity, tax allowances are generally likely to favour rich taxpayers compared with direct spending.

Empirical evidence allowing a comparison to be made between different countries is difficult to find. It is not so easy to isolate the impact of regulation from other policy measures. Feigenbaum and Jenkinson (1984) found that in the United States both the grants-in-aid and tax credit programs had a positive effect on per capita preservation expenditure and no major differences were found in the effectiveness of these tools. Benhamou (2004) points out that in France an extensive system of financial support does exist for private expenditure on conservation; overall, private owners of listed buildings benefit from listing, as shown by the difference in value between listed and non-listed buildings.<sup>28</sup> This differential is smaller in the United Kingdom because, whereas French regulation concerns only visible parts such as the façades, in the United Kingdom it refers to the entire building. No evidence exists for Italy as far as the impact of heritage policy on prices is concerned, but it can be presumed a priori that a disincentive effect is likely to exist.

## 8.2. *Sustainability in heritage financing*

Turning to the sustainability issue, we note that the likely unintended consequence of a conservationist stance might be the crowding out of private investment for conservation, if public spending is not directed toward compensating owners for the financial burden involved. To what extent this occurs is an empirical matter but the effect would be to reduce the overall amount of available resources for heritage conservation. On the other hand, if private investment is not crowded out by public spending, a conservationist stance will produce increasing pressure on public funds because, as noted above, listing is likely to expand; whether such pressure is sustainable depends among the other

<sup>28</sup> If, as appears to be generally the case, listing increases the market prices of buildings, owners benefit and might therefore be expected to bear a greater share of the cost of conservation; see discussion in Benhamou (2004) and Creigh-Tyte (2000).

things on the stock of heritage as well as on the strength of financial constraints. If the size of the stock is huge relative to GDP, a sustainability issue arises; heritage is likely to deteriorate, which is the opposite effect to that desired. This effect is likely to be self-perpetuating, producing increasing pressure on public funds and further decay. The same negative result is reached if private owners, because of the stringency of regulation, undertake their activities without complying with existing rules; the likelihood of this outcome is greater the lower the owners' risk of punishment. In all these situations, effective monitoring of the impacts of regulation is difficult because of the amount of resources required.

The above considerations suggest that no unique conclusion can be drawn as to the desirable path in the policy mix, since the solution will depend crucially on particular countries' features. [Netzer \(1998\)](#) offers an emblematic example comparing USA and Italy with respect to the ratio of annual heritage capital consumption to GDP, reaching the conclusion that in Italy the costs required to maintain the capital stock are unaffordable. He estimates that in Italy the capital consumption allowances would amount to nearly 35 percent of GDP while in the USA it would be only 0.05 percent of GDP. In fact, total public spending on culture in Italy comprises a low percentage of GDP (0.67 percent), which is around the average of OECD countries; almost half (45 percent) of such expenditure is devoted to heritage rather than to other cultural goods [[Pasquali \(2003\)](#)]. Thus Italy's public effort, though financially comparable with other countries, is inadequate to the heritage conservation task, calling at least for the introduction of an opportunity cost criterion into the decision-making process, whereby any decision in regard to heritage should take into account the costs and benefits stemming from different stances (conservationist vs. non-conservationist).<sup>29</sup> However, the public decision-making process in Italy has not been designed to take into account such a constraint; indeed, it offers an example of the "fiscal illusion" issue mentioned above. Listing applies only to private heritage while until recently any public item older than 50 years was considered by definition part of heritage, to be conserved regardless of its condition. Only recently has the possibility been introduced of selling these latter items provided that, after a complex bureaucratic procedure, heritage authorities agree that the item to be sold lacks any historical or artistic value and impose only weak constraints on the buyer in terms of allowed uses. Public opinion as well as experts have reacted strongly against such a possibility, claiming that Italian heritage was at risk but without suggesting alternative ways to find resources for the prevention of heritage decay. Interestingly enough, the relevant law was originally proposed by the Ministry of Treasury, with the clear objective of raising money from the sale of public property; a different and less conservationist mentality within the heritage authorities might have allowed for a more coherent policy based on clear cultural priorities and on more precise guarantees [[Rizzo \(2005\)](#)].

<sup>29</sup> However, as has been pointed out in Section 6 above, this is no easy matter.

How conservationist should public policy be? If a conservationist stance makes it difficult to pursue income-generating uses, it might have a negative impact on the conservation of historical districts etc. and therefore on the promotion of sustainable local economic development. The arguments put forward above would suggest that the larger the size of heritage stock, the less conservationist public policy should be. On the other hand, it might be argued that if attention is paid to other potential sources of support for heritage, such as forms of supra-national intervention or of international philanthropy, a conservationist stance might be perceived as a signal of quality, thereby creating the impression that the support is warranted. However, the extent of such an effect in reality does not seem significant enough to justify a reorientation of domestic heritage policy.

### 8.3. *Philanthropy and heritage*

We turn finally to the role of the private sector in the areas of philanthropy, sponsorship and voluntary action in support of heritage.<sup>30</sup> To encourage private giving in its various forms, governments mainly provide indirect support to individuals and business, via tax exemptions; however, direct support is also sometimes involved, as occurs when matching grants are given to arts institutions on condition that an equivalent amount of private support is raised. Public policies follow different patterns in different countries [Mussoni (2003)]. In some countries, such as the United States, the United Kingdom and recently Italy, equal tax treatment is accorded to patronage and sponsorship, while in other countries, such as France and Germany, tax treatment differs. Moreover, most countries, including the United States, the United Kingdom, France and Germany (but not Italy) provide tax incentives not only to monetary giving but also to in-kind giving. Of course, different tax devices have different effects. A tax system which is neutral with respect to the donor/sponsor choice is likely to strengthen private influence on cultural decisions; this may be perceived as a challenge to the development of culture as the autonomous outcome of the democratic collective decision making process, insofar as the control over the use of resources (of voters through their representatives) might be partially given up because private donors/sponsors would decide priorities. In the specific heritage field, as noted above, private decisions might be driven by the prestige of the monument or the institution concerned; for instance, in Italy empirical evidence shows that private contributions have privileged well-established institutions to pursue highly visible restoration activities.

The different policy measures in different countries give rise to varying patterns of private giving. As far as corporate support for the arts is concerned, Kirchberg (2003) notes that differences, though still consistent, have reduced in recent years. Mazza (1994) suggests that a possible explanation of differences in private giving among countries may lie in the fact that when the provision of cultural goods is primarily a government responsibility, social recognition for donations is low; therefore, sponsors

<sup>30</sup> See further Chapter 36 by Schuster and Chapter 37 by Katz in this volume.

will not have sufficient economic incentive to support the arts, and only patrons, who are not interested in social approval, will provide their support.<sup>31</sup> Such an argument is in line with empirical evidence in Italy, a country in which traditionally heritage conservation is considered a public duty; [Mussoni \(2003\)](#) reports that the overall amount of private giving in the first two years of the implementation of a new law encouraging private cultural support has been more modest than expected. Moreover, only 37.4 percent of private contributions have been directed to heritage while 62.6 percent have been devoted to support for the performing arts.

The United States case is widely taken as an example to stress how social norms or shared social recognition of the importance of the arts provide the fertile humus in which tax incentives act as effective tools to stimulate private contributions. This does not mean that the same tax scheme will exert similar effects everywhere because private support in whatever form is the result of a complex array of elements specific to each country.

## 9. Conclusions

In this chapter we have endeavoured to show how economic theory and public policy analysis can illuminate decision-making relating to cultural heritage. Several themes run through our discussion. First, we have argued that from an economic point of view the appropriate conceptualisation of heritage is as capital. In theoretical terms this opens up rich possibilities for analysing the productivity of heritage, and for formulating strategies for heritage investment; the distinctively cultural character of the capital assets under consideration adds a specific and challenging dimension to these analyses. In policy terms, regarding heritage as a capital asset places it alongside other items of the economy's capital stock for which governments have some responsibility, and conditions the application of the principles of public finance to this area.

Second, our concern throughout has been with sustainability. Our theoretical discussion has been particularly concerned to formulate conditions for defining sustainable development paths for cultural heritage. These considerations have informed our treatment of heritage policy, where maintenance of the cultural heritage stock has been seen as one of the critical issues facing policy-makers in many countries today.

Third, we have stressed that efficient institutional structures are an essential element in delivering sustainable policy outcomes. Questions of the appropriate mix of policy instruments, the allocation of functions between levels of government, and encouraging a positive role for private sector participation are all matters that are facilitated by good institutional design.

Finally, to come back to the beginning, we can reiterate the importance of information, especially via education in the arts and culture field. Public policy supporting

<sup>31</sup> For further discussion of sponsorship, see [O'Hagan and Harvey \(2000\)](#) and [Leclair and Gordon \(2000\)](#).



heritage can provide long term educational benefits, raising social appreciation for the arts and perhaps eventually reducing the need for increased public subventions in this area.

## References

- Acheson, K., Maule, C. (2004). "Convention on cultural diversity". *Journal of Cultural Economics* 28 (4), 243–256.
- Alesina, A., Baquir, R., Easterly, W. (1999). "Public goods and ethnic divisions". *Quarterly Journal of Economics* 114 (4), 1243–1284.
- Alesina, A., Tabellini, G. (2004). "Bureaucrats or politicians?" CESifo, Working Paper Series No. 1130 Harvard Institute of Economic Research Discussion Paper No. 2009; IGIER Working Paper No. 238.
- Arrow, K.J., et al. (1993). "Report of the NOAA panel on contingent valuation". *Federal Register* 58, 4601–4614.
- Australia ICOMOS (1999). "The Burra Charter". See <http://www.icomos.org/australia/burra.html>.
- Barbier, E.B., et al. (1994). *Paradise Lost? The Ecological Economics of Biodiversity*. Earthscan, London.
- Becker, G.S. (1964). "Human Capital". Columbia University Press, New York.
- Benhamou, F. (2003). "Heritage". In: *Towse (2003, pp. 255–262)*.
- Benhamou, F. (2004). "Who owns cultural goods? The case of built heritage". In: *Ginsburgh (2004, pp. 187–202)*.
- Bourdieu, P. (1986). "Forms of capital". In: Richardson, J.G. (Ed.), *Handbook of Theory and Research for the Sociology of Education*. Greenwood, New York, pp. 241–260.
- Braden, J.B., Kolstad, C.D. (Eds.) (1991). *Measuring the Demand for Environmental Quality*. North-Holland, Amsterdam.
- Brau, R., Lanza, A., Pigliaru, F. (2003). "How fast are the tourism countries growing? The cross-country evidence". FEEM Working Paper No. 85.2003.
- Bridgman, B. (2004). "Multiethnic democracy". Departmental Working Paper 2004-02, Dept. of Economics, Louisiana State University, Baton Rouge.
- Cannon-Brookes, P. (1996). "Cultural economic analysis of art museums: A British curator's viewpoint". In: *Ginsburgh and Menger (1996, pp. 255–274)*.
- Cellini, R. (2004). "Il turismo nei programmi di crescita e sviluppo". In: La Rosa, S., Quadrio Curzio, A., (Eds.), *Il Turismo in Sicilia* F. Angeli. In press.
- Chanel, O., Gerard-Varet, L., Ginsburgh, V. (1996). "The relevance of hedonic price indices: The case of paintings". *Journal of Cultural Economics* 20, 1–24.
- Chastel, A. (1986). "La notion de patrimoine". In: Nora, P. (Ed.), *La Nation*, vol. II, pp. 405–450.
- Costanza, R., Daly, H. (1992). "Natural capital and sustainable development". *Conservation Biology* 6 (1), 301–311.
- Creigh-Tyte, S. (2000). "The built heritage: Some British experience". *Recherches Economiques De Louvain* 66 (2), 213–230.
- Dasgupta, P.S., Heal, G.M. (1979). *Economic Theory and Exhaustible Resources*. Cambridge University Press, Cambridge.
- Ellis, A. (2003). "Valuing culture". Discussion Paper, Demos, London.
- El Sarafy, S. (1991). "The environment as capital". In: Costanza, R. (Ed.), *Ecological Economics: The Science and Management of Sustainability*. Columbia University Press, New York, pp. 168–175.
- Feigenbaum, S., Jenkinson, T. (1984). "Government incentives for historic preservation". *National Tax Journal* 37 (1), 113–119.
- Fondazione Eni Enrico Mattei (2003). "Economic growth, innovation and cultural diversità". Mimeo.
- Freeman, A.M. (1993). *The Measurement of Environmental and Resource Values: Theory and Methods*. Resources for the Future, Washington, DC.

- Frey, B. (2003). "Public support". In: *Towse (2003, pp. 232–236)*.
- Frey, B., Oberholzer-Gee, F. (1998). "Public choice, cost–benefit analysis, and the evaluation of cultural heritage". In: *Peacock (1998, pp. 27–53)*.
- Giardina, E., Rizzo, I. (1994). "Regulation in the cultural sector". In: *Peacock and Rizzo (1994, pp. 125–142)*.
- Ginsburgh, V.A. (Ed.) (2004). *Economics of Art and Culture*. Elsevier, Amsterdam.
- Ginsburgh, V.A., Menger, P.M. (Eds.) (1996). *Economics of the Arts: Selected Essays*. Elsevier, Amsterdam.
- Grafton, R.Q., et al. (2004). *The Economics of the Environment and Natural Resources*. Blackwell, Oxford.
- Grampp, W. (1996). "A colloquy about art museums: Economics engages museology". In: *Ginsburgh and Menger (1996, pp. 221–254)*.
- Guerzoni, G., (1997). "Cultural heritage and preservation policies: Notes on the history of the Italian case". In: *Hutter and Rizzo (1997, pp. 107–132)*.
- Hartwick, J.M. (1977). "Intergenerational equity and the investing of rents from exhaustible resources". *American Economic Review* 67 (5), 972–974.
- Hartwick, J.M. (1978). "Substitution among exhaustible resources and intergenerational equity". *Review of Economic Studies* 45, 347–374.
- Hartwick, J.M. (1995). "Constant consumption paths in open economies with exhaustible resources". *Review of International Economics* 3 (3), 275–283.
- Holden, J. (2004). *Capturing Cultural Value: How Culture Has Become a Tool of Government Policy*. Demos, London.
- Hutter, M., Rizzo, I. (Eds.) (1997). *Economic Perspectives of Cultural Heritage*. Macmillan, London.
- Hutter, M., Throsby, D. (Eds.) (2006). *Beyond Price: Value in Culture, Economics and the Arts*. Cambridge University Press, New York.
- Jansson, A.M., et al. (Eds.) (1994). *Investing in Natural Capital – The Ecological Economics Approach to Sustainability*. Island Press, Washington, DC.
- Kirchberg, V. (2003). "Corporate arts sponsorship". In: *Towse (2003, pp. 143–151)*.
- Klamer, A., Zuidhof, P. (1999). "The values of cultural heritage: Merging economic and cultural appraisals". In: *Economics and Heritage Conservation*. Getty Conservation Institute, Los Angeles, pp. 23–61.
- Leclair, M.S., Gordon, K. (2000). "Corporate support for artistic and cultural activities: What determines the distribution of corporate giving?" *Journal of Cultural Economics* 24 (3), 225–241.
- Levy, B., Spiller, P.T. (Eds.) (1996). *Regulations, Institutions and Commitment: Comparative Studies of Telecommunications*. Cambridge University Press, New York.
- Mazza, I. (1994). "A microeconomic analysis of patronage and sponsorship". In: *Peacock and Rizo (1994, pp. 35–53)*.
- Mazza, I. (2003). "Public choice". In: *Towse (2003, pp. 379–388)*.
- Mazza, I., Rizzo, I. (1998). "Devolution and heritage: Some political implication". Mimeo, University of Catania.
- Mazza, I., Rizzo, I. (2000). "Public decision-making in heritage conservation". Mimeo.
- McCarthy, K.F., Ondaatje, E.H., Zakaras, L., Brooks, A. (2004). *Gifts of the Muse: Reframing the Debate about the Benefits of the Arts*. The RAND Corporation, Santa Monica, CA.
- McKinley, T. (1998). "Measuring the contribution of culture to human well-being: Cultural indicators of development". In: *World Culture Report*. UNESCO, Paris, pp. 322–332.
- Moorhouse, J.C., Smith, M.S. (1994). "The market for residential architecture: 19th century row houses in Boston's South End". *Journal of Urban Economics* 35, 267–277.
- Mueller, D. (2003). *Public Choice III*. Cambridge University Press, Cambridge.
- Mussoni, M. (2003). "Il mecenatismo e la valorizzazione economica della cultura". *Rivista di Politica Economica V–VI*, 133–182.
- Navrud, S., Ready, R.C. (Eds.) (2002). *Valuing Cultural Heritage: Applying Environmental Valuation Techniques to Historic Buildings, Monuments and Artifacts*. Edward Elgar, Cheltenham.
- Netzer, D., (1998). "International aspects of heritage policies". In: *Peacock (1998, pp. 135–154)*.
- Neumayer, E. (2003). *Weak versus Strong Sustainability: Exploring the Limits of Two Opposing Paradigms*, second ed. Edward Elgar, Cheltenham.

- Nijkamp, P. (1995). "Quantity and quality: Evaluation indicators for our cultural-architectural heritage". In: Coccossis, H., Nijkamp, P. (Eds.), *Planning for Our Cultural Heritage*. Avebury, Aldershot, pp. 17–37.
- Noonan, D.S. (2003). "Contingent valuation and cultural resources: A meta-analytic review of the literature". *Journal of Cultural Economics* 27 (3–4), 159–176.
- O'Hagan, J., Harvey, D. (2000). "Why companies sponsor art events. Some evidence and a proposed classification". *Journal of Cultural Economics* 24 (3), 205–224.
- Pasquali, F. (2003). "La spesa pubblica per la cultura in Italia: Dimensioni e dinamiche". *Economia della Cultura* 4, 455–463.
- Pattanaik, P.K. (1998). "Cultural indicators of well-being: Some conceptual issues". In: *World Culture Report*. UNESCO, Paris, pp. 333–340.
- Peacock, A. (1994). "The design and operation of public funding of the arts: An economist's view". In: *Ginsburgh (1994, pp. 167–184)*.
- Peacock, A. (1995). "A future for the past: The political economy of heritage". *Proceedings of the British Academy* 87, 187–243.
- Peacock, A. (Ed.) (1998). *Does the Past Have a Future? The Political Economy of Heritage*. Institute of Economic Affairs, London.
- Peacock, A. (2004). "The credibility of cultural economist's advice to governments". In: *Ginsburgh (2004, pp. 165–178)*.
- Peacock, A., Rizzo, I. (Eds.) (1994). *Cultural Economics and Cultural Policies*. Kluwer Academic, Dordrecht.
- Pearce, D.W. (1993). *Economic Values and the Natural World*. MIT Press, Cambridge, MA.
- Pearce, D.W., Atkinson, G.D. (1993). "Capital theory and the measurement of sustainable development: An indicator of 'weak' sustainability". *Ecological Economics* 8, 103–108.
- Pearce, D.W., Turner, R.K. (1990). *Economics of Natural Resources and the Environment*. Johns Hopkins University Press, Baltimore.
- Pignataro, G., Rizzo, I. (1997). "The political economy of rehabilitation: The case of the Benedettini Monastery". In: *Hutter and Rizzo (1997, pp. 91–106)*.
- Rizzo, I. (2003). "Regulation". In: *Towse (2003, pp. 408–414)*.
- Rizzo, I. (2004). "The relationship between regional and national policies in the arts". In: *Ginsburgh (2004, pp. 203–219)*.
- Rizzo, I. (2005). "La regolazione nel settore dei beni culturali: Alcuni spunti di riflessione". In: *Trupiano, G. (Ed.), La Valorizzazione dei Beni Culturali*. Franco Angeli, Milano, pp. 17–34.
- Rizzo, I., Towse, R. (Eds.) (2002). *The Economics of Heritage: A Study in the Political Economy of Culture in Sicily*. Edward Elgar, Cheltenham, pp. 17–34.
- Schuster, J.M. (1997). "Deconstructing a tower of Babel: Privatization, decentralization and devolution as ideas in good currency in cultural policy". *Voluntas* 8 (3), 261–282.
- Schuster, J.M. (2004). "Making a list: Information as a tool of historic preservation". In: *Ginsburgh (2004, pp. 221–240)*.
- Schuster, J.M., de Monchaux, J., Riley II, C.A. (Eds.) (1997). *Preserving the Built Heritage: Tools for Implementation*. University Press of New England, Hanover.
- Shockley, G.E. (2004). "Government investment in cultural capital: A methodology for comparing direct government support for the arts in the US and the UK". *Public Finance and Management* 4 (1), 75–102.
- Solow, R. (1974a). "Intergenerational equity and exhaustible resources". *Review of Economic Studies* 41, 29–46. Symposium.
- Solow, R. (1974b). "The economics of resources or the resources of economics". *American Economic Review* 64 (2), 1–14.
- Solow, R. (1986). "On the intergenerational allocation of natural resources". *Scandinavian Journal of Economics* 88 (1), 141–149.
- Solow, R. (1994). "An almost practical step toward sustainability". In: *National Research Council, Assigning Economic Value to Natural Resources* National Academy Press, Washington, DC, pp. 19–29.
- Throsby, D. (1997). "Seven questions in the economics of cultural heritage". In: *Hutter and Rizzo (1997, pp. 12–30)*.

- Throsby, D. (1999). "Cultural capital". *Journal of Cultural Economics* 23, 3–12.
- Throsby, D. (2001). *Economics and Culture*. Cambridge University Press, Cambridge.
- Throsby, D. (2003). "Determining the value of cultural goods: How much (or how little) does contingent valuation tell us?" *Journal of Cultural Economics* 27 (34), 275–285.
- Throsby, D. (2004). "Sweetness and light: Cultural diversity in the contemporary global economy". In: Bechler, R. (Ed.), *Cultural Diversity*. British Council, London, pp. 40–53.
- Throsby, D. (2005). "On the sustainability of cultural capital". Macquarie University Department of Economics Research Paper No. 10/2005, Sydney.
- Tiebout, C. (1956). "A pure theory of local expenditures". *Journal of Political Economy* LXIV (5), 416–424.
- Towse, R. (Ed.) (2003). *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham.
- Ulibarri, C.A. (2000). "Rational philanthropy and cultural capital". *Journal of Cultural Economics* 24, 135–146.
- UNESCO (1972). *Convention for the Protection of the World Cultural and Natural Heritage*. Document 17 C/106, 15 November, UNESCO, Paris.
- UNESCO (2003). *Convention for the Safeguarding of the Intangible Cultural Heritage*. UNESCO, Paris.
- Victor, P.A. (1991). "Indicators of sustainable development: Some lessons from capital theory". *Ecological Economics* 4, 191–213.
- World Commission on Culture and Development (1995). *Our Creative Diversity*, UNESCO, Paris.
- World Commission on Environment and Development (1987). *Our Common Future*. Oxford University Press, Oxford.
- Zan, L. (2001). "Perché i musei veneziani sono in attivo". *Il Giornale dell'Arte*, No. 201, July.
- Zolberg, V. (2000). "Privatization: Threat or promise for the arts and humanities?" *International Journal of Cultural Policy* 7 (1), 9–28.

## THE ECONOMICS OF MUSEUMS

BRUNO S. FREY and STEPHAN MEIER

*University of Zurich, Switzerland*

### Contents

Abstract	1018
Keywords	1018
1. Introduction	1019
2. Demand for museum services	1020
2.1. Private demand	1020
2.2. Social demand	1022
2.2.1. External effects	1022
2.2.2. Effects on markets	1023
3. Supply of museum services	1024
3.1. Cost structure	1024
3.2. Cost functions	1025
3.3. Firm structure	1026
4. Museum behaviour	1027
4.1. Neoclassical approach	1027
4.1.1. A representative model	1027
4.1.2. Critique	1028
4.2. Institutional approach	1029
4.2.1. Public museums	1030
4.2.2. Private museums	1030
4.2.3. Museums dependent on donations	1031
4.3. Museum behaviour in three important areas	1032
4.3.1. Collection management	1032
4.3.2. Pricing	1033
4.3.3. Commercial activities	1034
5. Current trends in the museum world	1035
5.1. Superstar museums	1036
5.2. Special exhibitions	1038
5.2.1. Demand aspects	1038
5.2.2. Supply aspects	1040

5.2.3. The future of special exhibitions	1042
6. Conclusion	1042
References	1042

## **Abstract**

Museums fulfill many important functions in the art world and visits to museums are becoming an important leisure and holiday activity. This chapter surveys research about the functioning of museums from an economic point of view. Museum services are shaped by demand and supply factors and by the institutional setting constraining the decision makers in a museum. This chapter argues that the institutional factors, e.g., whether a museum is private or public, influence greatly how the museum is run with respect to the management of the collection, price setting, or the focus on commercial activities. Two current trends, the evolution of superstar museums and the growing importance of special exhibitions, are analysed from an economic point of view.

## **Keywords**

museums, collection management, pricing, commercial activities, superstar museums, special exhibitions

*JEL classification:* Z1

## 1. Introduction

Museums are becoming more important nowadays than ever before. They play a significant role in people's leisure activities and constitute important tourist attractions. Consumers spend substantial amounts of money visiting museums, both in terms of admission fees and expenditures in museum restaurants and shops. These expenditures have a strong effect on local economies, especially in popular tourist areas. There are many types of museums, classifiable according to four different criteria:

- *Content*: Museums may contain art, historical artifacts, scientific objects, and/or many other exhibits of general and sometimes very specific interest.
- *Size*: Some museums are huge and draw thousands of visitors a day from far and near; others are small, have few visitors, are run by amateur staff, have very restricted opening hours and are only of local interest.
- *Age*: Some museums lay claim to a long and distinguished history and are often situated in very old buildings, while others are newly founded and may impress visitors with their spectacular architecture.
- *Institutional form*: Traditionally European museums have been public, even forming part of the normal government administration. But there have always been private museums. Most museums are neither completely public nor private but lie somewhere in between. Almost all private museums receive some form of government subsidy, often in the form of contributions made by donors, who can claim tax exemption.

Despite these various differences, however, all museums share some particularities and similar functions.<sup>1</sup> This chapter analyses the various types of museums and points out where different aspects are crucial for understanding the behaviour of museums.

The term "Economics of Museums" may be understood in two different ways:<sup>2</sup> First, a museum may be looked at as an economic unit, or as a firm providing certain services. The analysis then focuses on the relationship between the input (exhibits, manpower, etc.) and output measured, for example, in terms of revenue. Moreover, the effect of museums on the economy may be analysed, e.g., how much employment is generated and

<sup>1</sup> A museum might have five different functions: to collect, conserve, study, interpret and exhibit [Noble (1970)]. These five functions could be condensed into three: preservation, research and communication [Weil (2002)]. Ginsburgh and Mairesse (1997) take an empirical look at the mission statement of Belgian museums and propose an alternative definition of a museum.

<sup>2</sup> The Economics of Museums has been the topic of a number of publications, including Montias (1973), Peacock and Godfrey (1974), Feldstein (1991), Bayart and Benghozi (1993), Frey (1994), Martin (1994), Robbins (1963), O'Hagan (1995, 1998b), Johnson and Thomas (1998), Schuster (1998a, 1998b), Benhamou (1998), Meier and Frey (2003), Maddison and Foster (2003) and Weil (1987). It has also been covered in more general surveys [Throsby (1994); Blaug (2001)], in monographs and textbooks [Frey and Pommerehne (1989); Frey (2000); Heilbrun and Gray (2001); Benhamou (2000)] and in readers in the field of cultural economics [Blaug (1976); Peacock and Rizzo (1994); Ginsburgh and Menger (1996); Towse (1997a)]. Early contributions in German are by Kindermann (1903), and in English by Robbins (1963, 1971), Baumol and Bowen (1966) and Peacock (1969).

how much added value is created in other sectors. Second, applying an economic way of thinking to museums involves assuming that individuals pursue their utility within the constraints imposed by institutions and the environment, especially where resources are scarce. This methodology has been applied to many different areas, such as politics, law, history, sports, or religion.<sup>3</sup> The economics of museums thus clearly distinguishes itself from other ways of studying museums, in particular the sociology of museums or the art historic points of view.<sup>4</sup>

The economic approach to museums may rely on standard or rational choice theory derived from neo-classical economics; individuals are then taken to be completely rational and selfish, and the analysis focuses on market relationships, which are assumed to function well. Political economy or public choice economics studies the behaviour of governments and public administrations as they affect museums, not only through subsidies and taxes but also through the web of regulations. Other types of economic theory may also be applied. In the context of museums, a particularly important variant is to take psychological aspects into account: individuals are not totally rational and are sometimes subject to anomalies, and they may to some extent be other-regarding and act in a pro-social way.

This chapter proceeds by looking first at the demand for museum services and then at the supply. Museum behaviour is analysed from a neoclassical and then from a more institutional perspective. The next section is devoted to public policy issues connected with museums, and the last section discusses current trends in the museum world from an economic point of view.

## 2. Demand for museum services

There are two types of demand for museums. The first is the private demand exerted by the visitors. These may be persons interested in the exhibits as a leisure activity or as part of their profession as art dealers or art historians. The visit may be undertaken by an individual or family, or may be part of an organised outing, e.g., by schools or firms. The second type of demand comes from persons and organisations benefiting from a museum. This social demand is based on external effects and/or effects on economic activity.

### 2.1. *Private demand*

By far the largest number of museum visits can be attributed to leisure time activity; since visits by the specialists mentioned above play a relatively minor role, they can be ignored here. The number of visits can be analysed by a traditional demand function,

<sup>3</sup> See, for example, Becker (1976), Hirshleifer (1985), Kirchgässner (1991) and Frey (1999).

<sup>4</sup> For example, Bourdieu (1979), Moulin (1986), Di Maggio (1986), Foster and Blau (1989) and Blau (1995).



capturing the major factors determining the number of visits in any given period of time. There are three major determinants relating to prices or costs:

(i) *Admission fee*. Econometric estimates for a large number of different museums in various countries suggest that the demand for museum services is price inelastic.<sup>5</sup> However, most studies are limited to case studies of one or two museums. For example, Goudriaan and Van't Eind (1985) found an average price elasticity of  $-0.1$  to  $-0.2$  for four Dutch museums. Darnell, Johnson and Thomas (1992) found a higher elasticity of  $-0.55$  for one particular museum in Great Britain. Luksetich and Partridge (1997), using US data from the 1989 *Museum Survey*, estimated demand functions for different types of museums. Their estimated price elasticity ranged from  $-0.12$  to  $-0.26$ , depending on the type of museum. The elasticity for art museums was found to be  $-0.17$ . Zoos, science museums and natural history museums showed the largest price sensitivity, probably due to greater competition from other leisure activities. Overall, the low price elasticities suggest that museums could generate significant increases in revenues through increasing admission fees.

(ii) *Opportunity cost of time*. For persons with high income or who are self-employed, the opportunity cost of time is higher than for people of low income or on fixed working hours who are therefore expected to visit museums more often, all other things being equal. The opportunity cost of a museum visit not only depends on the time actually spent in the museum, but also on how much time is required to get there. For tourists, the opportunity costs of time tend to be lower than for local inhabitants, because they often visit a city with the express purpose of visiting the respective museums. Econometric estimates have found no clear link between income and attendance [Luksetich and Partridge (1997)]; this is in line with Gapinski's (1986) findings for the lively arts. The increased opportunity costs of time for wealthy people attending art performances offset the positive income effect; the two effects have to be separated in order to find a positive income effect and a negative opportunity cost effect on demand.<sup>6</sup>

(iii) *Price of alternative activities*. These are, most importantly, alternative leisure activities such as going to other cultural events (theatre, cinema), taking part in sports, dining out in a restaurant, spending time with friends at home, etc. Even within the industry, one museum may be an alternative to another museum. But complementary costs also systematically influence the number of museum visits; for example, the costs incurred through travel, accommodation and meals may be important. These complementary costs constitute a high percentage of the total costs of a visit; Bailey et al. (1998) estimated them at more than 80 percent. Cross-elasticities have been found to be empirically significant for the arts,<sup>7</sup> but there have been no estimations of demand

<sup>5</sup> Apart from the question of how sensitive the demand is to price increases, there has been extensive discussion about the effect of charging an admission fee at all. For a general discussion about the question of what to charge, see O'Hagan (1995).

<sup>6</sup> See Withers (1980) for estimations for the performing arts.

<sup>7</sup> See Gapinski (1984, 1986) for estimations for the performing arts.

functions for museum services incorporating such variables. Income is another “classical” determinant affecting the demand for museum visits. Econometric estimates reveal an income elastic demand.<sup>8</sup> Estimates of income effects are often ambiguous, because the rising opportunity cost of time goes hand in hand with higher income. There is also a high correlation between income and education; better-educated people have the human capital necessary to benefit more fully from museums than people with lower education.<sup>9</sup> It can be suggested that this factor plays a larger role for museums of modern art and history than for museums of science and technology, especially transport museums (railways, cars, or space travel).

There are many other determinants to be included in a well-specified museum demand function.<sup>10</sup> One is the quality of the collection or the special exhibition mounted. Luksetich and Partridge (1997) estimated that the value of the collection increases attendance figures, especially for art museums. Or, as Oster and Goetzmann (2001, p. 9) state: “In fundamental terms, these results suggest that art matters.” Other determinants are how attractive the building is and the level of amenities provided by the museum, i.e. the general atmosphere, the extent of congestion in front of the exhibits, the cafés and restaurants, and the museum shop. The marketing efforts made by a museum also matter. A final determinant of the rate of museum visits is that of individual preferences, which are difficult to measure independently. Econometric studies of museum demand functions often indirectly capture individual preferences by introducing past visits as a determinant. In all empirical estimates, this factor has proved to be highly significant and important: persons who visited a museum in the past are more likely to do so in the present and future.

## 2.2. Social demand

Museums have effects on society which go beyond the experiences of the actual museum visitors themselves. These social effects include externalities and the effects on markets.

### 2.2.1. External effects

Museums create social value for which they are not compensated in monetary terms. As a consequence, museums tend not to produce these values, or in too low an amount. Five types of non-user benefits can be distinguished in the literature:

- *Option value*: People value the possibility of enjoying the objects exhibited in a museum sometime in the future.

<sup>8</sup> See, e.g., Withers (1980).

<sup>9</sup> For the influence of art lessons on museum visits, see Gray (1998).

<sup>10</sup> Many studies analyse surveys about museum visitors in order to see who actually visits the museums [see, e.g., Dickenson (1997)] and how much visitors benefit from their visits [Ashworth and Johnson (1996)].

- *Existence value*: People benefit from knowing that a museum exists, but do not themselves plan on visiting it now or in the future.
- *Bequest value*: People derive satisfaction from the knowledge that their descendants and other members of the community will be able to enjoy a museum in the future if they choose to do so.
- *Prestige value*: People derive utility from knowing that a museum is highly valued by persons living outside their community – they themselves need not actually like the museum, nor even visit it.
- *Education value*: People are aware that a museum contributes to their own or to other people's sense of culture and value it because of that.

Museums may also produce negative external effects, the costs of which are borne by other people. An example is the congestion and amount of noise museum visitors inflict on the local community.

The above non-user benefits and costs have been empirically measured by using three different techniques:

- An obvious possibility is to conduct representative surveys of both visitors and non-visitors to a museum. The questionnaires have to be carefully designed in order to elicit the true willingness to pay for the various social values produced by a museum. Contingent valuation studies, first developed to capture environmental values, have done a good job of capturing cultural values.<sup>11</sup>
- Other techniques rely on the revealed behaviour of individuals, for example, by estimating the extent to which property values increase in a city containing a museum, assuming that people are willing to pay more for a house or apartment situated in a location with a museum compared to an equivalent house or apartment in a location without such a museum. The same “compensating variation” can be computed by analysing wages, if persons are willing to work for lower compensation in a location housing a museum. The compensating variation method has been used, for example, by [Clark and Kahn \(1988\)](#).
- A third approach for capturing social values is to analyse the outcome of popular referenda on expenditures for museums. In Switzerland this approach has been successfully implemented in identifying options, such as the existence and bequest values of buying two Picasso paintings for a museum [[Frey and Pommerehne \(1989, Chapter 10\)](#)]. In the case of the performing arts, [Schulze and Ursprung \(2000\)](#) have analysed a referendum in Switzerland to gauge the amount of support for the opera house in Zurich. They were also able to identify external effects.

### 2.2.2. *Effects on markets*

Museums produce monetary values for other economic actors by creating additional jobs and commercial revenue, particularly in the tourist and restaurant industries. These

<sup>11</sup> See, for example, [Martin \(1994\)](#). The extensive empirical literature is surveyed in [Noonan \(2002\)](#); for a critical discussion from a behavioural point of view, see [Sunstein \(2002\)](#).

expenditures induce further expenditures and a multiplier effect results. Impact studies<sup>12</sup> measuring the additional market effects created are popular with politicians and administrators because they provide them with reasons for spending money on museums. However, these studies have to be interpreted very carefully:

- Impact studies tend to focus on the wrong issue. The *raison d'être* of museums is to produce the unique service of providing a certain type of cultural experience to its visitors, as well as providing non-user benefits.
- It is *not* the task of a museum to stimulate the economy; there are generally much better means of achieving that goal. A theme park, for example, may be better able to stimulate the economy.

### 3. Supply of museum services

#### 3.1. Cost structure

Museums have cost structures that differ from those of other firms in the service industry in at least four respects:

- *High fixed costs.* Museums in general operate with high fixed costs: buildings, exhibits, staff, insurance, technical outfits, etc. cannot be changed in the short run. Thus in the short run the operating costs of museums tend to be independent of output. In other words, variable costs constitute a relatively low percentage of total costs, and hence museums face rapidly decreasing unit costs when the number of visitors increases.
- *Marginal costs are close to zero.*<sup>13</sup> If a museum sets up an exhibition, the basic operating costs consist in opening the museum on a particular day, which to a large extent is independent of the number of visitors. In these circumstances the marginal costs of an extra visitor are close to zero. However, in some circumstances (for example, so-called “blockbuster” exhibitions), an additional visitor may impose congestion costs. Maddison and Foster (2003) analysed the congestion costs at the British Museum using contingent valuation techniques; they estimated that the cost imposed by the marginal visitor was £8.05, an estimate that seems to be exceptionally high.
- *Dynamic cost.* It might be claimed that museums face the same economic dilemma as most cultural organisations, namely the cost disease<sup>14</sup> whereby they are subject

<sup>12</sup> See, e.g., Seaman (1997, 2002) and, for two special exhibitions, Wall and Roberts (1984).

<sup>13</sup> Even if the output unit is the number of hours or days per year for which the museum is open, this statement probably holds. The British Museum tried to cut costs by closing some sections, but they did not achieve a big cut in costs – at least not in the short run [see Anon. (2002); Art Newspaper (2002)].

<sup>14</sup> For a detailed survey of Baumol's Cost Disease, see Towse (1997b). For a critique, see Cowen (1996) and Peacock (1993, pp. 66–70). See also Chapter 11 by Baumol and Chapter 15 by Brooks in this volume.

to a productivity lag that produces constant financial problems. However no empirical study exists analysing this claim for museums. Indeed, productivity advances in the museum industry seem to be possible: items can be shown on the Internet; surveillance can be undertaken by cameras; increased reliance may be placed on volunteers; activities may be outsourced; or institutional settings may be changed, like introducing New Public Management for public museums, or privatising them entirely. All these changes work against the potential effects of the cost disease.

- *High opportunity costs.* Museums own, through their collections, endowments which are often highly valuable. For most museums the value of their holdings is by far their greatest asset. The objects in a museum's collection generate not only storage and conservation costs but also opportunity costs. The opportunity cost of keeping a work of art is the return generated by its alternative use, i.e. by selling the work of art and investing the money in options. Other opportunity costs refer to the building and its alternative uses.<sup>15</sup> Most museums do not place a value on their collection in their accounts. In Great Britain, not placing a value is even a condition of registration with the Museums and Galleries Commission [Bailey and Falconer (1998, p. 173)]. Museums understate their true capital costs by not taking opportunity costs into account [Grampp (1989, p. 171)]. Neglecting opportunity costs can partly be explained by a rational reaction of the museum directorate to action undertaken by the political sector.<sup>16</sup>

### 3.2. *Cost functions*

It is important to know how museum costs vary with regard to output and input. One of the few museum cost functions to be estimated is that of Jackson (1988).<sup>17</sup> This study took various activities of the museum into account and analysed their influence on costs. The log-linear model is:

$$\ln TC = \ln a + b \ln Q + y \ln W + s \ln K + r_1 EX + r_2 ED + r_3 CN + r_4 MB + r_5 AC, \quad (1)$$

where  $TC$  is total operating cost,  $Q$  is the total attendance figure,  $W$  is the wage rate paid per worker, and  $K$  is the cost of capital measured as the ratio of promotional expenditures such as development, membership and advertising, to contributions from all public and private sectors. Because a museum can engage in various activities, the study looked at how priorities set by the museum influence costs. Therefore,  $EX$  are exhibition expenses as a fraction of total operating costs,  $ED$  are educational expenses,  $CN$  are

<sup>15</sup> In some cases, the opportunity cost of the land may be quite high, as museums are often situated in commercially attractive locations; see Rosett (1991) for more details.

<sup>16</sup> This question will be discussed in more depth in Section 4 below.

<sup>17</sup> For a cost function for performing arts, see, for example, Lange et al. (1985).

conservation and preservation expenses, and *MB* are expenses in connection with membership activities. Because quality strongly influences costs in the performing arts,<sup>18</sup> the study tried to capture quality by looking at which museum had been accredited with the American Association of Museums; *AC* is a dummy variable equal to 1 if accredited and 0 otherwise. This is only a rough proxy for quality.

The results, based on data from the Museum Program Survey 1979 for 326 US museums, yielded two interesting results: First, museum operation appeared to be characterised by economies of scale. Operational costs changed more slowly than attendance figures in small museums with up to 99,000 visitors a year. However, for bigger museums, diseconomies of scale were at work. Average cost curves for (art) museums were downward sloping with low attendance levels and rose only after the annual number of visitors exceeded 100,000. This result backs up the statement about museums being a decreasing cost industry. Second, an increase in expenses for membership activities as a fraction of total operating expenses decreased total costs. This may be due to the fact that a more active group of members stimulates voluntary work and that the cost of capital can be decreased as fundraising becomes easier.

Overall it can be said that more research is needed for us to fully understand the cost functions of museums.<sup>19</sup>

### 3.3. Firm structure

Museums can take different organisational forms. They are mostly either private for-profit organisations, private non-profit organisations, or public organisations run in a non-profit-making way. In Europe and the United States, the non-profit organisational form predominates for museums. Various hypotheses have been put forward to explain the dominance of non-profit firms in the museum world and the arts in general.<sup>20</sup> According to Weisbrod (1977), non-profit organisations were established due to an unsatisfied demand for public goods. Alternatively, the cost structure of museums may be responsible for establishing non-profit organisations.

Most museums face a demand curve lying below the average cost curve. This makes it impossible to set a price at which total admission receipts cover the total costs of the museum. If price discrimination is not applicable or only of limited use, Hansmann (1981) argued that arts organisations can still ask individuals for voluntary price discrimination. Visitors volunteer to pay more than the official admission price and thus become donors. The non-profit form is superior to the for-profit enterprise when it comes to getting donations, because consumers lack exact information about the quality of the

<sup>18</sup> See Throsby (1977) and Globerman and Book (1974).

<sup>19</sup> For a related branch of research on efficiency measurements in museums, see Mairesse and Vanden Eeckhout (2002) and the literature cited there.

<sup>20</sup> For a selection of articles dealing with non-profit firms in the arts, see Di Maggio (1986). For a general survey about non-profit firms and altruistic behaviour, see Rose-Ackerman (1996). See also Chapter 15 by Brooks in this volume.

good and service provided. It is impossible to draw up a contract which gives donors complete protection from being exploited. As a consequence, donors prefer non-profit firms, where the possibility of managers exploiting donors and consumers is limited.<sup>21</sup>

Smolensky (1986) argued that educational externalities, and not decreasing costs, led to the non-profit form of museums. In Europe, governments started to support museums because of these educational externalities while in the United States “public provision was rejected as a socialist solution” [Smolensky (1986, p. 768)]. The non-profit form that was subsequently established was a hybrid, applicable not only to museums but to performing arts organisations, universities, libraries and hospitals.

#### 4. Museum behaviour

Two theoretical approaches for analysing the behaviour of museums are presented here: the neoclassical approach, which assumes rational actors maximising the utility of a museum in a benevolent way, and the institutional approach, emphasising how various institutional settings (e.g., the dependence on public support) influence the behaviour of the museum management. These approaches are then used to analyse three major activities of museums: collection management, pricing policy and commercial activities.

##### 4.1. Neoclassical approach

###### 4.1.1. A representative model

Throsby (1994) presented a model of the behaviour of performing arts firms which can be applied to museums. The model assumes that there is no distinction between ownership and control of the firm. The directorate of the museum maximises the firm’s utility function. Assuming that a museum’s objective is non-profit making, the budget constraint requires zero net revenue. It has been proposed that the museum’s utility is related to the number of visitors to the museum ( $y$ ) and the quality of the exhibitions ( $q$ ). This assumes that the quality of the museum service can be measured. The museum management thus decides to maximise

$$U = U(y, q) \tag{2}$$

subject to

$$p(y)y + g(q) + h(y) - c(y, q) = 0. \tag{3}$$

The museum gains revenues from entrance fees ( $p$ ), which is a function of the number of visitors ( $y$ ); the level of donations and government grants ( $g$ ), which depend exclusively on the quality of the museum; and the revenue from ancillary goods from the shop

<sup>21</sup> For a similar argument, see Glaeser and Shleifer (2001).

and the restaurant or café ( $h$ ), which depends on the number of visitors. Costs depend on both output and quality.

The first-order conditions can be written as:

$$U_y/\lambda + p_y y + p(y) + h_y = c_y, \quad (4a)$$

$$U_q/\lambda + g_q = c_q, \quad (4b)$$

$$p(y)y + g(q) + h(y) = c(y, q). \quad (4c)$$

The subscripts indicate partial derivatives and  $\lambda$  is the multiplier on the constraint.

This formulation produces two insights: First, directors of a non-profit museum receive extra utility from a higher number of visitors. They therefore set the entrance fee in such a way that marginal revenue from entrance fees and ancillary goods is less than marginal costs. This result may explain why museums set too low a price compared to the revenue-maximising condition [e.g., Luksetich and Partridge (1997)]. Second, museums engage in quality beyond the point where marginal grant income is equal to the marginal cost of increasing the quality by one unit. This behaviour is due to the extra utility the museum gets from an increase in quality. According to the model, museums tend to provide too high quality at too low a price, compared to a revenue-maximising firm.

The objectives of the museum, the quality of the exhibition and the number of visitors are crucial elements in the above model. Hansmann (1981) analysed the extreme cases of a museum interested only in quality, a steady flow of visitors or budget. For example, the quality maximising firm sacrifices the number of visitors for the sake of quality. But different forms of public grants also affect the result. While lump-sum subsidies would lead to an increase in quality for the quality maximiser, the increase in the number of visitors has a less certain effect. It only happens if the increase in the flow of visitors does not increase the cost of quality, and if the marginal visitor attracted has an unusually marked taste for quality. Different behaviour results if the museum is supported in matching grants with the donations it receives. In this model, a subsidy does not only increase donations, but will induce the museum to adjust quality and price (and therefore visitor flows) to a level closer to maximising consumer welfare.<sup>22</sup>

#### 4.1.2. Critique

The model presented above assumes that managers of museums behave in a benevolent way and are driven only by a cultural aspiration favourable to the owner of the museum (e.g., the public, private donors and/or a foundation). But this assumption may be criticised in two respects: First, managers of museums and chief curators are likely to behave in a more selfish way than assumed by the model. Second, it may be that

<sup>22</sup> Many studies show that grants which match donations increase willingness to donate (see further Chapter 35 by Netzer and Chapter 36 by Schuster in this volume); see also List and Lucking-Reiley (2002) and Ribar and Wilhelm (2002).



museum managers are primarily interested in their reference group and try to maximise their respective reputations; in the absence of the right incentives, they will not produce the quality and quantity maximising the firm's or consumers' utility.

#### 4.2. *Institutional approach*

Museum managers (as well as all other actors in society) may be assumed to be primarily, or even exclusively, concerned with their own utility. They may therefore be taken to be interested in their own income and the prestige they get within their reference group, consisting mainly of art lovers and the international museum community. A second source of utility has to do with agreeable working conditions and job security. But the museum management is not free simply to pursue its own goals, because it faces certain constraints on its actions. Differences in these institutionally determined restrictions may explain the museum management's behaviour.

The financial resources available are the most important constraint on museum management. Other constraints such as limited space or legal and administrative burdens imposed by the public bureaucracy or labour unions can also weigh heavily. Income sources differ considerably between museums. Some depend mostly on public grants, while others rely more on private money (donations and sponsorship, or income generated from entrance fees, shops and restaurants).<sup>23</sup> From a politico-economic point of view, the institutional set up and the nature of funding museums has a dramatic influence on the behaviour of its management.<sup>24</sup>

Most museums lie somewhere between the extremes of purely public and purely private museums.<sup>25</sup> In the last couple of years, more public museums have moved in the direction of private museums because state support has decreased, especially in Europe [NEA (2000)]. Governments, as a consequence, have given the museum managers more independence. Both the discretionary room and the pressure to generate more income of their own has increased. Nevertheless, the institutional setting remains crucial for the behaviour of the museum management. The fact that public museums change their behaviour notably once they have more independence underlines the power of institutional factors.

In the following sections we distinguish between three types of museums: public museums, private museums, and museums dependent on donations. The three models of museums generate predictions for the behaviour of agents, e.g., managers, donors, and visitors.

<sup>23</sup> Rosett (1991) presents evidence on the financing of US museums, which support the notion of heterogeneous funding of museums.

<sup>24</sup> See Frey and Pommerehne (1989), Rosett (1991) and Meier and Frey (2003); see also the theory of non-profit organisation in Weisbrod (1998), James (1983), Schiff and Weisbrod (1991), and for a principal-agent model, Prieto Rodriguez and Fernandez Blanco (2002).

<sup>25</sup> See Schuster (1998a), van Hemel and van der Wielen (1997) and Meier and Frey (2003).

#### 4.2.1. *Public museums*

Directors of purely public museums rely exclusively on public grants. The government allocates them sufficient funds to cover the expenses considered necessary for fulfilling their tasks. While they are expected to keep within their budget, if a deficit does occur, it will be covered by the public purse. This institutional setting provides little incentive to generate additional income and to keep costs at a minimum. The management does not allocate energy and resources to generate additional income, because any additional revenue produced goes back into the national treasury. Moreover, with a surplus, the public grants would correspondingly decrease, which acts like an implicit tax of 100 percent on profits. Instead, the museum management emphasises non-commercial aspects, such as referring to intrinsic “artistic”, “scientific”, or “historical” values. This application of non-commercial standards helps the museum management to achieve their goal of prestige, top performance and pleasant working conditions. Even if museum revenue was not automatically to go back to the public purse, Maddison (2004, p. 89) shows that “(s)tatistically analyzing data drawn from a panel of UK museums, evidence is found that increases in non-grant incomes do indeed result in a statistically significant reduction in future government subsidies”. Based on this institutional point of view, one would therefore expect that:

- Public museums will not sell any paintings from their art collection because the management is prevented from using the income generated as it would wish, and selling part of the collection would leave management vulnerable to criticism from outside (be it by politicians or by public administrators) because “artistic” activities have now become monetised [Frey (1994); Montias (1973)].
- Directors of public museums will have little interest in the number of visitors, because they are not dependent on income from entrance fees or shops. Therefore, exhibitions will be designed to please an insider group of art “freaks”.
- Visitors’ amenities in public museums will be poorly developed, and little attention will be paid to the profitability of museum shops, restaurants and cafeterias.

#### 4.2.2. *Private museums*

Managers of purely private museums have a strong incentive to increase their income, because their survival depends on revenues produced by entrance fees, the restaurant, shop surpluses and additional money from sponsors and donors. If private museums manage to generate a surplus, the management can use it for future undertakings. Consequently, it is to be expected that private museums will behave in the following ways:

- They will rely on the market when managing their collection. Museums will actively sell paintings that no longer fit into the collection and use the money to buy new works of art.
- They will actively seek to gain additional revenue from museum shops, restaurants and cafeterias, and will be prepared to host non-artistic events such as corporate meetings in their facilities.

- Private museums will be concerned with attracting visitors. “Blockbuster” exhibitions will help the museum to earn revenue, because the preferences of a larger group of people will be catered to. The management will make a huge effort to have the exhibitions well arranged from a didactic point of view and appeal to a large crowd.
- Private museums will emphasise amenities to visitors, such as cafeterias and clean restrooms.

#### 4.2.3. *Museums dependent on donations*

In some countries, contributions to non-profit museums are deductible under the income tax rules for individuals and corporations.<sup>26</sup> A reduced marginal tax reduces the willingness to donate because the implicit price of doing so rises. The tax-deductible status, if chosen by the museum, affects behaviour fundamentally. There is an incentive to avoid profits by charging low or “social” prices (which strengthens the legitimacy of tax-deductible status). There is also an incentive to take out profits in the form of various kinds of payments, showing up as costs.

Museums depending on donations have an incentive to attract donors, and they devote a great deal of effort and resources to this end. Donors can exercise some measure of control over the activities of museums.<sup>27</sup> Museums dependent on financial and/or in-kind donations are expected to behave in the following way:

- Donors will directly influence museum policy by interfering in the programming, or they can set strictly binding constraints on the ways in which works they donate can be used. Most donors have clear ideas on how the works of art donated should be exhibited. Donors in general will also want to prevent paintings donated from ever being sold, which imposes considerable opportunity costs on museums. Donors can benefit from museums publicising their contribution, thus enhancing their prestige [Glazer and Konrad (1996); Harbaugh (1998)]. Museums have developed an elaborate system of honours, ranging from appropriate attributes (“benefactor”, “patron”, “contributor”, etc.), to naming rooms, wings and even whole buildings after the donor.
- Museums must make the impression that donations are well used, so that donors will have the feeling that they are contributing to a worthwhile cause. It is crucial that the art institution has a good reputation with the public and the media to encourage a regular flow of donations. This compels the museum management to act efficiently. Contracts cannot completely control the museum managers. Donors will therefore prefer to deal with non-profit firms acting under a “non-redistribution constraint” (i.e. prohibiting the personal appropriation of profits). Removing the profit goal avoids the problem that managers cheat on the donors to a certain extent [Hansmann (1981)].

<sup>26</sup> For an overview of the legal possibilities of deducting donations to the arts from taxes, see Schuster (1985, 1986), and Chapter 36 in this volume.

<sup>27</sup> As discussed in Glaeser (2001, p. 39) and Oster and Goetzmann (2001).

### 4.3. *Museum behaviour in three important areas*

#### 4.3.1. *Collection management*

In most art museums throughout the world, a considerable portion of the holdings of paintings is not exhibited and not accessible except perhaps to specialists. What constitutes a major part of a museum's wealth does not appear on any balance sheet; the bookkeeping procedures of art museums often do not even mention that the paintings collected are of any value, although at today's art market prices collections of even minor museums are likely to be worth millions of dollars and in the case of major museums many hundreds of millions of dollars.<sup>28</sup> The museum managers know, of course, that their holdings are very valuable. Three reasons can be put forward which may explain this behaviour:

(i) The government may impose a legal constraint on selling. Many, or even most, public museums in continental Europe are prohibited from selling works from their collection, a practice known as de-accessioning.<sup>29</sup> This practice is often allowed in the United States and to a lesser extent also in Britain [Grampp (1996)], but as O'Hagan (1998a, p. 171) notes: "The real opposition arises from the museum personnel and not from the law." Even in the United States where it is legal to sell paintings the curators argue that it is not ethically right to do so unless one improves the collections in the process. However, Temin (1991) for instance argues that museum directors should be allowed to be more flexible in using the money of de-accessioned paintings, e.g., for operation expenses.

(ii) Voluntary contracts between the museum management and the donors of paintings may act as a constraint. In these circumstances the management is faced with a trade-off between receiving additional paintings and having to accept certain restrictions [Thompson (1986); Weil (1987)]. Today, few museums are prepared to accept such restrictions being attached to a donation [Weil (1987)].

(iii) The most convincing explanation for the behaviour observed refers to institutional differences. For public museums, the museum directorate has no incentive to sell the holdings it has in storage<sup>30</sup> for two major reasons [Frey (1994)]. First, when a painting is sold, the revenue gained is not added to the museum's disposable income but, according to the rules of the public administration in most countries, goes into the general public treasury. Even if this were not the case, it is likely that the budget allocated to the museum would be correspondingly reduced. This institutional setting undermines the incentives to manage the collection on the market. Second, selling paintings means that the existing stock of art is at least partially monetised, which may encourage outside

<sup>28</sup> Most art museums hold a large part of their paintings in storage rooms – up to 80 percent of the collection; see, e.g., Lord, Dexter and Nicks (1989).

<sup>29</sup> For a discussion on the legal aspects of de-accessioning art, see White (1996).

<sup>30</sup> Pommerehne and Feld (1997) also found differences in buying paintings by public and private institutions; for example, public museums paid more in art auctions than private investors did, *ceteris paribus*.

interference by politicians and parliamentarians into the museum's business [O'Hare and Feld (1975)] since the museum directorate's "performance" becomes easier to evaluate. As long as the criteria for evaluation are exclusively of an art historic nature, the museum community is to a substantial extent able to define its performance itself. This is a useful and successful survival strategy that museum administrations do not voluntarily give up.

By way of contrast, private American art museums actively sell and buy art as they see fit; for example, in the period 1988–1989, 88 museums sold 1284 lots worth \$29.6 million, and 93 museums bought 142 lots worth \$37.5 million [Cantor (1991, p. 21)]. The director of the Getty Museum stated that "this practice . . . (is) the key to shaping the collections by the staffs of many major big city museums with large collections, and others too" [Feldstein (1991, p. 26)]. Note that if the name of the donor is attached to the painting, the donor's resistance to de-accessioning may be reduced.

Lending policy is a different but related phenomenon. There is a norm not to exchange works of art using the price mechanism [Caves (2000, pp. 345–347)].<sup>31</sup> Even private museums follow this rule.<sup>32</sup>

#### 4.3.2. Pricing

Museums differ in the way they set entrance fees. There is an extensive discussion on whether to charge or not to charge,<sup>33</sup> a discussion that probably goes back to Sir Hans Sloane, whose donation at his death in 1753 led to the founding of the British Museum, but with the explicit restriction not to charge an entrance fee. Even today, most British museums do not charge their visitors for admission. In the United States too, there are some museums, at least the national ones, which do not levy an explicit entrance fee. Two main arguments are put forward in favour of free admission. First, there are some positive externalities connected with a museum, as discussed above, and therefore the museum should be compensated for this service by tax money. Those who actually visit a museum probably benefit most from the museum; hence an entrance fee should be levied over and above the contribution from general taxation. There does not seem to be any evidence that this measure hits low-income groups disproportionately [O'Hagan (1998a, p. 178)]. Second, the low or zero marginal costs of a visitor suggest that a zero price is efficient. However, as mentioned above, the assumption of zero marginal costs can be criticised for various reasons.

There are a variety of pricing options besides free admission: donation boxes with and without price suggestions, seasonal tickets with zero marginal pricing or a free-day

<sup>31</sup> Russian museums are an exception. Western museums are prepared to waive their rule of not paying money for lending works of art, because they acknowledge that Russian museums are extremely short of cash.

<sup>32</sup> Note that there are many advantages to having instead a market relying on barter; for an overview, see Heilbrun and Gray (2001, pp. 202–209).

<sup>33</sup> For an overview, see, e.g., O'Hagan (1995), Heilbrun and Gray (2001) and Bailey and Falconer (1998).

policy.<sup>34</sup> In addition, price discrimination, often supported by economists [e.g., Frey (1994)], can be undertaken in times of high demand and/or with respect to the type of visitor. For example, many museums, even those who do not charge for their permanent collection, have higher entrance fees for special exhibitions. Additionally, the museum could charge more at weekends and less during summer holidays. Tourists could be charged more than local residents. Visitors who want to spend little time in a particular museum could be charged less. In periods of high demand when the art museum is stretched to capacity, a high and a low entrance fee could be set; the higher-priced entrance will have a shorter queue and will be used by visitors with high opportunity costs, often persons with above-average income, whereas the lower-priced entrance will be used by visitors not wanting to spend too much money, and having plenty of time at their disposal.<sup>35</sup> Such price differentiation is advantageous for both categories of visitors (the one gets in more quickly, while the other pays less) as well as for the museum, which can raise its revenue.

The question of how pricing influences the finances of museums depends not only on the price elasticity of demand. Charging can also influence the flow of public subsidies and donations. Moreover, pricing decisions also impact on the amount of income generated with ancillary goods, like revenue from the shop and restaurant. In some cases, the government enforces targets which the museum managers have to comply with, such as a given number of visitors or a given amount of revenue. Darnell (1998) analysed the effect of such targets on admission fees for the museum. In the case of inelastic demand curves, the museum may face the problem that there is no fee attracting enough visitors and bringing in enough revenue at the same time. The demand curve may be shifted (e.g., by advertising more or improving the quality of the visitors' experience) to make the two targets mutually compatible. However, this model does not incorporate the possibility of raising revenue from sponsors, donations or ancillary goods.

The complementarities between admission fees and sales in museum stores and cafeterias affect optimal pricing strategy.<sup>36</sup> The empirical result in Steiner (1997) did not suggest that an additional free day maximises revenue, since decreased admission revenue was not compensated for by larger sales in shops and restaurants.

#### 4.3.3. Commercial activities

In addition to the core activities of museums directly related to the works exhibited or stored, most museums also engage in ancillary activities. The revenues from these

<sup>34</sup> Museum passes have become quite common in Europe and America. A pass allows free entrance in every museum in a given city or region. Ginsburgh and Zang (2001, 2003) focused in their analysis on how the revenue of such a pass can be distributed to the participant museums. Based on theoretical arguments drawn from game theory, they proposed distribution using the Shapley value.

<sup>35</sup> Oberholzer-Gee (2002) presented evidence in a field experiment where money was offered in order to be able to jump the queue. While a majority would like to jump the queue, only a minority in the queue were prepared to accept the money.

<sup>36</sup> For a general theoretical discussion of the interdependence between entrance fees and ancillary goods, see Marburger (1997).

activities can make a large contribution towards the operational expenses.<sup>37</sup> Museums operate museum shops, restaurants and cafés, sell catalogues, make money from parking lots, organise cultural trips, etc.<sup>38</sup> When the first museum shop was established by the Metropolitan Museum of Art in New York in 1908 [Weisbrod (1988, p. 109)], it was rather the exception than the rule. Today, many American museums not only operate their own shops, but also run off-site stores even in a totally different city, as does the New York Metropolitan Museum of Art.

Why do museums engage in ancillary activities? Is the museum world increasingly commercialised? Directors of museums do not necessarily want to produce ancillary goods as such; often they only serve to generate revenue for the core activity. But institutional factors may force museum directors to engage in such activities.<sup>39</sup> Weisbrod (1998, p. 58) cites the example of the British Museum where, in 1996, the government announced it would reduce its subsidies; the museum then started to discuss new possibilities of raising revenues.<sup>40</sup> However, the empirical evidence on commercialisation is ambiguous. Heilbrun and Gray (2001, p. 210) state that “Earned income accounted for only 16.1 percent of the total in 1993 but rose to 25.9 percent in 1997.” In contrast, Anheier and Toepler (1998, p. 240) concluded from their more in-depth study: “Our data suggest that art museums have not become significantly more commercial in recent years.” Segal and Weisbrod (1998) found for the arts industry that a decrease in donations (or public grants) increases ancillary activities. Much more research is needed to gain more precise knowledge.

Does the commercialisation of museums lead to a new type of museum manager? Anecdotal evidence suggests that at least in Europe arts organisations increasingly demand that their managers have some business experience.<sup>41</sup> This trend is especially marked in the so-called superstar museums to be discussed in the next section.

## 5. Current trends in the museum world

Two developments relating to museums are worth special attention: superstar museums and special exhibitions.<sup>42</sup>

<sup>37</sup> See, for instance, Heilbrun and Gray (2001, p. 211) and Anheier and Toepler (1998).

<sup>38</sup> Many books offer advice on how to maximise profit from specific services such as the management of museum stores; see, for example, Theobald (2000).

<sup>39</sup> See the data in Frey and Pommerehne (1989).

<sup>40</sup> In the case of the British Museum, and general problems of extended management discourse within museums, see Zan (2000).

<sup>41</sup> See also Anon. (2001).

<sup>42</sup> This section follows Frey’s ideas closely [Frey (1998); Frey and Busenhart (1996)].

### 5.1. Superstar museums

There are a few really well-known and world-famous museums. They can be called “superstar museums” because they have a special status setting them far apart from other museums. They are distinguished by five characteristics:

- Superstar museums are a “*must*” for tourists. They are featured prominently in guidebooks, and have achieved a cult status almost everyone is aware of. European examples include the Hermitage (St. Petersburg), the Vatican Museums (Rome), the Uffizi (Florence), the Prado (Madrid), the National Gallery (London), the Kunsthistorische Museum (Vienna), the Rijksmuseum (Amsterdam), and the Louvre (Paris). In the United States there are certainly few tourists who would not visit the Metropolitan Museum of Art and/or the Museum of Modern Art when in New York, the National Gallery of Art when in Washington, or the Art Institute when in Chicago.
- Superstar museums have *large numbers of visitors*. Over the last decades, these museums have experienced a dramatic increase in the number of visitors. In contrast, many small museums have fewer and fewer visitors.
- Superstar museums feature *world-famous painters and world-famous paintings*. Rosen (1981) originally developed the superstar idea, emphasizing that differences in income often far exceed differences in talent and performance. This also applies to artists and painters. The great disparity among artists is a striking feature of all the studies on the distribution of income among artists.<sup>43</sup> The collections in large museums comprise works by thousands of artists; only a few of them are known to art lovers, let alone to the average visitor. Museums wanting to attract a large crowd have to concentrate on a few renowned artists. Some paintings are known to virtually everyone in the Western world (and far beyond), but the number of such paintings is rather small. Examples are the “Nightwatch” in Amsterdam’s Rijksmuseum, or “Las Meninas” in the Prado. The quintessential superstar painting is Leonardo’s “Mona Lisa”; the Louvre has responded by indicating the most direct way to get to the Mona Lisa from the entrance. Even the Vatican Museum now posts the (more or less) direct way to another world famous work of art, Michelangelo’s frescos in the Sistine Chapel. From the visitors’ point of view, even very large museums are closely associated with, or defined by, very few (often one or two) paintings – the superstar phenomenon. Museums are not only the proud owners of these masterpieces, but they are also their captives. They are forced to exhibit them, but this also means that in comparison their other paintings lose prominence. There may be a slight spillover of interest to less renowned pieces in the collection, but the main effect of the superstar works is to draw attention away from the rest of the collection.

<sup>43</sup> See, e.g., Filer (1986) or Frey and Pommerehne (1989, Chapter 9).



- Superstar museums often have an *architectural design* making the building itself a world-famous artistic feature. Examples are Frank Lloyd Wright's Guggenheim Museum in New York; the Centre Pompidou in Paris; Mario Botta's San Francisco Museum of Modern Art; Frank Gehry's Guggenheim Museum in Bilbao, and Richard Meier's Getty Center in Los Angeles.
- Superstar museums are *commercialised* in two ways: A significant part of their income derives from the revenue of the museum bookshops and museum restaurants and they have a major impact on the local economy.<sup>44</sup>

The importance of these five characteristics varies amongst the different superstar museums referred to above. Ideally, they would meet all of the criteria; the Musée du Louvre is an example of one that does (the architectural feature being Ming Pei's pyramid at the entrance). Other superstar museums are very strong with respect to some characteristics, while barely meeting others. Examples would be the Getty Museum in Los Angeles, which excels with respect to the architecture (including its location) but does not have as many world famous artists and paintings as other superstar museums. Another example is Amsterdam's Rijksmuseum, which is not particularly noted for its architecture, especially when compared to the Louvre.

Superstar museums are able to exploit economies of scale by reaching out to a large number of people. They are not only featured in newspapers, on the radio and TV, but can raise enough money to produce their own videos and virtual museums. These sorts of costs are essentially independent of the number of consumers and therefore favour the major museums, because the set-up costs are normally too high for smaller institutions. While the latter will certainly catch up (a homepage will soon be a matter of course for all museums), the major museums will have the funds to improve their scope and quality so as to maintain their lead. Superstar museums have started to branch out by establishing museum networks. Thus, for example, the London Tate Gallery has spawned satellite museums at Liverpool and St. Ives and, of course, the New Tate, and the Prado has started to lend out about one third of its holdings to museums in the provinces.

Superstar museums find themselves in a new competitive situation. Their reference point shifts from other museums in the city or region to *other* superstar museums. This competition between the superstars extends over a large area, including commercial activities and sponsors.

Superstar museums must make a huge effort to stay in that category. Frantic activities are therefore often undertaken: special exhibitions are organised in the hope that they turn out to be blockbusters, visitors' amenities are improved (e.g., a larger variety

<sup>44</sup> In the case of the Guggenheim Museum in Bilbao, a discussion emerged as to how much this superstar changed the economy. While Plaza (1999) showed that the number of visitors to the Basque region increased dramatically due to the Guggenheim Museum, Gómez (1998, 2001) emphasised that one should be more cautious in analysing the effect on urban regeneration, because it is still too early to assess the economic impact of the museum on the city. For a broader discussion of "museum cultural districts", see Santagata (2002) and Chapter 31 in this volume.

of fancy restaurants) and new buildings with stunning architectural designs are added (e.g., in the case of New York's Museum of Modern Art). The superstar status tends to transform museums into providers of "total experience". This new role stands in stark contrast to the traditional notion of museums as preservers of the past.

The "total experience" offered by superstar museums, and demanded by the huge crowds of visitors, must meet two conditions:

- Art must be placed in the context of history, technology and well-known events in politics and entertainment, such as motion pictures. Superstar museums are constantly under pressure to be "special", i.e. to also somehow embed the permanent collection in a context attractive to a large numbers of visitors.
- Superstar museums must be able to provide everything, not unlike entertainment parks; the activities offered extend beyond cafés, restaurants and museum shops. Many superstar museums have already gone far in this respect; the Louvre, for example, opened a commercial precinct called "Le Carrousel du Louvre", a large underground shopping mall. Activities of superstar museums supply all sorts of educational activities (not only for children but also adults), and most importantly, plain entertainment.

## 5.2. *Special exhibitions*

There is hardly an art museum not running, or at least preparing, a special exhibition of some sort. Such an exhibition may feature one particular artist (often in commemoration of his or her birth or death) or a group of artists, or it may focus on a period or a genre of paintings, or it may establish a connection to some historical event [Belcher (1991, p. 49)]. Some special exhibitions are composed solely of paintings from the holdings of the organising museum, but most of them bring together works of art from different museums and private collections. Once put together, large temporary exhibitions frequently travel to other museums coordinated with the organisers. Some exhibitions are shown in various countries. Often, important museums have several exhibitions simultaneously which they have either organised themselves or taken over from other organisers. The boom in special exhibitions poses a challenge to art economists, because of the glaring contrast to the financial depression in which many museums find themselves. Even in some of the world's leading museums, wings are temporarily closed, and opening hours are reduced in order to save money. Curators are concerned that they have less and less money available for the restoration and conservation of their collection.

### 5.2.1. *Demand aspects*

On the demand side, special exhibitions have some special features worth noting:

- *High income effect.* Consumers tend to spend an increasing amount of their income on visiting specially arranged art exhibitions. Scattered empirical evidence exists showing that econometrically estimated income elasticities of demand for

museum visits in general are larger than one (see above). Special exhibitions thus find themselves in the comfortable position of being in a growing market.

- *Attracting new visitor groups.* As has been well documented in cultural sociology [e.g., Klein (1990)], a large percentage of the population rarely if ever visits museums (except museums of technology and transport). This applies in particular to population groups with low education which are also short of cultural tradition.<sup>45</sup> The situation is clearly different for special cultural events, which are widely advertised and which are made attractive to new groups. As special exhibitions normally take place in museum premises, they still face the difficulty of attracting new groups. This is partly overcome by “dressing-up the museum”.<sup>46</sup> special exhibitions are marked by huge banners and other advertising strategies, and even the museum entrances (which to non-museum goers can otherwise look menacing) are made welcoming. Extensive promotion also plays a role.
- *Focusing attention.* A special exhibition seeks to attract consumers by presenting them with some extraordinary cultural experience. They specialise on some particular artist (e.g., Rembrandt or van Gogh), some period (e.g., Renaissance paintings), some topic (e.g., courtly paintings), some genre (e.g., mannerist paintings), or some type of presentation (e.g., portraits). As a result, the visitors interested in such forms of art come together, often from faraway locations. Special exhibitions, in particular the blockbusters, may even be compared with major sports events such as the Olympic games or world championships. Public attention is drawn away from regular activities towards a special and unique (or at least rare) event.
- *Newsworthiness.* Special exhibitions are *news*, and attract the attention of television, radio and the print media, which is otherwise impossible to get to the same degree, and especially free of charge, while the permanent collection is hardly newsworthy.<sup>47</sup> Large exhibitions devoted to iconic artists such as Rembrandt, Van Gogh or Picasso mobilise the press and thrust the museum organisers into the limelight.<sup>48</sup>
- *Low cost to visitors.* Special exhibitions are closely linked with tourism.<sup>49</sup> A considerable number of visitors come from out of town, from another region, and often from a foreign country. Being able to combine a cultural event with tourism lowers the individual’s cost of attending in various ways. In the case of the increasingly popular package tours, the consumers only have to take the initial decision and all the rest is taken care of by the travel agent. In the case of culture, where it is often burdensome to acquire the tickets from outside, the reduction of decision and transaction costs is substantial.

<sup>45</sup> See Blau (1989) and Di Maggio and Useem (1978).

<sup>46</sup> See also Elsen (1986).

<sup>47</sup> See, e.g., Bayart and Benghozi (1993, p. 210).

<sup>48</sup> See also Elsen (1986, p. 20).

<sup>49</sup> See, e.g., Getz (1989) and O’Hagan (1992, p. 65).

- *Low price elasticity of demand.* The strong attraction of special exhibitions to tourists also affects the price elasticity of demand. Tourists relate the entrance fee to their expenditures for the overall trip. A given price increase is then, in comparison, perceived to be relatively small and does not have much impact on demand.<sup>50</sup> This effect is supported by empirical evidence. Attendance figures at the Museum of the Palazzo Ducale in Venice, for example, have been fairly stable, although admission fees for the exhibitions presented in recent years have on average increased by more than 10 percent. In fact the number of people visiting the Palazzo Ducale seems to be in direct proportion to the number of people visiting the centre of Venice [ICARE (1994)].
- *High demand by business.* Special exhibitions offer museums ample opportunities to make money. Indeed, there is a large literature documenting the monetary profitability of such cultural events.<sup>51</sup> These opportunities not only extend to the tourist industry but also to firms catering for the production of festivals and exhibitions. In the case of special exhibitions, there is also a benefit to publishers of books and catalogues.

### 5.2.2. Supply aspects

There are also various special determinants on the supply side of special exhibitions which contribute to their boom.

- *Low production cost.* The absolute cost of many special exhibitions is certainly high, but it is low compared to the sum of money they would require if all the resource inputs used were attributed to them. Important resources are taken from the permanent venues and only additional costs are covered by the special artistic events. Museum employees are involved in organising and running special exhibitions, but the corresponding cost is not attributed to the special event [Montebello (1981)]. Some cost factors, though substantial, often only appear in disguised and long-term form. One such cost is the neglect of cataloguing and maintaining the permanent collection.<sup>52</sup> But also the museum rooms, where special exhibitions take place, do not enter into the costs accounted for, as the opportunities forgone are not part of the bookkeeping.
- *More scope for artistic creativity.* Museum directors are bound by artistic conventions; for example, the particular hanging of pictures at many museums has become part of the cultural heritage, and it is next to impossible to rearrange the permanent collection to any significant extent. Special exhibitions offer a chance to avoid such historical restrictions. One of the major tasks and potentials of an art exhibition is to arrange the art objects in such a way that they create new insights and effects. In

<sup>50</sup> For the general argument, see Thaler (1980); for museum admission fees, see Blattberg and Broderick (1991).

<sup>51</sup> For example, Feldstein (1991), Fronville (1985) and Di Maggio (1985).

<sup>52</sup> See Börsch-Supan (1993) for several pertinent examples.

addition, the actual assembly of art objects coming from many different permanent collections provides a much sought-after challenge to museum directors, curators, exhibition and graphic designers, conservators, editors and managing officers, to demonstrate their artistic creativity and sense of innovation, and possibly to raise controversy – aspects which are highly valued by museum people for their own sake, but also because it is beneficial for their careers.

- *Evading government and trade union regulations.* Cultural institutions' freedom to act is restricted by two major institutions, the government and trade unions. Government restrictions go far beyond budgetary affairs. They hinder the art institutions' way of acting and performing in a myriad of ways. Thus, pricing policy is greatly restricted, as well as opening times.<sup>53</sup> In view of the strong hand exercised by the government, and its persistence due to a long tradition, engaging in special events presents a major possibility for getting round these regulations. So, for example, special exhibitions provide an opportunity for directors of art museums to appropriate at least part of the extra revenue generated. Being an extraordinary event, the museum directors are in a good bargaining position vis-à-vis the public budgetary authorities to maintain some discretion over these funds, and not to be heavily "punished" by a reduction in future budget allocations. In addition, one of the most stringent public regulations imposed on public art institutions relates to government sector employment; the virtual impossibility of dismissing inefficient or downright destructive employees, of promoting and paying employees according to performance, and of adjusting working hours to actual needs are major factors in reducing creative endeavours and turning art institutions into mere bureaucracies. Special exhibitions make it possible to avoid at least some employment restrictions, especially as most of the respective employees are only part-time and temporary, are not union members, and are therefore not legally bound by trade union regulations.
- *More sponsoring.* Politicians and public officials have an interest in special exhibitions which enable them not only to respond to the demands of the art world and the local business community, but also to appear in the media as "patrons of the arts" (with tax payers' money). Business is also more prepared to sponsor special exhibitions than regular activities, where legal provisions often hinder sponsoring. The most important reasons for the attractiveness of these events to sponsors are the higher media attention they attract, and the fact that an individual firm has more control over the funds contributed and sees less of it wasted by an inefficient bureaucracy, as is frequently the case with opera houses or art museums. Sponsors "want a well-defined, high quality event aimed at specific audience" [Anon. (1989)]. For the reasons given above, the corporate sponsors also feel that their contributions add to cultural output, and do not simply induce the government to provide less subsidy.

<sup>53</sup> For many examples, see, e.g., Börsch-Supan (1993, pp. 11, 15).

### 5.2.3. *The future of special exhibitions*

As special exhibitions become the rule rather than the exception, there is pressure to have them carry their full costs, and to subject them to the same government and trade union regulations as other museum activities. Even if the rapid rise in special exhibitions cannot be expected to persist, they will continue to have a strong and lasting impact on the art world. On the demand side, they have opened up art to an increasing number of people. This “popularisation” may not be in the interests of some art suppliers and art lovers, but from the point of view of caring for individual preferences, it is a considerable achievement. On the supply side, the increased competition between producers of art has transformed career patterns at museums and has led to a new relationship between potential and actual art consumers. By subjecting art producers at least partly to the market, it has also favoured more efficient forms of organisation and production in the world of art. These trends are likely to continue.

## 6. Conclusion

This chapter has treated various aspects of the production of museum services. From an economic point of view, two different approaches have been distinguished: first, we have shown that museums may be looked at as an economic unit where inputs and outputs can be analysed; second, we have demonstrated how the economic way of thinking can be applied to museums and individuals (directors, curators, politicians, visitors, donors, etc.) connected with them. The chapter has discussed the demand and supply side of museums, the behaviour of museums, and the phenomena of superstar museums and special exhibitions as two recent trends in the museum world. We have emphasised that the behaviour of museum management is guided by the institutional setting. According to this theory, the main source of funds has a major impact on the behaviour of the museum. The management’s decisions to raise income through ancillary services, or to manage their collection on the market, or to set entrance fees depend to a large extent on the ownership of the museum. In looking at museum behaviour, we make a distinction between private and public museums and museums dependent on donations.

A worthwhile goal for future research is to understand more fully how the changing environments in which museums are acting influence the behaviour of the museums. How will museums adapt to the new situation? The rise of superstar museums and the reliance on special exhibitions are two such notable developments. But the future may well bring new environments. Museums will no doubt rise to the occasion.

## References

- Anheier, H.K., Toepler, S. (1998). “Commerce and the muse: Are art museums becoming commercial?” In: Weisbrod, B.A. (Ed.), *To Profit or Not to Profit: The Commercial Transformation of the Nonprofit Sector*. Cambridge University Press, Cambridge, pp. 233–248.

- Anon. (1989). *The Economist*, August 5th.
- Anon. (2001). "When merchants enter the temple". *The Economist*, April 19th.
- Anon. (2002). "Sorry, we're closed: Hard times at the British Museum". *The Economist*, June 20th.
- Art Newspaper (2002). "Deeply in the Red". *The Art Newspaper*, February 1.
- Ashworth, J., Johnson, P. (1996). "Sources of 'value for money' for museum visitors: Some survey evidence". *Journal of Cultural Economics* 20, 67–83.
- Bailey, S.J., Falconer, P. (1998). "Charging for admission to museums and galleries". *Journal of Cultural Economics* 22 (2–3), 167–177.
- Bailey, S.J., Falconer, P., Foley, M., McPherson, G., Graham, M. (1998). *To Charge or Not to Charge: Full Report*. Glasgow Caledonian University, Glasgow.
- Baumol, W.J., Bowen, W.G. (1966). *Performing Arts: The Economic Dilemma*. Twentieth Century Fund, New York.
- Bayart, D., Benghozi, P.-J. (1993). *Le tournant commercial des musées en France et à l'étranger*. La Documentation Française, Paris.
- Becker, G.S. (1976). *The Economic Approach to Human Behavior*. Chicago University Press, Chicago.
- Belcher, M. (1991). *Exhibitions in Museums*. Leicester University Press and Smithsonian Institution Press, Leicester/Washington, DC.
- Benhamou, F. (1998). "The contradictions of désétatisation: Museums in France". In: Boorsma, P.B. (Ed.), et al., *Privatisation and Culture: Experience in the Arts, Heritage and Cultural Industries in Europe*. Kluwer Academic, Dordrecht, pp. 95–110.
- Benhamou, F. (2000). *L'économie de la culture*. La Découverte & Syros, Paris.
- Blattberg, R.C., Broderick, C.J. (1991). "Marketing of art museums". In: Feldstein, M. (Ed.), *The Economics of Art Museums*. University of Chicago Press, Chicago, pp. 327–346.
- Blau, J.R. (1989). "Culture as mass culture". In: Foster, A.W., Blau, J.R. (Eds.), *Art and Society: Readings in the Sociology of the Arts*. State University of New York, Albany, NY, pp. 141–176.
- Blau, J.R. (1995). "Art museums". In: Carroll, G.R., Hannan, M.T. (Eds.), *Organizations in Industry: Strategy, Structure and Selection*. Oxford University Press, Oxford, pp. 87–114.
- Blaug, M. (Ed.) (1976). *The Economics of the Arts*. Martin Robertson, London.
- Blaug, M. (2001). "Where are we now in cultural economics?" *Journal of Economic Surveys* 15 (2), 123–143.
- Börsch-Supan, H. (1993). *Kunstmuseen in der Krise: Chancen, Gefährdungen, Aufgaben in Mageren Jahren*. Deutscher Kunstverlag, Nördlingen.
- Bourdieu, P. (1979). *La distinction: Critique sociale du jugement*. Editions de Minuit, Paris.
- Cantor, J.E. (1991). "The museum's collection". In: Feldstein, M. (Ed.), *The Economics of Art Museums*. University of Chicago Press, Chicago, pp. 17–23.
- Caves, R.E. (2000). *Creative Industries: Contracts Between Art and Commerce*. Harvard University Press, Cambridge, MA.
- Clark, D.E., Kahn, J.R. (1988). "The social benefits of urban cultural amenities". *Journal of Regional Science* 28, 363–377.
- Cowen, T. (1996). "Why I do not believe in the cost-disease: Comment on Baumol". *Journal of Cultural Economics* 20 (3), 207–214.
- Darnell, A.C. (1998). "Some simple analytics of access and revenue targets". *Journal of Cultural Economics* 22 (2–3), 189–196.
- Darnell, A.C., Johnson, P.S., Thomas, R.B. (1992). "Modelling visitor flows at the Beamish museum". In: Johnson, P.S., Thomas, R.B. (Eds.), *Choice and Demand in Tourism*. Mansell, London, pp. 167–174.
- de Montebello, P. (1981). "Exhibitions and permanent museum collection – competition or correspondence". In: Institut für Museumskunde (Ed.), *Ausstellungen – Mittel der Politik?* Gebr. Mann Verlag, Berlin, pp. 153–163.
- Dickenson, V. (1997). "Museum visitor surveys: An overview, 1930–1990". In: Towse, R. (Ed.), *Cultural Economics: The Arts, the Heritage and the Media Industries*, vol. 1. Edward Elgar, Cheltenham, pp. 272–281.
- Di Maggio, P. (1985). "When the profit is quality: Cultural institutions in the marketplace". *Museum News* 63 (5), 28–35.

- Di Maggio, P. (Ed.) (1986). *Nonprofit Enterprise in the Arts*. Oxford University Press, New York.
- Di Maggio, P., Useem, M. (1978). "Cultural democracy in a period of cultural expansion: The social composition of arts audiences in the United States". *Social Problems* 26, 179–197.
- Elsen, A. (1986). "Assessing the pros and cons". *Special Section: Museum Blockbusters, Art in America*, June.
- Feldstein, M. (Ed.) (1991). *The Economics of Art Museums*. University of Chicago Press, Chicago.
- Filer, R.K. (1986). "The 'starving artist' – myth or reality? Earnings of artists in the United States". *Journal of Political Economy* 94, 56–75.
- Foster, A.W., Blau, J.R. (Eds.) (1989). *Art and Society. Readings in the Sociology of the Arts*. State University of New York Press, Albany.
- Frey, B.S. (1994). "Cultural economics and museum behaviour". *Scottish Journal of Political Economy* 41, 325–352.
- Frey, B.S. (1998). "Superstar museums: An economic analysis". *Journal of Cultural Economics* 22, 113–125.
- Frey, B.S. (1999). *Economics as a Science of Human Behaviour*. Kluwer Academic, Dordrecht.
- Frey, B.S. (2000). *Arts and Economics*. Springer, Heidelberg.
- Frey, B.S., Busenhardt, I. (1996). "Special exhibitions and festivals; culture's booming path to glory". In: Ginsburgh, V.A., Menger, P.-M. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier/North-Holland, Amsterdam, pp. 275–302.
- Frey, B.S., Pommerehne, W.W. (1989). *Muses and Markets: Explorations in the Economics of the Arts*. Blackwell, Oxford.
- Fronville, C.L. (1985). "Marketing for museums: For-profit techniques in the non-profit-world". *Curator* 28 (3), 169–182.
- Gapinski, J.H. (1984). "The economics of performing Shakespeare". *American Economic Review* 74 (3), 458–466.
- Gapinski, J.H. (1986). "The lively arts as substitutes for the lively arts". *American Economic Review* 76 (2), 20–25.
- Getz, D. (1989). "Special events: Defining the product". *Tourism Management* 10 (2), 125–137.
- Ginsburgh, V., Mairesse, F. (1997). "Defining a museum: Suggestions for an alternative approach". *Museums Management and Curatorship* 15 (4), 15–33.
- Ginsburgh, V., Menger, P.-M. (Eds.) (1996). *Economics of the Arts: Selected Essays*. Elsevier/North-Holland, Amsterdam.
- Ginsburgh, V., Zang, I. (2001). "Sharing the income of a museum pass program". *Museum Management and Curatorship* 19 (4), 371–383.
- Ginsburgh, V., Zang, I. (2003). "The museum pass game and its value". *Games and Economic Behavior* 43, 322–325.
- Glaeser, E.L. (2001). "The governance of not-for-profit firms". Harvard Institute of Economic Research, Discussion Paper No. 1954.
- Glaeser, E.L., Shleifer, A. (2001). "Not-for-profit entrepreneurs". *Journal of Public Economics* 81 (1), 99–115.
- Glazer, A., Konrad, K.A. (1996). "A signaling explanation of charity". *American Economic Review* 86 (4), 1019–1028.
- Globerman, S., Book, S.H. (1974). "Statistical cost functions for performing arts organizations". *Southern Economic Journal* 40, 668–671.
- Gómez, M.V. (1998). "Reflective images: The case of urban regeneration in Glasgow and Bilbao". *International Journal of Urban and Regional Research* 22 (1), 106–121.
- Gómez, M.V. (2001). "A reply to Beatriz Plaza's 'The Guggenheim–Bilbao museum effect' ". *International Journal of Urban and Regional Research* 25 (4), 898–900.
- Goudriaan, R., Van't Eind, G.J. (1985). "To fee or not to fee: Some effects of introducing admission fees in four museums in Rotterdam". In: Owen, V.L., Hendon, W.S. (Eds.), *Managerial Economics for the Arts*. Association for Cultural Economics, Akron, OH, pp. 103–109.
- Gramp, W.D. (1989). *Pricing the Priceless: Art, Artists and Economics*. Basic Books, New York.



- Grampp, W.D. (1996). "A colloquy about art museums: Economics engages museology". In: Ginsburgh, V.A., Menger, P.-M. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier/North-Holland, Amsterdam, pp. 221–254.
- Gray, C.M. (1998). "Hope for the future? Early exposure to the arts and adult visits to art museums". *Journal of Cultural Economics* 22 (2–3), 87–98.
- Hansmann, H.B. (1981). "Nonprofit enterprise in the performing arts". *Bell Journal of Economics* 12 (2), 341–361.
- Harbaugh, W.T. (1998). "What do donations buy? A model of philanthropy based on prestige and warm glow". *Journal of Public Economics* 67, 169–284.
- Heilbrun, J., Gray, C.M. (2001). *The Economics of Art and Culture*. Cambridge University Press, Cambridge.
- Hirshleifer, J. (1985). "The expanding domain of economics". *American Economic Review* 75, 53–68.
- ICARE, International Centre of Studies on the Economics of the Arts (1994). *I musei Veneziani. Indagine sulle strutture permanenti e temporanee. Offerta culturale della città di Venezia, Venezia*.
- Jackson, R. (1988). "A museum cost function". *Journal of Cultural Economics* 12, 41–50.
- James, E. (1983). "How nonprofits grow: A model". *Journal of Policy Analysis and Management* 2, 350–365.
- Johnson, P., Thomas, B. (Eds.) (1998). "Special issue on the economics of museums". *Journal of Cultural Economics* 22, 75–207.
- Kindermann, C. (1903). *Volkswirtschaft und Kunst*. Fischer, Jena.
- Kirchgässner, G. (1991). *Homo Oeconomicus: Das ökonomische Modell individuellen Verhaltens und seine Anwendung in den Wirtschafts- und Sozialwissenschaften*. Mohr (Siebeck), Tübingen.
- Klein, H.J. (1990). *Der gläserne Besucher: Publikumsstrukturen einer Museumslandschaft*. Gebr. Mann Verlag, Berlin.
- Lange, M.D., Bullard, J., Luksetich, W.A., Jacobs, P. (1985). "Cost functions for symphony orchestras". *Journal of Cultural Economics* 9 (2), 71–85.
- List, J.A., Lucking-Reiley, D. (2002). "The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign". *Journal of Political Economy* 110 (1), 215–233.
- Lord, B., Dexter, G., Nicks, J. (1989). *The Cost of Collecting: Collection Management in UK Museums*. Her Majesty's Stationery Office, London.
- Luksetich, W.A., Partridge, M.D. (1997). "Demand functions for museum services". *Applied Economics* 29, 1553–1559.
- Maddison, D. (2004). "Causality and museum subsidies". *Journal of Cultural Economics* 28, 89–108.
- Maddison, D., Foster, T. (2003). "Valuing congestion costs in the British Museum". *Oxford Economic Papers* 55, 173–190.
- Mairesse, F., Vanden Eeckaut, P. (2002). "Museum assessment and FDH technology: Towards a global approach". *Journal of Cultural Economics* 26, 261–286.
- Marburger, D.R. (1997). "Optimal ticket pricing for performance goods". *Managerial and Decision Economics* 18, 375–381.
- Martin, F. (1994). "Determining the size of museum subsidies". *Journal of Cultural Economics* 18, 255–270.
- Meier, S., Frey, B.S. (2003). "Private faces in public places: The case of a private art museum in Europe". *Cultural Economics* 3, 1–16.
- Montias, J.M. (1973). "Are museums betraying the public's trust?" *Museum News* 51, 25–31.
- Moulin, R. (1986). "Le marché et le musée: La constitution des valeurs artistiques contemporaines". *Revue Française de la Sociologie* 27, 369–395.
- National Endowment for the Arts (2000). *International Data on Government Spending on the Arts*. NEA Research Division Note No. 64, Washington, DC.
- Noble, J.V. (1970). "Museum manifesto". *Museum News* April, 27–32.
- Noonan, D. (2002). "Contingent valuation studies in the arts and culture: An annotated bibliography". Working Paper Series, Cultural Policy Center Working Paper No. 11, University of Chicago.
- Oberholzer-Gee, F. (2002). "Selling the right to jump the queue: Pricing decisions and competing principles of fairness". Working Paper Series, Harvard Business School.
- O'Hagan, J.W. (1992). "The Wexford Opera Festival: A case for public funding?" In: Towse, R., Khakee, A. (Eds.), *Cultural Economics*. Springer, Berlin, pp. 61–66.

- O'Hagan, J.W. (1995). "National museums: To charge or not to charge?" *Journal of Cultural Economics* 19, 33–47.
- O'Hagan, J.W. (1998a). *The State and the Arts: An Analysis of Key Economic Policy Issues in Europe and the United States*. Edward Elgar, Cheltenham.
- O'Hagan, J.W. (1998b). "Art museums: Collections, deaccessioning and donations". *Journal of Cultural Economics* 22 (2–3), 197–207.
- O'Hare, M., Feld, A.L. (1975). "Is museums speculation in art immoral, illegal and insufficiently fattening?" *Museum News* 53, 25–31.
- Oster, S., Goetzmann, W.N. (2001). "Does governance matter? The case of art museums". Working Paper Series, NBER, Washington, DC.
- Peacock, A.T. (1969). "Welfare economics and public subsidies to the arts". *Manchester School of Economics and Social Studies* 37, 323–335.
- Peacock, A.T. (1993). *Paying the Piper: Culture, Music and Money*. Edinburgh University Press, Edinburgh.
- Peacock, A.T., Godfrey, C. (1974). "The economics of museums and galleries". *Lloyds Bank Review* 111, 17–28.
- Peacock, A.T., Rizzo, I. (1994). *Cultural Economics and Cultural Politics*. Kluwer Academic, Dordrecht.
- Plaza, B. (1999). "The Guggenheim–Bilbao museum effect". *International Journal of Urban and Regional Research* 23, 589.
- Pommerehne, W.W., Feld, L. (1997). "The impact of museum purchase on the auction prices of paintings". *Journal of Cultural Economics* 21, 249–271.
- Prieto Rodríguez, J., Fernandez Blanco, V. (2002). "Optimal pricing and grant policies for museums". Working Paper Series, Instituto de Estudios Fiscales, Madrid.
- Ribar, D.C., Wilhelm, M.O. (2002). "Altruistic and joy of giving motivations in charitable behavior". *Journal of Political Economy* 110 (2), 425–457.
- Robbins, L. (1963). "Art and the State". In: Robbins, L.C. (Ed.), *Politics and Economics: Essays in Political Economy*. Macmillan, London, pp. 53–72.
- Robbins, L. (1971). "Unsettled questions in the political economy of the arts". *Three Banks Review* 91, 3–19.
- Rose-Ackerman, S. (1996). "Altruism, nonprofits and economic theory". *Journal of Economic Literature* 23, 701–728.
- Rosen, S. (1981). "The economics of superstars". *American Economic Review* 71, 845–858.
- Rosett, R.N. (1991). "Art museums in the United States: A financial portrait". In: Feldstein, M. (Ed.), *The Economics of Art Museums*. University of Chicago Press, Chicago, pp. 129–177.
- Santagata, W. (2002). "Cultural districts, property rights and sustainable economic growth". *International Journal of Urban and Regional Research* 26 (1), 9–23.
- Schiff, J., Weisbrod, B.A. (1991). "Competition between for-profit and nonprofit organizations in commercial markets". *Annals of Public and Cooperative Economics* 62 (4), 619–639.
- Schulze, G.G., Ursprung, H.W. (2000). "La Donna e mobile – or is she? Voter preferences and public support of the performing arts". *Public Choice* 102, 131–149.
- Schuster, J.M. (1985). *Supporting the Arts: An International Comparative Study*. Massachusetts Institute of Technology, Cambridge, MA.
- Schuster, J.M. (1986). "Tax incentives as arts policy in Western Europe". In: DiMaggio, P.J. (Ed.), *Nonprofit Enterprise in the Arts: Studies in Mission and Constraint*. Oxford University Press, New York, pp. 320–360.
- Schuster, J.M. (1998a). "Neither public nor private: The hybridization of museums". *Journal of Cultural Economics* 22, 127–150.
- Schuster, J.M. (1998b). "Beyond privatization: The hybridization of museums and the built heritage". In: Boorsma, P.B., et al. (Eds.), *Privatization and Culture: Experience in the Arts, Heritage and Cultural Industries in Europe*. Kluwer Academic, Dordrecht, pp. 58–81.
- Seaman, B.A. (1997). "Arts impact studies: A fashionable excess". In: Towse, R. (Ed.), *Cultural Economics: The Arts, the Heritage and the Media Industries*. Edward Elgar, Cheltenham, pp. 723–755.
- Seaman, B.A. (2002). "CVM vs. economic impact: Substitutes or complements". Mimeo, Georgia State University.

- Segal, L.M., Weisbrod, B.A. (1998). "Interdependence of commercial and donative revenues". In: Weisbrod, B.A. (Ed.), *To Profit or Not to Profit: The Commercial Transformation of the Nonprofit Sector*. Cambridge University Press, Cambridge, pp. 105–127.
- Smolensky, E. (1986). "Municipal financing of the U.S. Fine Arts Museum: A historical rationale". *Journal of Economic History* XLVI (3), 757–768.
- Steiner, F. (1997). "Optimal pricing of museum admission". *Journal of Cultural Economics* 21, 307–333.
- Sunstein, C.R. (2002). "Coherent and incoherent valuation: A problem with contingent valuation of cultural amenities". Working Paper Series, The Cultural Policy Center at the University of Chicago.
- Temin, P. (1991). "An economic history of American art museums". In: Feldstein, M. (Ed.), *The Economics of Art Museums*. University of Chicago Press, Chicago, pp. 179–194.
- Thaler, R.H. (1980). "Toward a positive theory of consumer choice". *Journal of Economic Behavior and Organization* 1, 39–60.
- Theobald, M.M. (2000). *Museum Store Management*. Altamira Press, Walnut Creek.
- Thompson, N.L. (1986). "Financially troubled museums and the law". In: DiMaggio, P.J. (Ed.), *Nonprofit Enterprise in the Arts: Studies in Mission and Constraint*. Oxford University Press, New York, pp. 214–242.
- Throsby, D. (1977). "Production and cost relationships in the supply of performing arts services". In: Tucker, K.A. (Ed.), *Economics of the Australian Services Sector*. Croom Helm, London, pp. 414–432.
- Throsby, D. (1994). "The production and consumption of the arts: A view of cultural economics". *Journal of Economic Literature* 33, 1–29.
- Towse, R. (1997a). *Cultural Economics: The Arts, the Heritage and the Media Industries*. Edward Elgar, Cheltenham.
- Towse, R. (Ed.) (1997b). *Baumol's Cost Disease: The Arts and Other Victims*. Edward Elgar, Cheltenham.
- van Hemel, A., van der Wielen, N. (Eds.) (1997). *Privatisation/Désétatisation and Culture: Limitations or Opportunities for Cultural Development in Europe? Conference Reader for the Circle Round Table*, Amsterdam.
- Wall, G., Roberts, C. (1984). "The economic impact of the Tutankhamun and van Gogh exhibitions". In: Hendon, W.S., Grant, N.K., Shaw, D.V. (Eds.), *The Economics of Cultural Industries*. Association for Cultural Economics, University of Akron, pp. 66–76.
- Weil, S.E. (1987). "Deaccession practices in American museums". *Museum News* 65, 44–50.
- Weil, S.E. (2002). *Making Museums Matter*. Smithsonian Institution Press, Washington.
- Weisbrod, B.A. (1977). *The Voluntary Nonprofit Sector: An Economic Analysis*. D.C. Heath and Company, Lexington, MA.
- Weisbrod, B.A. (1988). *The Nonprofit Economy*. Harvard University Press, Cambridge, MA.
- Weisbrod, B.A. (1998). "Modeling the nonprofit organization as a multiproduct firm: A framework for choice". In: Weisbrod, B.A. (Ed.), *To Profit or Not to Profit: The Commercial Transformation of the Nonprofit Sector*. Cambridge University Press, Cambridge, pp. 47–64.
- White, J.L. (1996). "When it's OK to sell the Monet: A trustee-fiduciary-duty framework for analysing the deaccessioning of art to meet museum operating expenses". *Michigan Law Review* 94 (4), 1041–1066.
- Withers, G.A. (1980). "Unbalanced growth and the demand for performing arts: An econometric analysis". *Southern Economic Journal* 46 (3), 735–742.
- Zan, L. (2000). "Management and the British Museum". *Museum Management and Curatorship* 18 (3), 221–270.

## CULTURE IN URBAN AND REGIONAL DEVELOPMENT\*

TRINE BILLE

*AKF, Institute of Local Government Studies, Denmark*

GÜNTHER G. SCHULZE

*University of Freiburg, Germany*

### Contents

Abstract	1052
Keywords	1052
1. Introduction	1053
1.1. On the relationship between economic and cultural development	1053
1.2. The concepts of regional culture and regional development	1054
2. How can arts and culture generate urban and regional development?	1057
2.1. Short-run effects	1057
2.1.1. Measuring short-run spending impacts – economic impact studies	1057
2.1.2. Applications 1: The impact of a single event or institution	1059
2.1.3. Applications 2: The impact of the arts on a town, a municipality or a region	1061
2.1.4. The role of tourism	1062
2.1.5. Assessment and critique	1064
2.2. Long-run effects	1066
2.2.1. Location choice	1066
2.2.2. Educational and creativity values	1068
2.3. Cultural and urban planning	1069
3. How does regional economic growth influence the development of the cultural sector?	1071
3.1. Market prices and non-market values	1071
3.2. The demand for cultural goods	1073
3.2.1. Observable demand behavior	1073
3.2.2. Non-market demand	1075
3.3. The supply of cultural goods	1079
3.3.1. Production process in the arts	1079
3.3.2. Baumol's cost disease	1080

\* We are grateful to Bruce Seaman and David Throsby for helpful comments.

3.3.3. Innovation, globalization and cultural change	1081
3.3.4. Trade	1082
3.4. Development and public support for the arts	1083
3.5. Assessment: The influence of development on culture	1084
4. Location of arts and cultural industries	1086
4.1. Delineation of the cultural industries	1086
4.2. Agglomeration economies and the location of arts and culture	1087
4.2.1. Proximity to market	1087
4.2.2. Economies of scale and scope in production	1088
4.2.3. Input markets	1089
4.2.4. Creativity spillovers and the location of artists	1089
4.3. The particular role of cities	1090
5. Concluding remarks: Regional cultural and economic development – a complex simultaneity	1091
References	1093

## Abstract

This chapter critically assesses, from an economic viewpoint, the role of the arts and culture in urban and regional development and growth. This includes the analysis of short run spending impacts, and longer term effects on location quality and creativity. In addition, the specific possibilities for using arts and cultural activities as a focal point in strategies for urban revitalization are discussed including the role of tourism, the non-market demand for arts and culture as an element in local willingness to support urban and regional development policies, questions of sustainability, etc. The reverse influence of regional economic development on the demand for and supply of culture in the region is analyzed and the location of arts and cultural industries is discussed highlighting the role of agglomeration of cultural industries.

## Keywords

regional and urban development, arts and culture, economic development, economic impact, cultural tourism

*JEL classification:* Z11, O18, R1, R5, H7

## 1. Introduction

### 1.1. On the relationship between economic and cultural development

What is the relationship between economic and cultural development at the urban or regional level? How does culture change in the course of economic development at these levels and conversely what contribution can culture make to overall development? These are important questions not only for urban, regional and cultural economics but also for cultural policy-makers and urban planners, given that culture has become an integral part of regional development strategies. Urban planning has increasingly become intertwined with cultural planning, and cultural programming is used as part of strategies to revitalize inner cities and old industrial areas.

This chapter contains a critical assessment, from an economic viewpoint, of recent research on the role of the arts and culture in urban development and redevelopment and in regional growth. It addresses the relationships between regional economic development and development of the cultural sector. This interdependence obviously has a dual causality – development of the cultural sector has an effect on overall economic development and at the same time general economic development affects the cultural sector in specific ways. These two directions of causality are explored in this chapter. First, we ask what is the contribution of cultural activities to economic growth and *how* do arts and culture generate urban and regional development? We thus focus on outcomes in terms of employment and income growth as well as on transmission channels through which these effects are realized. Examples of such transmission channels are cultural tourism which stimulates the tourism industries more generally, and improved location quality through better arts and cultural services that attract mobile firms or highly skilled labor. Of particular interest in this context is the question as to whether cultural industries tend to cluster, and if so why. If there are agglomeration advantages for cultural industries they will lead to an unbalanced growth of the cultural sector, resulting, for instance, in an urban–rural divide in the provision of cultural services; if this were to occur, the contribution of cultural industries to regional economic growth would be unevenly distributed.

Second, we investigate the reverse causality: How does regional economic growth influence the development of the cultural sector? Does the cultural sector benefit from the overall economic growth in a city, a metropolitan area or a region, and if so, what are the mechanisms through which this improvement is brought about? We can distinguish three lines of explanation: a demand-side effect, a supply-side effect, and altered support for the arts from the government and private sponsors. The demand-side effect is straightforward: in the course of economic development people become richer and typically also better educated; because demand for culture has been shown to depend positively on income and education this increased demand should result in a better provision of arts and culture as more cultural activities become commercially profitable. The supply-side effect is more intricate and works indirectly. On the one hand, economic development provides better infrastructure that makes remote areas more accessible for

traveling orchestras and the like and allows inhabitants to travel more easily to cultural centers, thereby effectively improving the cultural services provided for them. This also applies to broadcast cultural services such as radio and television sets become more widespread. On the other hand, technological progress is unevenly distributed across industries; many cultural sub-sectors such as orchestras, theaters, painters and fine art are much less subject to (labor saving) technical progress than, for instance, manufacturing. This leads to a rising *share* of the workforce employed in arts and culture and a rising relative price for cultural goods produced by these sub-sectors, other things being equal (the so-called “Baumol’s cost disease”). Finally, in many countries public subsidies to the arts account for the lion’s share of the revenues of performing arts institutions (theaters, orchestras, operas, ballets, etc.) and museums. Such subsidies could be expected to increase with regional economic development, thereby tending to improve the supply of arts and culture. The reasons are not only increased budgets of regional governments, but also increased demand for culture due to economic development (see above) which, depending on the political system, in turn may lead to higher subsidies for arts and culture through the political process.<sup>1</sup>

### 1.2. *The concepts of regional culture and regional development*

We distinguish three definitions of culture:

- *Culture as aspect*, i.e. the community of ideas, values, norms and habits that a society has.
- *Culture as sector*, i.e. if culture is defined by a series of genres like painting, sculpture, theater, dance, music, literature, poetry, film, video, architecture, etc., the cultural sector consists of those institutions, firms, organizations and individuals who work with these genres; in total the cultural sector so defined comprises a wide spectrum from commercial businesses on the one hand to subsidized cultural institutions and idealistic organizations on the other.
- *Culture as art*, where the word “art” includes an implicit quality valuation; for instance, not all paintings would be classified as art according to some quality criteria.

This chapter’s point of departure is the definition of culture as a sector. Thus, there are no underlying quality valuations regarding the definition, but only a specification of genre. For our purposes the cultural sector contains both the wide commercial culture and the narrower traditional culture, comprising artistic modes of expression like theater, music, visual art, architecture, design, arts and crafts, media, film/video, multimedia, literature and cultural heritage; it could also include fashion, advertising, computer software, games and toys, amusement parks, etc. Researchers in this field refer to arts and culture as an economic activity. That does not necessarily imply that these activities are

<sup>1</sup> On the other hand, however, the increase in relative prices for art tends to erode purchasing power of cultural budgets. The net effect on real subsidies for culture is thus an empirical issue.

predominantly motivated by economic considerations (income generation) nor does it require that the results of these activities are traded in the market place. In fact many cultural products, including items of regional or national heritage, are not sold and bought freely on the market. Yet in the creation of arts and culture significant flows of financial resources are involved.

We disregard leisure time cultural activities which are (almost) free such as unsponsored neighborhood or community cultural activities, music groups that play for fun only, and so on. This is clearly an omission as these activities are an integral part of regional culture. Presumably the extent and structure of these activities will change in the course of economic development and thus will change regional and local culture. However they have not been subject to extensive economic research and therefore will be disregarded in what follows.

How do we define economic *development*? First and foremost, economic development implies economic growth, i.e. an increase in goods and services produced by a (regional) economy in a given period of time. This typically goes in tandem with increased employment and enhanced standards of living.<sup>2</sup> Economic growth is thus the motor for economic development in a broader sense; it typically increases life expectancy at birth, quality of health care, reduces infant mortality and allows for shorter working hours including a reduction in child labor.<sup>3</sup> It frees more time and resources for cultural consumption and, through increased market demand, mobilizes more resources for its production. According to [Kindleberger and Herrick \(1977\)](#) economic development implies both more output and changes in the technical and institutional arrangements by which it is produced and distributed. Development implies changes in functional capacities – in physical coordination, for example, or learning capacity (or ability of the economy to adapt). Thus economic development might be viewed as a longer term increase in the capacity of a local economy to enhance the quality of life to its residents; this will require increases in human productivity and a balance among the economy's component sub-regions that increases its ability to coordinate economic activity and adapt to changing circumstances.

[Throsby \(2003, p. 183\)](#) has introduced a related concept of cultural sustainability which he defines as arising from the broader notion of sustainable development:

... a concept that marries the ideas of sustainable *economic* development, meaning development that will not slow down or wither away but will be in some sense self-perpetuating, and ecological sustainability, meaning the preservation and enhancement of a range of environmental values through the maintenance of ecosystems in the natural world. Furthermore, the term “sustainable development” embraces an interpretation of “economic development” that supersedes former notions of economic growth measured only in terms of increases in per capita GDP,

<sup>2</sup> The deficiencies of GDP as a measure of well-being are well understood by now and need not be detailed here. In particular GDP measurement disregards externalities (such as environmental degradation) and most non-monetized transactions.

<sup>3</sup> Cf. [Easterly \(2001\)](#) and references cited therein.



and replaces it with the wider concept of “human development”, focused on the individual as both the instrument and the object of development, and measured by a variety of indicators of quality of life and standards of living that go well beyond measuring simply material progress.<sup>4</sup>

As a definition of cultural sustainability seems impossible to him, Throsby (2001, 2003) argues that a list of criteria may be used instead by which to assess sustainable management of cultural capital: material and non-material well-being, intergenerational equity, intragenerational equity, maintenance of diversity, precautionary principle, maintenance of cultural systems and recognition of interdependence. The final criterion, in essence, draws together the entire concept of sustainability when applied to culture, providing an overall framework within which the other more specific principles can be seen to operate.

The concept of economic growth is of course simplest to apply in empirical studies, the concepts of economic development and cultural sustainability being more and more complex – and less operational. Most studies of arts and culture in urban and regional development thus apply the simple measures of economic growth such as income and employment. Broader concepts like cultural sustainability are interesting, as they evaluate the impact of culture in a broader perspective, including wider aspects of development. Therefore it would be interesting in principle to see more studies in the future applying concepts like cultural sustainability to measure the impact of arts and culture on urban and regional development. However, economic growth is a more clearly defined concept and thus much more operational,<sup>5</sup> and in any case it can be seen as the main driving force behind economic development and cultural sustainability.

This chapter proceeds as follows. First we will analyze how arts and culture promote urban and regional development in the short run by increasing overall demand for regional goods and services (Section 2.1), and in the longer run through increased attractiveness of a region for firms, people and economic activity in general (Section 2.2). Then we look at long run educational and creativity values of arts and culture for a region. We switch perspective in Section 3 to an analysis of how economic development will change arts and culture; we look in Section 3.2 at the changing demand for cultural goods in the course of development, and in Section 3.3 at changing supply that is affected by a productivity difference for the art sector (Baumol’s cost disease), technological change and globalization of culture. Remarks on changing patterns of public

<sup>4</sup> The most commonly quoted definition of sustainable development is that put forward by the UN World Commission on Environment and Development (“the Brundtland Commission”) in 1987, which specified sustainable development as “development that meets the needs of the present generation without compromising the ability of future generations to meet their own needs”. A subsequent UN commission, the World Commission on Culture and Development, which reported in 1995, carried these ideas through to the arena of cultural development, where again the long-term needs of future generations for access to cultural resources can be seen to be important.

<sup>5</sup> In particular it is not clear how the various aspects of cultural sustainability shall be weighed in order to derive a measurement for cultural sustainability.

support for the arts and a final assessment round off this section. The location of arts and culture is discussed in Section 4 where we look at reasons for agglomeration of cultural industries and the relationship of urban versus suburban and hinterland development. Concluding remarks follow in Section 5.

## 2. How can arts and culture generate urban and regional development?

The role of arts and culture as a way of generating urban and regional development has been widely canvassed for some time. In this section we will look at the economic aspects of this, the methods applied and the magnitude of the effects involved. Arts and culture can generate urban and regional development through effects that can be divided into short-run spending impacts and long-run growth impacts. The short-run spending impacts are mainly due to the fact that the arts and culture can attract visitors – local and non-local consumers, who spend money in the local area (Section 2.1). The long-run growth impacts discussed in Section 2.2 can be divided into two different types: arts and culture as a localization factor for people, companies and investments (Section 2.2.1), and the impact of culture on creativity, etc. (Section 2.2.2). Arts and culture can also play an important role for both image and knowledge of a town or a region; this in itself is not important for economic development, but it could be a prerequisite for the creation of positive economic effects by attracting tourists, inhabitants and firms. In Section 2.3 we look into strategies for cultural and urban planning that take such a broader perspective on urban and regional development.

### 2.1. Short-run effects

The short-run spending impacts are due to the fact that the arts and culture can attract visitors – local and non-local consumers – who spent money on culture, but sometimes also on related goods and services like food, beverages, accommodation, shopping, etc. If there are visitors who have come to a city or a region because of the cultural activities on offer, their consumption constitutes an increase in aggregate demand that can be attributed to the arts. These kind of short-run spending impacts are mainly based on cultural tourism, but can also be due to the possibility that the region's own population spends less elsewhere.

#### 2.1.1. Measuring short-run spending impacts – economic impact studies

The most frequently applied method to study the short-run spending effects of the arts is economic impact analysis. This approach is a well-established and frequently-used methodology in economics, in which the impact of one activity on the rest of the economy is calculated using traditional multiplier models. In cultural economics many such studies have been carried out during the last 25 years,<sup>6</sup> with Cwi and Lyall (1977) being

<sup>6</sup> See Radich (1987, 1993) for an extensive overview.

one of the pioneering applications. In economic impact studies, the short-term spending impact of the arts on consumption, income and employment is measured, normally differentiating between direct, induced and indirect effects:

- *The direct effects* are the employment and income generated in a locality by the arts activities themselves.
- *The induced effects* (or linkage effects) – the customer effects – measure the spending of visitors to cultural events or institutions on transport, food and drink and perhaps accommodation.
- *The indirect effects* are multiplier effects, which are associated with both the direct and the induced effects.

The total impact can be calculated as:

$$\text{Total Impact} = \text{Direct Effects} + \text{Induced Effects} + \text{Indirect Effects},$$

$$\text{where Indirect Effects} = (\text{Direct Effects} + \text{Induced Effects}) \times \text{Multiplier}.$$

This approach typically assumes that there is sufficient free capacity in the local area in the form of unemployment.<sup>7</sup> Note that a short-run impact model has to be distinguished from a general equilibrium model, in which the required (cultural) production competes with other sectors in input and output markets. General equilibrium models can give very different results from impact analyses, particularly because they typically focus on the longer run.

Impact studies have been classified under the term “multiplier analysis”, although there are different types: export base, traditional Keynesian, input-output and proportion multiplier models [Cooper et al. (1993)]. The magnitude of the multiplier depends on interregional linkages; it will be lower in smaller, less self-sufficient geographical regions than in larger, more self-sufficient ones.

A sophisticated economic impact study should attempt to answer the question: “How much would short-run economic activity decline in a specific region if X were no longer to exist?” In other words the analysis should only deal with the net effects on consumption, income and employment, not with the gross effects, which means that care should be taken to distinguish between net injections into the region from tourists or other external sources and diversions of local spending. Extra expenditure attracted from outside the region is what matters most; thus in impact studies in the cultural sector it is spending by cultural tourists that is the main measured source of short-run impact on the regional economy. A cultural establishment or event which draws a considerable number of tourists will have a much higher impact than one which caters more to the interests of the locals. In other words, the more tourists that can be attracted by the arts and culture, the higher the economic impact.<sup>8</sup> This implies that there is a contrast between the local and the national level. Spending which happens to be additional at a

<sup>7</sup> Otherwise, multiplier effects would be lower and price effects for cultural products would have to be taken into account.

<sup>8</sup> See further Section 2.1.4 below.

local or a regional level (and therefore leading to economic growth) might not be additional at a national level, for example, if domestic tourist spending in one region is mostly at the expense of other regions; in the latter case the economic effects will be higher at the local/regional level than at the national level.

Besides, alternatives must be considered. If these kinds of studies are used as arguments for public support for the arts and culture<sup>9</sup> it must also be realized that it is not enough to show that the arts and culture can generate income and employment, because all economic activity does that. It is necessary (though difficult) to investigate whether the arts stimulate economic development more than if the subsidies had been used for alternative purposes such as developing infrastructure or subsidizing sports events.

Economic impact studies in the cultural sector have been conducted at different levels including the following:

- a single cultural event, e.g., a festival, or individual permanent cultural institutions;
- the entire cultural life of a town, municipality or a region; or
- the cultural life of an entire state or nation.

For our present purposes, analyses of the cultural life of an entire state or nation<sup>10</sup> are not very relevant, and so we concentrate in the following two sections on presenting several examples of studies in the other two categories.

### *2.1.2. Applications 1: The impact of a single event or institution*

Few comparisons or generalizations can be drawn from studies dealing with the economic impact of a single cultural event or individual permanent cultural institution, because the impact differs from event to event or institution to institution. Examples of these studies are the economic impacts of: eight cultural institutions in Baltimore [Cwi and Lyall (1977)], the Staatsoper, Volksoper, Burgtheater and Akademietheater in Vienna [Abele and Bauer (1984)], the Edinburgh Festival [Vaughan (1979)], the Stratford Festival in Ontario [Mitchell and Wall (1989)], three Canadian theater companies [Mitchell (1989)], the open Air Museum at Beamish [Johnson and Thomas (1992)], the Renoir and Barnes Art Exhibit [Stanley et al. (2000)] and the illustrative cases discussed in more detail in the following paragraphs.<sup>11</sup>

Christoffersen and Lyk-Jensen (1994) show the importance of cultural tourism in their impact study of a Danish art institution. They use an input-output model for the local area, and their study is one of the few that tries to compare the economic impact of an art institution with alternative ways of using the money. The cultural institution is Brandts Klaedefabrik, an old textile factory in Odense (a large provincial town in

<sup>9</sup> Economic impact studies often had that function; see, for example, van Puffelen (1996).

<sup>10</sup> For example, Myerscough (1988a) and Hummel and Berger (1988). Lately, creative industries mapping have been conducted in several countries based on national statistics [e.g., Department of Culture, Media and Sports (1998, 2001)]. See further Section 4.1 in this chapter.

<sup>11</sup> See also the guidelines for calculation of the economic impact of America's orchestras published by the American Symphony Orchestra League [Cooper (1997)].

Denmark) housing exhibition premises including Brandt's Textile Factory Art Gallery, the Museum of Photographic Art, Denmark's Museum of Graphic Art, and the Danish Press Museum. A two-phase analysis was carried out. First, demand with regard to the cultural establishment was determined by interviewing visitors. Second, the regional and national economic effects were calculated using economic models and a regional input-output table. The economic impact of the art institution was compared with alternative ways of using the money by way of two supplementary analyses: the effects of Brandts Klaedefabrik were compared first to the effects of another cultural establishment in Odense, the house of Hans Christian Andersen (the famous Danish writer of fairy tales), and second to the effects of a general expansion of public-funded activities.

The results indicate that the calculated effects of Brandts Klaedefabrik on economic conditions scarcely differ from the effects of a general expansion of the public sector distributed evenly among all public expenditure outlets. However, more significant economic consequences can be calculated for Hans Christian Andersen's House than Brandts Klaedefabrik using a demand-oriented approach as a starting point. The fact that the two institutions have such different economic effects from this perspective is mainly due to Hans Christian Andersen's House being a cultural establishment which attracts a very considerable number of tourists, while Brandts Klaedefabrik caters more to the interests of local inhabitants; in a demand orientated approach, a public activity will emerge with better results if there is actual production of goods or services for the export market (including tourism), as is very much the case for Hans Christian Andersen's House.

An economic impact study of the Wexford Festival in Ireland [O'Hagan (1989, 1992)] also shows the importance of attracting tourists for the analysis to come up with big multiplier effects. At the Wexford Festival (an opera festival in Ireland) attendees from outside the Wexford area accounted for 85 percent of all opera ticket sales during the study period. Attendees from outside the Republic of Ireland accounted for almost 33 percent of all ticket sales and 50 percent of total expenditures. The vast majority (96 percent) of those traveling to Wexford during the time of the Festival came primarily to attend the Festival. On this basis O'Hagan concludes that the Wexford Festival makes a significant contribution to Irish tourism by attracting a sizeable flow of foreign tourists to Wexford and Ireland, by attracting an equally large flow of domestic (Irish) tourists to the economically depressed Wexford area, and by generating these tourists flows in an off-peak period of the year.

A further illustration of the impact of a single event is provided by Gazel and Schwer (1997) who studied the economic impact of three "Grateful Dead" performances in Las Vegas in 1995. They estimated the economic impact for the local economy to range between US\$17 and 28 million and between 346 and 589 sustained jobs.

A hybrid phenomenon is that of the "cultural mega-event" such as the European Capitals of Culture program or the World Expositions. These are events of limited duration and consequently part of their economic impact is limited to the duration of the event, as for festivals; however, a substantial second part of the expenditure has longer-lasting effects as it increases the capital stock of the city or region where the event takes place.

The extent of these extra returns is difficult to measure. Consider, for example, the European Cultural Capital Cities program. Palmer (2004) estimates that out of the 2.133 billion euros spent on the 21 cities in this program during the period 1995–2004, 737 million were operating expenditures and 1.396 billion were capital expenditures that went into new provision and upgrading of cultural capital, urban revitalization, and also into infrastructure such as improving transportation facilities. In the latter case, it is difficult to measure to what extent these infrastructure projects were *additional* in the sense that they would not have been undertaken in the future if the cultural event had not taken place. Furthermore, it is difficult to assess what other investments these projects may have crowded out and what the effects of these alternative investments would have been.

In a very thorough analysis of the EXPO 2000 (the World Exposition) in Hanover, Brandt et al. (2001) included only those projects that had not been planned before and moved up due to the EXPO. They calculated a primary and total impact of the EXPO and showed how the geographical delineation influences the results. The primary impact of the EXPO amounted to 7.5 billion DM in Lower Saxony, and 11.2 billion DM in Germany, the total effect summed to 8.4 billion DM for Lower Saxony and 13.4 billion DM for Germany. The employment effect was 65,000 and 100,000 man-years, respectively. The long-term effects – a better transportation infrastructure, better infrastructure for the Hanover fair, the newly created industrial zone at the former EXPO area, image and learning effects, etc. – were very hard to quantify.

### 2.1.3. Applications 2: The impact of the arts on a town, a municipality or a region

We consider two studies dealing with the entire cultural life of a town, a municipality or a region: a study of the New York–New Jersey Metropolitan Area [Scanlon and Longley (1984)] and a study of the economic impact of the arts in Glasgow [Myerscough (1988b)]. Other such studies include van Puffelen (1987) on art and culture in Amsterdam.

Scanlon and Longley (1984) gathered data in the New York–New Jersey Metropolitan Area from 1580 non-profit cultural institutions including museums, orchestras, dance and opera companies and theaters, 54 commercial Broadway and Off-Broadway theaters, 335 galleries and auction houses and film and television producers. Furthermore, 5600 questionnaires were handed out to visitors to performing and visual art events. On the basis of these data the economic impact of culture on the metropolitan region was estimated using an input-output model. They found that the arts had an impact of \$5.6 billion on the regional economy and generated more than \$2 billion in personal income and over 117,000 jobs for the area in the study period. Of the \$5.6 billion regional impact, \$1.3 billion was created by the 1580 non-profit institutions; \$360 million by art galleries and auction houses; \$650 million by Broadway and Off-Broadway theaters, plus Broadway road companies; and \$2 billion by local film and television productions. Moreover, the arts constituted a major export industry for the region by generating \$1.6 billion in expenditure by visitors who came primarily or extended their stay for arts and

culture and from touring companies. The more than 1900 arts institutions in the region inspired and entertained an annual audience of 64 million, 13 million of whom were visitors from outside the region. The regional arts institutions and the proportion of their non-resident patrons who visited the area primarily for the arts generated a total of \$150 million in regional income taxes and sales taxes revenues. Industries that benefited most in order of importance were: real estate; business and professional services; wholesale and retail trade; eating and drinking establishments; hotels and personal services; utilities; transportation; medical and educational services; and finance and insurance. Lastly, the arts at that time were shown to be a larger industry than advertising, hotel and motel operations, management consulting, or computer and data processing services. These results seem to show that cultural life is of great importance to the New York economy. No doubt this is true, but it is also worth remembering that in many ways New York is the cultural center of the USA, and thus not representative. Ten years later a second study was carried out on the same region, which showed that the total impact was 75 percent greater than in 1982, but its structure was almost unchanged (The Port Authority of New York and New Jersey, 1993).<sup>12</sup>

The economic contribution of culture to the Glasgow region<sup>13</sup> was assessed by Myerscough (1988b) for the year 1985–1986. He found that cultural life in Glasgow had a gross turnover of £204 million in that year, a value-added of £79.5 million and employed 8000 persons or 1.2 percent of total employment in the Glasgow region. Of the 6.5 million visitors to cultural events, 2.1 million went to cinemas, pop, rock, folk music and jazz concerts, 3.2 million visited museums and galleries and 1.2 million went to theaters and classical concerts; of the 4.4 million who went to museums, galleries, theaters or concerts, 64 percent lived in the Glasgow region, 17 percent were one-day visitors and 19 percent were tourists. Among the one-day visitors to museums and galleries, 66 percent stated that culture was the most important reason for their visit to the region, while only 16 percent of the tourists stated that culture was the most important reason for their stay. The total employment effect of cultural life in Glasgow was 14,731 direct and indirect jobs, amounting to 2.25 percent of the total number of people employed in the Glasgow region. Note that care must be taken in interpreting the figures from Myerscough's study to distinguish between gross and net effects (see Section 2.1.5).

#### 2.1.4. *The role of tourism*

As the above studies show, cultural tourism is decisive for the economic impact that a cultural institution or cultural life of a region has on the local or regional economy. However it is very difficult to define what should be included in cultural tourism. How many tourists can really be regarded as arts tourists? The problem is to distinguish those

<sup>12</sup> See also Heilbrun and Gray (2001).

<sup>13</sup> The Glasgow region was subject of one of three regional analyses of the economic importance of culture, which formed the basis of Myerscough's national analysis of the importance of culture to the British economy [Myerscough (1988a)]; the two others were carried out in Merseyside and Ipswich.



visitors who are in the region because of the cultural activities offered there from those who would have gone there anyway; furthermore, for those who came for multiple reasons it is necessary to determine to what extent they have prolonged their stay in order to enjoy the cultural amenities. Even though people say that the main reason for their visit is art and culture, it cannot be concluded that they would not have visited the place anyway. Conversely, in the case of visitors whose main aim is not art and culture, it cannot be excluded that the cultural activities on offer have not had some influence on their decision. These difficulties suggest the exercise of caution in interpreting the results of tourism-related studies. For example, lack of an exact distinction between “solely-culturally-motivated” tourists and those who are “combined-motivated” may bias the results. Another possible source of exaggeration lies in the overestimate of the complementary expenses incurred by the consumption of the cultural goods themselves such as transport and lodging expenses [Bonet (2003)].

The European Association for Tourism and Leisure Education (ATLAS) conducted a survey in 1997 with a sample of 8000 visitors to 20 different cultural locations in Europe. More than 50 percent claimed to have toured a museum during their trip, 40 percent had visited a monument, 30 percent had seen an exhibition, and 23 percent had seen a live performance. Between 20 and 30 percent of those surveyed could be classified as cultural tourists in the sense that they chose a destination with the primary objective of enjoying a cultural activity or product, and they identified themselves as cultural tourists [Bonet (2003)]. This study shows clearly that tourists visiting a cultural attraction should not be confused with cultural tourists. The biggest impacts of cultural tourism are presumably at regional or local level, since a larger percentage of one-day visitors than of tourists typically mention the arts as the main reason for their visit.

Another important question from an urban and regional development perspective is: which cultural assets are most likely to attract tourists and generate development and under which circumstances? An isolated attraction within a largely unknown or remote region has perhaps little hope of standing out in a dense market of tourist products and services. On the other hand, large-scale internationally famous cultural artifacts such as the Eiffel Tower in Paris, the Sydney Opera House, and the Statue of Liberty in New York operate as central tourist attractions, becoming symbols of their respective cities. There is, however, no statistical estimate dealing with the impact of a single large-scale cultural artifact and its contribution to tourism [Landry and Bianchini (1995)]. Obviously it is difficult to disentangle the attractiveness of a single cultural institution if it is embedded in a rich environment of culture and other amenities.

Bilbao in Spain may serve as a test case for the impact of a single internationally famous facility because of the Guggenheim Museum located there. Bilbao was not previously known for its tourism potential, and did not otherwise lend itself to large flows of tourism [Gomez (1998)]. Plaza (1999, 2000) has studied the influence of the Guggenheim Museum on tourism in Bilbao; although these studies show that the Guggenheim



Museum Bilbao exerts an important effect on the attraction of tourism, they do not calculate the economic impacts.<sup>14</sup>

Finally, tourism also can have negative effects on the local community, such as crowding effects, pollution and wear and tear. The latter can especially be the case in connection with some special items or places of cultural heritage, which are not able to stand intensive use.<sup>15</sup>

### 2.1.5. *Assessment and critique*

What kind of generalizations can be drawn from the different results of economic impact studies on the arts? First, these impact analyses show that the arts, like any other economic activity, have economic effects on the rest of the economy. If the arts create growth in aggregate demand, income will rise by more than the initial rise in demand, because of multiplier effects. The size of the multiplier depends on factors such as tax rates, the marginal propensity to consume, the marginal propensity to import, the area's size, etc. The results do not, however, indicate that the multiplier effect is greater in the arts sector than in other sectors of the economy, although the economic effects on the regional economy could indeed be greater if art activities attract a lot of tourists.

Second, the economic effects are greater for some cultural activities than for others. In general, theaters and concerts have a larger turnover than museums and galleries, but museums and galleries attract more day visitors than theaters and concerts. If we take into account that tourists spend more money per visit than one-day visitors who in turn spend more than the local population, and the fact that visitors at museums and galleries spend more money on shopping and on food and beverage consumption during their visit, we arrive at the conclusion that the economic effects are in general bigger for museums and galleries than for theaters and orchestras.

Although economic impact studies are very well established by now, many authors have criticized them on the interpretation of their results.<sup>16</sup> Recently, Seaman (2003) has summarized the sources of errors linked to impact studies as follows:

- direct base error (errors of counting the direct effects);
- induced base error (errors of counting the ancillary spending, or the induced effects);
- multiplier error (errors of calculating the multiplier effects);
- supply constraint error (errors arising if the assumptions about the capacity of the local area are not correct);
- ex-post verification error (the failure to verify whether any observed changes such as closures of local arts organizations had economic effects consistent with prior economic impact claims); and

<sup>14</sup> Note that the Guggenheim Museum itself carries out economic impact studies of their activity, which are published as part of their annual accounts.

<sup>15</sup> For a consideration of heritage conservation and tourism in urban regeneration, see Stabler (1996).

<sup>16</sup> For example, Bille Hansen (1995), van Puffelen (1996), Graves (1987), Peacock (1991) and Seaman (1987, 2001, 2003).

- policy interpretation error (the drawing of wrong policy conclusions from results of impact studies, as discussed further below).

As a consequence of the heavy critique of economic impact studies some authors have even suggested that no more impact studies of the arts were needed [e.g., van Puffelen (1996)].<sup>17</sup> Especially it is not clear that a positive economic impact of cultural events or institutions can be used to justify public subsidies to the arts. Many economic impact studies have been carried out to provide extra arguments for government subsidies to the arts. These studies are, however, not very well suited for this purpose. Most importantly, they risk evaluating the arts on an incorrect basis, because the main purpose of the arts is not to attract tourists or to generate consumption, income and jobs. These economic impacts are extra gains, not the main goal. The most important impact of the arts – and the main argument for their public support – is in their cultural or social influence. If one only analyzes the impact of the arts in relation to the economic side effects, a wrong picture will emerge. In particular this could imply a risk of focusing on cultural events with the greatest short term economic impacts – for example, big concerts with star performers – with the aim of attracting a big (foreign) audience, instead of focusing on cultural activities that are to the benefit of the population and to its cultural development.<sup>18</sup>

The problem is that impact studies use traditional indicators of economic growth such as national income, consumption expenditures, employment, etc., thus reducing the definition of development to one of economic growth in the short term; as noted earlier, this is a narrow definition of development compared to one containing not only economic but also human and cultural (sustainable) development.<sup>19</sup> Moreover the period studied is of great importance; the demand effects calculated in economic impact studies occur at once and they can be estimated quite exactly, whereas the economic effects which are due to the substantial importance of cultural activities for society and human development only occur in the long term and are very difficult to estimate.

Nevertheless the present knowledge which the existing economic impact studies provide can probably be useful in a broader assessment of policies and strategies for development where other kinds of development (i.e. human and cultural development) are taken into account. Throsby (1997, 2001, 2003) has made some initial suggestions on how the concept of “culturally sustainable development” can be made operational. He proposes “the specification of the concept as a set of principles or criteria which

<sup>17</sup> Seaman (2002) has however argued, that what he calls “naïve” economic impact studies, may provide more accurate estimates of the “true” total economic impact (more broadly defined) than more defensible methodological approaches, like “sophisticated” impact studies or contingent valuation studies.

<sup>18</sup> Schuster (1995) analyzes the motivations for two large urban festivals, La Mercè in Barcelona and First Night in Boston, and identifies motives to develop urban infrastructure, to upgrade the image of the city, to attract tourism, and to address social problems by bringing different types of people together (p. 184). Obviously, creating enjoyment for the population is the dominant motivation. Only part of these motivations is captured by economic impact analyses, thus making them a very narrow concept.

<sup>19</sup> See Throsby (2001, 2003) and the discussion in Section 1.2 above.

provide a means of systematizing the assessment of policies and strategies for development and for evaluating the extent to which, in terms of each criterion, such policies or strategies may or may not be consistent with culturally sustainable development". The first step is simply listing a set of variables: economic, social and cultural indicators. Having determined the appropriate set of indicators, the next step is to identify the range of relationships that might be postulated between them. In this process the knowledge from the existing economic impact studies can be of great help in identifying and estimating the relationships between some of the cultural and economic indicators. In other words even though economic impact studies are very narrow – looking only at economic growth in the short term, not taking into account the long-term effects and the public-good characteristics of the arts – properly conducted economic impact studies can be important tools in identifying the linkages between the cultural sector and the wider economy and in estimating the magnitude of the relationships involved.

## 2.2. Long-run effects

### 2.2.1. Location choice

In addition to the short-run effects that can be measured in economic impact studies, cultural activities may have more indirect and longer-term effects. In particular a lively cultural atmosphere may attract people and firms to particular regions, for example, because people enjoy arts and culture (as they enjoy a beautiful environment) and are willing to forgo some of their income compared to other locations in exchange for better access to cultural institutions. Such a positive externality will make this location more attractive to firms which can save on their payroll costs.<sup>20</sup> Moreover they may find more creative people in these areas which enhances their possibilities to hire qualified workforce.

Thus culture can have an impact on urban and regional development by attracting people, companies and investments. For most individual cultural institutions these effects will be without importance, because they are too small to trigger such a process of development. But these effects can be important for some few very large institutions such as the Guggenheim Museum in Bilbao or, more importantly, for a vibrant cultural life created by many cultural institutions in cities like London, Berlin, and Paris. Moreover, other kinds of investments may be attracted by a high level of cultural activity, such as investments in better transportation infrastructure, etc., but these effects are in general difficult to verify.

In regard to the locational choices of *people*, we could observe the arts and culture leading to population growth in a geographical area if cultural institutions and events attracted new citizens to live there. The arts can be one factor among others affecting where a household decides to live. In fact studies based on interviews generally show

<sup>20</sup> That is the entry point for the hedonic approach, cf. Clark and Kahn (1988).

that the arts are not the first factor considered when employees are deciding where to live and work; rather, it is job opportunities, housing, family relations, schools, transport facilities, and sports and leisure facilities that are the deciding factors.<sup>21</sup>

In regard to the locational choices of *firms*, it is widely believed that there is a positive relationship between the arts environment and industrial development in a region. But what do we really know about this relationship? Arts may have some influence on the location of industrial firms because the cultural environment may be important for a company's possibilities for attracting key manpower, in addition to such factors as infrastructure, land prices, wage level, manpower, sales opportunities, etc.<sup>22</sup> The question is, however, how important the arts and culture are as a localizing factor? Some attempts have been made to determine this by means of interviews with company managers and key manpower.<sup>23</sup> The results of these studies show that the arts are not at all the first factor considered when companies are deciding where to locate; companies regard hard business factors – sales opportunities, infrastructure, manpower, wage level, land prices and tax level – as far more important than cultural factors. But other studies have found stronger relationships. For example, Kotkin (2000) identified the relationships between lifestyle amenities and the locational preferences of some high-technology industries for neighborhoods such as New York's Silicon Alley, San Francisco's SOMA and Mission Districts, and Seattle's Pioneer Square. Sommers and Carlson (2000) found that some 50 percent of high-technology firms and employment in Seattle is located in a high-amenity district surrounding the urban core.

Florida (2002a) has introduced a new perspective on economic growth. His theory is that companies move where the creative people are, because human creativity today and in the future is the most important resource for companies. Creative people shift jobs very fast, and it is important for companies to locate in a place which have a "critical mass" of creative people. Florida (2002a) has stressed, that the "creative class" now comprises more than 30 percent of the entire workforce.<sup>24</sup> He emphasizes the "three T's": tolerance, talent and technology, which in his view drive economic growth. Technology and talent (a concept closely related to human capital) are very well established determinants of economic growth. The new concept is tolerance. Tolerance is important because diversity and an open society attract all kinds of creative people who cannot be discriminated between on the basis of place of birth, race, sexual preference, dress, age, etc. As a measure of how open an urban area or a region is, Florida has constructed several indexes such as the gay index, the bohemian index and the melting-pot index, measuring the share of the region's population who are gays, bohemians or foreign-born people respectively [Florida (2002b)]. A second important ingredient is lifestyle

<sup>21</sup> See, for example, Rubernowits and Rubernowits (1990).

<sup>22</sup> For an extensive review of the empirical literature on the location of firms, see Bartik (1991).

<sup>23</sup> See, for example, Myerscough (1988a).

<sup>24</sup> This includes computer and mathematical occupations, architecture and engineering occupations, life, physical, and social science occupations, education, training, and library occupations, arts, design, entertainment, sports, and media occupations, management occupations, business and financial operations occupations, legal occupations, healthcare practitioners and technical occupations, high-end sales and sales management.

amenities that people care about such as sport facilities, parks and other environmental amenities and – notably – arts and culture!

The presence of such creative human capital in turn attracts and generates innovative, technology-based industries. To test this hypothesis Florida (2002b) uses different correlations, e.g., between geographic concentrations of bohemians (measured by an index) and concentrations of high human capital individuals, and between the bohemian index and high-technology industry. Florida concludes that the findings support the hypothesis: “The presence of a significant bohemian concentration in a region signals an environment that is open and attractive to high human capital individuals. This in turn stimulates the kind of creativity and innovation associated with high-technology industries.” There are, however, problems with causality in this conclusion. Even if the positive correlations are the result of some type of causality, neither the direction nor the magnitude of the effects can be determined. It could well be the case that some latent third factor attracts both bohemians *and* creative firms but that there is no causality between the latter two (endogeneity problem).<sup>25</sup>

Ottaviano and Peri (2005) study whether cultural diversity measured by the variety of native languages spoken enhances productivity. Using census data for 1970–1990 they find that wages and employment density of US-born workers are systematically higher in cities with richer linguistic diversity.

To conclude, the interaction between the arts and industrial development is complicated and our present knowledge in this field is relatively sparse. We need more knowledge about which kind of industries are particularly interested in the level of art and culture in a region, and about the magnitudes of the impact and the complex interaction with other factors of importance. Apart from this more general effect of culture on the location of investments, companies and people, a special effect can be mentioned, namely that a high level of cultural institutions in an area could especially attract small companies working within the cultural or creative area, because they can use the network and cluster advantages arising from highly developed cultural institutions in a local area. We deal with this issue in Section 4 below. Finally, it may be noted that while the issue of attracting businesses and people is interesting from the viewpoint of an individual city or region, from a general point of view it may be a zero-sum game if the growth in one place happens at the expense of another place. This raises the question of how many (public) resources should be used in the presence of interregional competition in order to attract tourists, inhabitants and businesses.

### 2.2.2. *Educational and creativity values*

Different kinds of learning effects can be connected to arts and culture. Most people would agree that the arts are important for the formation of identity, personality, attitudes, job motivation, creativity, etc. – factors that are also important for economic

<sup>25</sup> It could also be the case that bohemians and foreign-born people follow creative and fast growing firms rather than the opposite.

development. In other words, the arts can provide inspiration for creative processes and trigger actions directed towards change and innovation. Artistic and cultural activity in a region can thus indirectly lead to more innovative industrial development through the creativity in the arts sector spreading to other economic activities. Creativity – in the form of innovation and breaking through boundaries – is something universal, and impulses from the arts can be important for actors on the business stage – and thus for economic growth. In this way, the arts may act as a dynamo in the development of society. Educational and creativity effects can only occur in the long run and are obviously very difficult to verify or to measure, in part because we do not know the transmission channels through which such creativity and personality spill-overs could materialize. It is thus necessarily a somewhat vague concept.<sup>26</sup> Yet again it must be remembered that, as in the case of economic impact studies, such productivity spill-overs are not the prime reason for supporting local art and culture.

A further arts-related development concerns so-called creative alliances. During the last few years, new forms of co-operation between culture and business have been developed which are broader than traditional sponsorship agreements, and, for instance, include design and product development, marketing and organization development. In some countries organizations have been set up with government support that try to organize this link between the arts and industrial innovation and development. The idea is to build networks between artists and industrial companies and induce companies to use artists in their innovative processes. In England the organization Arts and Business has existed for several years with this purpose, and similar institutions have gradually been established in many other western countries. The outcome of establishing these creative alliances is, however, an area where not much research has been done.

### 2.3. Cultural and urban planning

Over the last 10–15 years – primarily in Europe, the USA and Australia – a new trend has arisen using concepts like “cultural planning”, “cultural programming” and “urban planning” with the aim of revitalizing cities (especially inner cities) as well as old industrial areas and waterfronts.<sup>27</sup> The goal is to foster the development of “creative cities” [Landry (2000; Bianchini and Landry (1995)], to bring about “cultural regeneration” and to attract “the creative class” [Florida (2002a)]. Different development strategies can be applied which often define culture very broadly and which combine the cultural dimension with the extension of transport infrastructure, modern housing and industrial buildings, shopping centers, etc. Prime examples of cities that used such concepts in Europe include Glasgow, Newcastle, Barcelona and Bilbao.

<sup>26</sup> Many of these long-run effects are by their very nature somewhat vague and hard to measure. That does not mean that they may not be important, but it implies that they will not easily serve as a justification for government support of arts and culture; however, growth promoting effects should not be the prime reason for public support of arts and culture.

<sup>27</sup> See, for example, Mercer (2002) and Evans (2001).

These development strategies are interesting because they focus on the interaction between culture and many other factors of importance for the development of cities and regions in a broader planning process, where the art and culture are integral parts of the revitalization and development of cities and regions.<sup>28</sup> However, the question is what we really know about the result of such efforts and strategies, including the importance of culture in this connection. In the special issue of the *International Journal of Cultural Policy* on the theme “Urban Space and the Uses of Culture”, Gibson and Stevenson (2004) write in the introduction: “What is of most concern is the lack of rigorous research done on the cultural, economic, political and social short- and long-term effects of cultural planning. What evidence is there that the massive public expenditure required for these redevelopment and re-imaging strategies actually produces outcomes that are in the public interest?”, and they finish by concluding: “As this issue makes clear, the complexities of the cultural, economic, social and political implications of such programs are as multiple as their applications. It is clearly not simply a matter of ‘add culture and stir’.”

Glasgow and Bilbao are two of the most ambitious and outstanding cities in Europe that have deliberately used arts and culture in urban development. Both cities have experienced processes of deindustrialization and a subsequent loss in employment. They share a common past and a very similar approach to improving their current position as declining industrial centers on the European periphery. In terms of economic performance, the reality of Glasgow does not look as promising as it has been portrayed. Glasgow’s image may have changed, but the economic recovery has failed to materialize. According to Gomez (1998) employment had actually fallen. But it is undeniable that the image of Glasgow, both within and outside the city, has been radically reconstructed. There is also agreement on the success of these policies in terms of the physical renaissance of Glasgow’s central area. Overall, as Throsby (2001, p. 125) points out, “there may also be long-run externalities with real economic potential if the enhancement of the cultural environment of a city leads to greater social cohesion, a stronger sense of civic pride, lower crime rates, increased economic dynamism and so on. These sorts of factors may be important in improving the profile or image of the city as a desirable location for in-migration of capital and establishment of new business.” Nevertheless as noted above, these effects are very difficult to verify and trace out.<sup>29</sup>

<sup>28</sup> Stern (1999) found that even after statistically controlling for other determinants, arts and cultural organizations had an important independent effect on neighbourhood revitalization: “Compared to neighbourhoods with the fewest arts organizations . . . neighbourhoods with many arts and cultural organizations . . . were more than twice as likely to revitalize.”

<sup>29</sup> See also Vidarte (2000).

### 3. How does regional economic growth influence the development of the cultural sector?

Much of the debate on culture and development has centered on the issue we have dealt with above, namely the contribution that the arts sector can make to overall development of a region. The opposite causality however is of equal importance: How does regional culture change in the course of regional economic development? Or to put it more pointedly: Do economically poorer regions also have poorer culture? While the answer to this question is extremely difficult as it inherently implies value judgments on various forms of art (for instance, is a performance in a simple theater at a local inn better or worse than a performance in a large multipurpose auditorium?), we may be able to identify major lines of development of local culture as economic development proceeds. This refers not only to changes in the level of cultural activity measured by overall expenditures, but also to its structure.<sup>30</sup>

#### 3.1. Market prices and non-market values

If we assume for a moment that prices for cultural activities are determined freely in the marketplace, we could simply ask how demand and supply conditions will change in the course of development; i.e. identify the determinants of the demand and supply curves and analyze how they are affected by overall development. That would give us a clear idea as to how cultural industries will change in the course of development. Free price determination may be a reasonable assumption for a number of goods such as movies, popular and folk music, although markets may not all be competitive.<sup>31</sup> We will thus look at demand and supply conditions and their likely change brought about by development. However, if we look at cultural services provided by museums, theaters, or cultural heritage sites, such an analysis, valuable as it may be, will disregard the strong role that the state has in these areas. The heavy state involvement in arts and culture arises partly for historical reasons,<sup>32</sup> but partly also because it is a response to the existence of non-market values generated by the arts and culture, i.e. these activities

<sup>30</sup> Different cultural activities can be made comparable through the price system in order to determine the total level of cultural activity. This implies value judgments as well because prices reflect marginal utility, although these value judgments are made by the consumers and not the researcher. Note however that these expenditures reflect only the utility of people actually *consuming* the cultural services and not the option or bequest value offered by cultural institutions (see below); moreover it presupposes that prices are determined freely on competitive markets, which for cultural institutions often is not the case, given the strong involvement of the state.

<sup>31</sup> Note that measures of industry concentration may not be good indicators for the degree of competitiveness in industries producing information goods, such as the movie industry, due to the extremely high risk they face; cf. de Vany (Chapter 19 in this volume).

<sup>32</sup> For instance, the reason that Germany has so many orchestras is that it had been a nation fragmented into many kingdoms and princedoms each of which had its own court orchestra for the pleasure of the king and his court, cf. Schulze and Rose (1998).



impart values other than *use values* that may make people treasure art and culture even if they do not use these services directly. These values are: option, existence, prestige, and bequest values:

- *Option value* describes the utility that individuals receive from preserving the possibility that they may be able to enjoy the particular art or cultural service (even if they do not actually do so or have never done so before);
- *Existence value* is the value individuals derive from knowing that a particular cultural service exists (for instance, a particular museum, monument, etc.) even if they never intend to visit the site;<sup>33</sup>
- *Prestige value* refers to the value that a populace derives from national or regional heritage sites or monuments which help to identify and maintain national or regional identity or pride;
- *Bequest value* refers to the value that individuals derive from the knowledge that their children will be able to enjoy that particular service.

To the extent that culture creates these values in addition to the use or consumption value, it has public-good characteristics – non-rivalry and non-excludability. Of course, some important cultural and heritage sites as well as art in public spaces have public-good characteristics deriving also from their use values, even if some of their usages may be restrictable. For instance, access to the Eiffel Tower itself is restricted, but it is seen throughout the city of Paris – in its function as a national symbol and a monument of art and technology in the city of Paris it is a pure public good. Similarly a monument or statue in a public place is a pure local public good (or sometimes bad). Furthermore, art and culture may arguably have educational values that are not reflected in the price system. In all these ways culture produces positive externalities, constituting an obvious case for government intervention; indeed governments have played a very substantial role by subsidizing cultural institutions directly, or indirectly through tax concessions, or through public ownership.<sup>34</sup>

The strong government involvement in arts and culture means in turn that prices do not reflect solely consumer preferences for actual consumption of the cultural services and industries' supply conditions, but that they are distorted by subsidies, tax concessions, state-ownership and administered prices, etc. Consequently the value of the services produced depends crucially on what the political process is that determines the form and intensity of state involvement.<sup>35</sup> It is therefore important to understand this process and how its results will change in the course of development. The political interventions in arts and culture do not replace market forces, but rather they alter them. To the extent that cultural policy is demand-determined, i.e. politicians actually

<sup>33</sup> Sub-aqueous archeology is a case in point as many people do not dive; cf. Whitehead and Finney (2003).

<sup>34</sup> See, inter alia, Baumol and Bowen (1966, Chapter 16), O'Hagan (1998), Frey (2003), and Chapter 32 by Peacock, Chapter 34 by van der Ploeg, Chapter 35 by Netzer and Chapter 36 by Schuster in this volume.

<sup>35</sup> Of course private sponsorship and philanthropy could address the problem of underprovision of arts and cultural goods through private markets that stem from the public-good characteristics of arts and culture (positive non-use values); see Chapter 37 by Katz in this volume.

do what their constituencies want them to do, the factors that determine demand for culture are actually reinforced. Demand-determined policies refer not only the factors that determine effective demand in the market place, but also to the non-use values that individuals attribute to art and culture. We take up these issues in the next section.

### 3.2. *The demand for cultural goods*

#### 3.2.1. *Observable demand behavior*<sup>36</sup>

The empirical literature on the demand for arts has become quite sizeable, including econometric studies and a number of visitor surveys.<sup>37</sup> Given Seaman's comprehensive survey in this volume,<sup>38</sup> we simply present here our reading of the empirical evidence which refers mostly to performing arts (theater, opera, ballet, classical music) and the cinema.

Highly educated people are strongly overrepresented in audiences of performing arts as are people with high income; however this effect is significantly more pronounced for education than for income.<sup>39</sup> Demand for arts depends negatively on prices, although the estimates for the own price elasticity vary from very inelastic to highly elastic. Studies vary greatly in degree of sophistication and aggregation. The studies that use aggregate data (for instance, total revenue divided by attendance as a measure for the relevant price) tend to have lower price elasticities. They are of lesser value because they are not able to portray adequately the behavior at the margin. There is no clear evidence however that price elasticities for disaggregated studies exceed unity in absolute terms. Nevertheless our reading is that the evidence overall tends to support price elastic behavior.<sup>40</sup> Demand for industrially produced cultural goods such as movies has been shown to be price elastic [Fernandez-Blanco and Banos-Pino (1997); Dewenter and Westermann (2005)].

While it is intuitive that art is a luxury good and visitor surveys confirm a higher representation of well-off people in performing arts attendance, the econometric literature has not been unambiguous. Felton (1992) finds evidence of the luxury character for ballet but not for the opera, whilst in a sophisticated study Pommerehne and Kirchgässner (1987) find income elasticities greater than 1 for both average and high income visitors of German theaters. Gapinski (1984) and Bille Hansen (1991) find likewise income

<sup>36</sup> We are grateful to Bruce Seaman for sharing his views with us on the determinants of demand for art; see further Chapter 14 in this volume.

<sup>37</sup> For instance, the US National Endowment for the Arts sponsored visitor surveys in 1982, 1985, 1992, 1997, and 2002; in Germany the "Kulturbarometer", a national visitors survey, is frequently carried out, with varying focus; in Denmark similar studies have been conducted [see Bille et al. (2005)].

<sup>38</sup> See also Levy-Garboua and Montmarquette (2003).

<sup>39</sup> See Baumol and Bowen (1966) for the US and UK; Ford Foundation (1974) for the US; and Wiesand (1995) for Germany, cf. also Gray (2003). Obviously, overrepresentation does not necessarily imply that highly educated or high income people constitute the majority of the audience.

<sup>40</sup> Only few studies report cross-price elasticities [e.g., Gapinski (1986); Krebs and Pommerehne (1995)].

elasticities larger than one for theaters. Even though the evidence is not unequivocal, we see the balance of the existing evidence to be in favor of the performing arts as a luxury good.

In terms of overall *exposure* to the arts, the young population is overrepresented but in terms of overall *attendance* for the performing arts the opposite applies, i.e. the old constitute a relative larger share of the overall audience in a given period and the young are underrepresented.<sup>41</sup> Furthermore, arts consumption decreases with distance from the arts venue [Gray (2003); Schulze and Ursprung (2000)], women attend the performing arts more often than men (except for Jazz and Dance) and in the US there is a racial pattern depending on the arts form [Gray (2003)]. Other important influences for the demand for arts are life-style variables such as political or sexual preferences, which are notoriously hard to capture.<sup>42</sup>

How do these results relate to the problem at hand, i.e. the effect of economic development on the demand for culture? Development will affect all the described parameters. Generally, economic development tends to make people become richer, and typically also better educated. Since education and income have been shown to increase the demand for culture we should expect richer regions and cities to have a richer culture, other thing being equal. Note however that it is not only per capita income that is indicative of the demand for culture, but also the industry structure of the regional economy and thus the socioeconomic structure of the population. An old industrial town may have a lower demand for culture compared with a town with equal per capita income that is home for service industries, universities, government administration, etc. Of course the distribution of income is important as well. If income is concentrated on a few people with high incomes, some elitist art institutions may exist but the overall level of cultural activities may be lower than for a region with more equal income distribution. Geography and infrastructure matter as well – a largely dispersed population will have a higher average travel time and travel costs, which will tend to reduce the demand for cultural activities. Lastly a city or a region tolerant vis-à-vis homosexuality (and possibly other ‘alternative’ life-styles) might attract those people who *ceteris paribus* tend to have a

<sup>41</sup> One reason for age, but particularly education is that preference for arts and culture is an acquired taste, where consumption capital is built up over time [Stigler and Becker (1977)], which is easier for more educated people. Consumption capital tends to be higher for older people as they acquisition period is larger, cf. Lévy-Garboua and Montmarquette (2003). Very significant impact on arts attendance has exposure to arts education such as music lessons, etc., see Smith (1998). Art exposure is measured as relative frequencies, i.e. the percentage of people who go to the arts at least once in a given time interval. The exposure is larger for the young than for the overall or the older population, attendance is not.

<sup>42</sup> German voters of conservative parties have been shown to prefer more traditional art forms (opera, ballet, symphony) whereas voters of the social democratic and the green party have a higher probability to be in favor of experimental art forms [Wiesand (1995); Schulze and Rose (1998)]. Homosexuals are *much* more likely to attend the arts than demographically similar heterosexuals [Lewis and Seaman (2004)].

higher demand for cultural services. Those areas will probably also tend to attract more artists, but this becomes a supply side factor.<sup>43</sup>

Note however that existing demand studies mostly refer to performing arts and to institutions such as ballet companies, theaters, opera houses and orchestras that are well established, have a permanent legal form, and sell their tickets on the market. As regions become richer, they might not only enjoy a 'richer' culture in terms of reported per capita spending on culture, they might also experience a substitution from 'subsistence' and non-market art and culture towards more established art forms that our statistics would not capture.<sup>44</sup> Inasmuch as such a substitution takes place, the increase in cultural services in the course of development as reported by demand studies would be overestimated.

### 3.2.2. Non-market demand

As noted earlier, individuals' valuation of cultural goods and services is only partially reflected in the market demand for culture, since cultural goods produce non-use values as well. If government support for the arts reflects these non-use values, we have to study how they are systematically related to factors that change in the course of development. An increase in national or regional income over time will typically give rise to increasing public expenditures on culture, such that regions with a larger income per inhabitant will have relatively larger cultural expenditure per inhabitant than regions with a smaller national income. In Bille, Hjorth-Andersen and Gregersen (2003) the public cultural expenses of the Nordic countries have been compared for a number of years. This analysis shows that:

... we can conclude that cultural expenses have an income elasticity of 1 or perhaps a little more. Or in other, less technical, terms: when the national economy is doing well, there is a spill-over effect on culture. Every time Danish society becomes DKK 10 billion richer, this will affect culture with a 0.5 to 0.6 percent increase or DKK 50 to 60 million. Thus, economic growth seems to be the right thing for a rich cultural life! Seen in this perspective, the Nordic countries seem remarkably similar. [Bille, Hjorth-Andersen and Gregersen (2003, p. 343)]

In general four approaches have been used to study the magnitude of the demand for non-market goods and the determinants of non-market values [Frey (1997)]: the hedonic approach, the travel cost approach, contingent valuation studies, and the analysis of referenda.<sup>45</sup> We consider each one in turn.

<sup>43</sup> "The experience economy" [Pine and Gilmore (1999)] actually seems to be a concept that builds on this growing demand for culture, events and experiences. This relatively new and highly canvassed concept includes the demand for culture, events and experiences in a broad sense.

<sup>44</sup> This point has been made convincingly by Goldstone (2003) for the international context referring to the statistics of the UN World Culture Report.

<sup>45</sup> Of course, a very promising fifth method is discrete choice experiments (DCE) that have been used extensively in environmental economics [Adamowicz et al. (1998); Adamowicz and Boxall (2001)] and also in

(i) *Hedonic approach*. The hedonic approach can be used to analyze how land and housing prices are influenced through proximity to cultural sites (or the beauty of the landscape) after controlling for other factors.<sup>46</sup> Alternatively, hedonic studies might analyze the wage discount that people are willing to accept in exchange for a workplace in an area with high supply of cultural services. These studies need to rely, however, on competitive housing or labor markets and must be able to control for all other variables in the wage or land price regression. All other unobservable factors must not be correlated with culture, which is hardly met in reality as locations with a vibrant culture mostly tend to have other attractive attributes as well, so that a classical identification problem arises. Within cultural economics very few hedonic studies have been undertaken.<sup>47</sup>

(ii) *Travel cost*. The travel cost approach recognizes that willingness to pay for cultural services is not measured only by market prices for tickets but also by costs for travel in terms of money and opportunity cost of travel time. The sum of these latter costs and the entrance fee constitute a lower bound for the willingness to pay for the cultural service under consideration. However, this approach needs to assume that the visit to the cultural site is the only purpose of the travel and that there is no pleasure derived from the travel itself. There is no reason to believe that these assumptions typically hold in reality.<sup>48</sup> Thus this approach, like hedonic pricing, has significant drawbacks; most importantly both are capable of measuring only use-values (not the non-use values). We thus focus our attention on the remaining two approaches.

(iii) *Contingent valuation*.<sup>49</sup> Contingent valuation studies are a survey-based methodology where a sample of a population is asked about their maximum willingness-to-pay

health economics [Ryan and Gerard (2003)]. Boxall et al. (1996) show that WTP is much higher in CVM than in DCE, Carlsson and Martinsson (2001) show that the WTP for a public good is almost the same for hypothetical and real payments. DCE seem a very promising avenue of approach also for non-market evaluation in cultural economics, but it has hardly been used so far. An exception is Boxall, Englin and Adamowicz (2002) for cultural heritage (aboriginal rock paintings).

<sup>46</sup> The hedonic approach regards a good as a bundle of underlying characteristics which are valued by consumers and therefore the price can be decomposed into valuations of these components. In our context proximity and access to cultural institutions is such a characteristic for housing prices or wage regressions. The hedonic approach was developed 40 years ago by Lancaster (1966); Rosen (1974) was the first to include elements of the quality of the environment in this "hedonic" decomposition, which ultimately leads to an indirect value for changes in the levels of such characteristics.

<sup>47</sup> See, for example, Clark and Kahn (1988).

<sup>48</sup> For instance, a day trip to La Scala in Milan with friends or family will be enjoyable in itself and might involve other attractions such as shopping, enjoying the urban flair, etc., with the opera visit being but a special highlight of such a trip; for a recent contribution on travel cost, including a review of the literature, see Poor and Smith (2004).

<sup>49</sup> Contingent valuation studies measure the benefits to the individual consumers, not the importance of arts and culture for urban and regional development. Since in this section we are concerned with how arts and culture will change in the course of development we need to study the determinants of demand for them including non-use demand. The contingent valuation method is a way to do that. We need to formulate hypotheses about how the determinants of demand are changing in the course of development and whether local governments will accommodate these demand changes in their support for arts and culture.

(WTP) for a specified hypothetical change in a (public) good or service. Values for the good are then inferred from the respondents' valuations. The method is called the contingent valuation method (CVM) because the responses depend on a hypothetical market which the interviewer describes to the consumer. Different elicitation formats are available, including the open-ended (respondents are asked directly for their maximum WTP or WTA),<sup>50</sup> dichotomous choice (where respondents are asked to answer yes or no to a specific amount stated by the interviewer) and choice experiments (where respondents are asked to choose between different choice sets, where attributes, including price, are varied). All elicitation formats have their pros and cons.<sup>51</sup>

CVM has been developed in environmental economics, but recently has to an increasing degree been used for valuation of cultural goods; most of the studies have been concerned with cultural heritage and historic sites.<sup>52</sup> Most studies show a quite high willingness-to-pay among both users and non-users of the cultural good, with the non-use value constituting the biggest part of the total value and with a higher average WTP among users than non-users. There is also an increasing WTP with income and education, a result consistent with theoretical expectations, and average WTP is decreasing with distance to the cultural good [e.g., Bille Hansen (1997, 2002)]. An example of a CVM study conducted at the urban or regional level is Alberini, Riganti and Longo (2003) who conducted a survey of Belfast residents to elicit people's preferences for regeneration projects that change the aesthetic and use character of specific urban sites. They found that individual choices could be explained by these attributes of the regeneration projects. Another example is a study eliciting the value of protecting and restoring the Nidaros Cathedral in Trondheim, Norway, which is the oldest medieval building in Scandinavia [Navrud and Strand (2002)]. In this study a contingent valuation survey of visitors to the cathedral was carried out in the summer of 1991. Santagata and Signorello (2000, 2002) conducted a CV study aimed at measuring holistically the total benefits accruing to the local residents from maintaining the Napoli Musei Aperi, a cultural public good provided in Naples. When dealing with cultural heritage or cultural goods of national importance, the study has to be conducted at the national level, to get an estimate of the total value of the good.<sup>53</sup>

<sup>50</sup> Willingness to pay (WTP) refers to a situation in which the respondent is asked how much she would be willing to pay for a certain good whereas willingness to accept (WTA) refers to a situation where the respondent is asked how much she must be given in compensation for the loss of a certain good. Typically WTA exceeds WTP.

<sup>51</sup> For an in-depth discussion of the theoretical and empirical issues involved, see Mitchell and Carson (1989), Branden and Kolstad (1991), Freeman (1993) and Bateman et al. (2002).

<sup>52</sup> See Navrud and Ready (2002) and Bille Hansen (1998). Some of the pioneering studies are Throsby and Withers (1983), Morrison and West (1986), Martin (1994) and Bille Hansen (1997, 2002). Noonan (2003) presents a meta-analysis of existing CVM studies of cultural goods, and states that 139 studies exist within cultural economic, among which 61 have been published since 2000. The *Journal of Cultural Economics* devoted a special issue to the critical appraisal of the use of CVM in cultural economics, including a number of interesting applications (vol. 27, issues 2–3).

<sup>53</sup> See, for example, Bille Hansen (1997) on the Royal Theatre in Copenhagen.

CVM studies are subject to many kinds of biases including hypothetical bias (leading to an overstatement of WTP because the scenario is only hypothetical), strategic bias (where respondents are acting strategically), “warm glow” bias<sup>54</sup> (where respondents are willing to pay for doing something good – not for the particular change specified in the scenario) and so on. CVM has been widely criticized,<sup>55</sup> and has created a vehement debate among economists, who traditionally prefer estimates that are derived from observed behavior rather than from stated preferences [Carson, Flores and Meade (2001)]. Nevertheless, in many cases CVM is the only method capable of estimating the total value (use and non-use value) of a good.

(iv) *Referenda*. Referenda have the distinct advantage that they tie individual evaluation to a concrete decision that voters know will be implemented, in contrast to CVM studies where subjects’ stated willingness to pay need not reflect their *actual* demand behavior. Furthermore, anonymous referenda are not subject to the same sorts of biases as affect CVM, in particular anonymous referenda do not suffer from a systematic answering bias which may occur in CVM if individuals respond strategically or attempt to present themselves in a favorable light vis-à-vis the interviewer.<sup>56</sup> The obvious disadvantage of referenda is that only few countries hold referenda about cultural issues.

Pommerehne (1982) analyzes two consecutive referenda in the semi-canton of Basle-City in 1973 and 1974 on possible increases in theater subsidies provided by the canton; the first proposal (involving a larger subsidy) was voted down, the second approved. Frey and Pommerehne (1989, Chapter 10) study a referendum held also in Basle-City in 1967 to purchase two Picasso paintings for the Basle Art Gallery; the bill was approved by 54 percent. Schulze and Ursprung (2000) analyze a 1994 referendum that transferred financial responsibility for the Zurich opera house from the municipality to the canton; it was approved by 73 percent.

The results of the three analyses are consistent in showing that income and education exert a significant positive influence on the percentage of yes votes in a district.<sup>57</sup> Taxable income and travel time or monetary travel costs significantly reduce the share of approving votes. The bequest motive, approximated by the number of children aged 0–15 or the birthrate, are significantly positive and so is the share of people generally

<sup>54</sup> See Diamond and Hausmann (1994).

<sup>55</sup> See, for example, Diamond and Hausmann (1994), Hanemann (1994) and Portney (1994).

<sup>56</sup> A further advantage of referenda is that they contain data for more individuals. There is an extensive literature on the pros and cons of referenda and contingent valuation studies: see, for example, Frey (1997), Schulze and Ursprung (2000, p. 134), Bishop and Heberlein (1986), Mitchell and Carson (1989), the special issue of the *Journal of Economic Perspectives* in Fall 1994, and the special issue on contingent valuation in the *Journal of Cultural Economics* 2003 (vol. 27, issues 3–4). We note that both approaches have their relative advantages; inasmuch as they result in a consistent picture of the determinants of demand for cultural goods they are complementary approaches.

<sup>57</sup> Note that referenda analyses take the voting districts as the unit of observation since individual voting is anonymous. Thus the endogenous variable is the log odds ratio of the share of yes votes over 1 minus this share; the exogenous variables used are educational profile (e.g., share of population with secondary education), average income in the district, etc.



interested in cultural or public affairs (as measured by the share of subscribers to non-tabloid newspapers).<sup>58</sup> Results on age and self-employed versus employed as proxies for opportunity costs of time do not produce consistently significant results of the same sign. Although variables associated with use and option values account for a large share of the variation in approval rates, these analyses show that non-use values play a significant role in explaining support for art.<sup>59</sup>

### 3.3. The supply of cultural goods

#### 3.3.1. Production process in the arts

How will supply conditions for arts and cultural goods change in the course of regional or urban development? Obviously, a unique answer is impossible because supply conditions differ fundamentally between art forms. Consider the following classification:<sup>60</sup>

- *Live performing arts* (music concerts of all styles, plays, operas, ballet and dance);
- *Visual arts* (paintings, drawings, sculptures);
- *Reproducible art* (literature, recordings, movies, digital art) with an industrial reproduction process.

This classification is primarily based on production process and not on the sensual experience or content. Visual arts such as sculptures, paintings, sketches, etc. are produced with no scale economies – there is no reduction in unit labor input as production increases.<sup>61</sup> Moreover, production technology prevents easy and perfect copying, which makes this art unique. Replications are distinguishable from (and thus not interchangeable with) the original.<sup>62</sup> There are hardly scale economies in the *live* performing arts either as a given play, concert, or opera has a fixed duration. Only time for rehearsal and preparation of the backdrop per performance can be decreased by increasing output. This art form is unique to the extent that a particular performance of a particular (group of) artist(s) is not interchangeable with a performance of the same opera, concert, etc. by different interpreter(s). Uniqueness is thus a matter of degree. The common element that characterizes the production of these art forms is “the handicraft attribute of their supply processes” [Baumol (1996)]. Production processes differ fundamentally for reproducible art. Whereas the creative process of writing a book, composing a concert and

<sup>58</sup> The latter variable may indicate a stronger existence or prestige value. There is no clear proxy for the option value alone because those people who use the cultural institutions more will presumably also attribute a higher option value to this institution. In other words variables such as income, education and distance will proxy use and option values simultaneously.

<sup>59</sup> Further contributions are Frey and Pommerehne (1995) for Switzerland, Getzner (2004) for Austria, and Rushton (2005) for the US.

<sup>60</sup> See further in Schulze (1999, Section 2).

<sup>61</sup> The absence of scale economies does not preclude learning curve effects or economies of scope.

<sup>62</sup> Of course that does not preclude fakes that may be very hard to distinguish for the non-expert; but if detected they would command a significantly smaller price than the original [Frank (2004)]; see further Chapter 8 by Benhamou and Ginsburgh in this volume.



performing it once, etc. does not exhibit scale economies, the reproduction of the final output – a book, a CD or DVD, a film roll, etc. – exhibits very strong scale economies, the marginal costs being almost negligible. That gives rise to serious copyright issues; the cultural industries, which produce these goods industrially with advanced technology, are typically copyright-protected [Towse (2003)].<sup>63</sup>

Obviously, economic development, brought about by technical progress and accumulation of capital, affects these fundamentally different types of production processes – individual production versus industrial production – very differently. We discuss these effects in turn, starting with the effects of technological development on live performing and visual arts.

### 3.3.2. *Baumol's cost disease*

Baumol's cost disease is an intuitive general equilibrium story of relative price effects caused by differences in productivity growth between sectors that are linked through an integrated labor market [Baumol and Bowen (1966)]. Consider one sector (e.g., the live performing arts) that experiences no change in labor productivity and another sector (manufacturing) that experiences large productivity gains. The productivity growth in the manufacturing sector leads wage rates in this sector to rise in terms of the manufacturing good, the numeraire, as labor unit costs have fallen due to the productivity gain.<sup>64</sup> If manufacturing goods constitute a significant share of overall consumption, real wages in that sector will have gone up as well. If (real) wages in the other – stagnant – sector are linked to (real) wages in the manufacturing sector, they need to follow suit.<sup>65</sup> This leads to an increase in unit costs in this sector because the wage increase cannot be offset by productivity gains. The rise in unit costs must be recouped by increases either in prices or in subsidies. The relative price of the stagnant sector's product has increased, either as market price or as shadow price. The logic of the argument, of course, still holds if there is productivity growth also in the arts sector; for the relative price effect to materialize only a (significant) productivity *differential* needs to exist.<sup>66</sup>

This is a general story that can be applied to the arts as well as to education, health care, etc. [Baumol (1996)]. For live performing and visual arts the potential of labor

<sup>63</sup> Note that artists can be part of both production processes: the author performing a reading is a live performing artist but at the same time his or her books are produced industrially; another example is a symphony orchestra, which gives live concerts and records compact disks.

<sup>64</sup> Obviously, this argument presumes wage flexibility.

<sup>65</sup> Wage rates need not be identical (as they would be if labor were perfectly mobile and homogeneous across sectors). It suffices that the relative wage differential remains constant, maybe out of a sense of 'appropriateness' or 'fairness', or because unions use wage trends in other sectors as a yardstick for their own demands. Even if the wage gap widened but the wages in the arts sector responded to the general wage increases, art would become relatively more expensive.

<sup>66</sup> See further Chapter 11 by Baumol and Chapter 15 by Brooks in this volume.

saving technical progress is relatively limited;<sup>67</sup> rehearsal times cannot be cut substantially without compromising quality, a Beethoven symphony requires a full symphony orchestra, and a Mozart string quartet lasts a fixed amount of time. Even though there have been modest productivity gains in the performing arts, they have never come close to those in many other sectors<sup>68</sup> such as the manufacturing and IT sectors. This has led to price increases in the arts that have continuously exceeded inflation rates.<sup>69</sup> Economic development thus brings about relative price effects, which tend to make those art forms more expensive that have the “handicraft attribute” [Baumol (1996)]. These relative price effects trigger a substitution away from these art forms according to the price elasticity of demand. At the same time development has made people richer and thus an income effect running counter to the substitution effect is simultaneously at work. The net effect for the art forms subject to the ‘cost disease’ depends on the relative strength of these two effects and could go either way, although for other art forms and cultural industries it is unambiguously positive.

The cost disease refers to the increase in relative prices of cultural goods brought about by the increase in *relative* production costs. This price increase however need not reflect the movement in *effective* prices of cultural goods for the consumer thanks to countervailing secondary supply side effects. If in the course of development transportation and information infrastructure improves, the opportunity costs of visiting cultural institutions will decline, i.e. it will be easier to get information about cultural events, book tickets and get to the site. Conversely, previously remote areas will become more accessible to touring theaters and orchestras, etc. thereby improving the supply of arts and culture in these areas. Lastly, regional development tends to increase the urban population and thereby puts a larger share of the population within reach of the cultural centers. We thus have two supply side effects that run in opposite directions.

### 3.3.3. Innovation, globalization and cultural change

Technical progress in cultural industries has led to a reduction in prices or an improvement in quality of industrially-produced cultural products. Music CDs, movies on video cassettes and DVD have declined in price as has the necessary equipment (CD and DVD players, etc.). At the same time TV stations have switched from terrestrial mode

<sup>67</sup> Note that this argument applies strictly not to the performing arts, but only to the *live* performing arts. Technical progress may apply to the reproduction process of this art form (CD pressing, etc.), even if the contribution of royalties to orchestras’ revenues has been limited [Heilbrun (2003)].

<sup>68</sup> The fact that performing arts have experienced productivity gains as well has been levelled as criticism against the ‘disease’. Indeed there are productivity gains in the performing arts, for instance, the number of actors has been declining [cf. Baumol and Baumol (1985); Heilbrun (2001)] and festivals have emerged as low cost alternative to more established art forms [Frey (1996)]. Such a critique against Baumol’s theory [e.g., Cowen (1996)], however, misunderstands the general equilibrium logic of Baumol’s argument – *relative* productivity growth matters for the relative price effect, not the absolute. For a discussion of the cost disease see also Towse (1997).

<sup>69</sup> See Baumol and Bowen (1966), Baumol (1996) and Heilbrun (2003) for ample empirical evidence.

to cable and satellite and are beginning to switch from analogue to digital transmission.<sup>70</sup> These technical improvements disseminate to consumers unequally as they are costly; more well-off people tend to have them earlier and they only become affordable to more people as scale economies lead to price decline. For those cultural products that are network-based such as cable TV and digital terrestrial TV, the regional base is important as a threshold needs to be surpassed before investments in these networks are economically viable. In the course of regional economic development local purchasing power for these products increases and with the rise in user base, the average costs decline. Therefore the introduction of new network-based technologies depends on local development. For instance, digital terrestrial TV has been introduced in Germany only in selected areas such as Berlin and the Ruhr area [Perino and Schulze (2005)]. That implies a larger availability not only of movies, but also of cultural programs as these programs are seen by a minority only and are thus offered only if enough channels are available; however with satellite transmission this effect is of lesser concern.

Technical change brings about substitution that affects the local cultural industries. With the spread of television local movie theaters have suffered from declining audiences.<sup>71</sup> Due to decreasing costs of broadcasting equipment and deregulation, local radio and television stations have mushroomed. The number of local stations is affected not only by available bandwidth but also by the user base as these stations must finance themselves through advertisement, the rates for which depend on the average number of contacts, i.e. the local user base. With regional development this user base increases and so does the supply of local radio and television programs.

In a related vein, technological developments have led to the spread of the Internet and made data transmission more efficient; fast Internet connections in turn have made (illegal) music downloads possible on a large scale and the advent of CD burners and MP3 players has made copying easy. Not only does this crowd out regular sales [Rob and Waldvogel (2004); Zentner (2003)], but also makes this music effectively cheaper. It tends to make artists more global than they have already been.<sup>72</sup> This relative price movement (and the greater accessibility) may shift consumption from local live performing arts towards recorded international music. Obviously the spread of computers with fast Internet connections rises with the level of development.

### 3.3.4. Trade

Globalization has brought an increased exchange in cultural goods. While the trade in works of art follows normal trading patterns in the sense that closer and larger economies trade more with each other, it deviates significantly from normal in that the

<sup>70</sup> For Germany see Perino and Schulze (2005).

<sup>71</sup> In Germany, for instance, cinema turnover declined sharply in the mid 1960s with the spread of TV sets (see *Filmstatistisches Jahrbuch*, 2003).

<sup>72</sup> In addition, it reduces the complementarities between live concerts and record sales, which has led to a surge in concert prices [Krueger (2005)].

positive effects of common language, geographical closeness and GDP per capita are significantly stronger for trade in works of art than for overall trade [Schulze (1999)]. This demonstrates that cultural proximity proxied by geographical proximity and common language exerts an important influence. Stated differently, foreign cultural products are valued with a discount leading to a preference for local or national art and culture; cultures that are 'closer' to one's own have a lower cultural discount than others. In the course of increased trade this cultural discount may become less and less as culture-specific consumption capital is built up, with the consequence that *local* culture may play a diminished role in local cultural life.<sup>73</sup> A clear demonstration of this effect is the changing composition of trade in movies in European countries. The market share of German movies in German theaters declined from around 47 percent in the late 1950s to 16 percent in 2001, while the American movies increased their share from 30 percent to 70–80 percent [Dewenter and Westermann (2005)]. The same may be true also for other forms of local, traditional culture such as local dances, literature, and music.

As noted above, attracting tourists is an important way in which the arts and culture can contribute to exports at the local and regional level. Similarly arts companies such as theaters and orchestras can tour abroad. In addition the arts and culture may be able to promote the sales of other goods in export markets because a good image (created by the arts offering) generates goodwill. However the few studies which exist on this subject indicate that a country's arts institutions have very little if any impact on the sale of goods to other countries. For example, Abele and Bauer (1984) found that no importance was attached to sympathy for the producing country in connection with the purchase of imported goods, where price and quality were deemed to be very important factors. The impact is probably more important at the national level than at the urban level, but the empirical evidence is very sparse.

### 3.4. Development and public support for the arts

An assessment of the impact of urban and regional development on the cultural sector needs to take account of how public support for the arts changes in the course of development. Government intervention in the markets for art and cultural products is very important in many countries; for instance, in continental Europe public support makes up significant portions of total revenue for theaters, orchestras, opera houses, and museums.<sup>74</sup> The level of public support differs widely between countries and between art and cultural sectors; moreover, forms of support and the relevant level(s) of government differ across countries. To illustrate, in 1994 the per capita levels of public support of the arts ranged from \$US112 for Finland and \$90 for Germany to \$9 and \$6 for Ireland and

<sup>73</sup> For the role of consumption capital cf. Stigler and Becker (1977), for cultural discount and international trade in arts see Schulze (2003), for the application to trade in movies see Wildman and Siwek (1988).

<sup>74</sup> See, for example, O'Hagan (1998, Chapter 6) and Schulze and Rose (1998).

the US, respectively.<sup>75</sup> The support has very different importance for the various cultural subsectors – for instance, in Germany public support for symphony orchestra and theaters (*‘Kulturorchester’*) covered up to 80 percent of total revenues of these institutions [Schulze and Rose (1998)], while most musicals are not supported at all. The level of government that provides the lion’s share of cultural expenditures differs as well: in Finland, France, and Germany, for example, more than half of all support comes from *local* governments (counties, municipalities, cities), whereas Ireland, Italy, Sweden, and the UK rely more than half on *national* support.<sup>76</sup> While the level of local government support reflects local conditions, especially the state of local development, central government support is more likely to reflect the nation state’s development.<sup>77</sup> Moreover, one may surmise that countries with a large share of central government support may allocate a larger share to the capital.<sup>78</sup>

Countries differ with respect to the instruments they use to support arts and culture; hence different decision-makers will be involved and the resulting allocations are also likely to differ. For example, while direct support depends on political decisions, tax concessions give wealthy individuals and corporations the power to decide where taxpayers’ money should go. Insofar as political decisions follow voters’ preferences, they should reinforce the determinants for demand for culture including the willingness to pay for the non-market valuation of the arts and culture. However it is very difficult to draw general conclusions if politicians enjoy discretionary scope; the change in public support for the arts in the course of development depends inter alia on the institutional framework, i.e. the level of government, the socio-economic characteristics of the electorate, the initial size of support and the available instruments for support.

### 3.5. Assessment: *The influence of development on culture*

For our purposes we have taken ‘development’ to have three manifestations – it brings about technological progress which is unevenly distributed, it increases per capita income and education levels, and it increases the share of economic activity traded in

<sup>75</sup> Arts Council of England (1998) reproduced in Heilbrun and Gray (2001, p. 254).

<sup>76</sup> Ibid.

<sup>77</sup> Schuster (2002) states that in the US direct support of arts and culture at the state level is more important than at the national level. He argues, among other things, that state support will pay more attention to audiences and participation, engage more in diversity and multiculturalism, tend to support more popular culture as opposed to fine arts and be based more on political influence compared to the national support; see further Chapter 35 by Netzer and Chapter 36 by Schuster in this volume.

<sup>78</sup> This hypothesis is based on casual observation only (but is easily testable); national politicians may favor the capital in which they live more than local politicians not living in the capital, not least because institutions of national importance tend to locate in the capital. Thus a redistribution of cultural funds from local governments to the center may result in higher cultural subsidies for the capital. In Germany this is clearly the case as the Undersecretary of Culture (*‘Kulturstaatsminister’*) supports heavily central foundations, museums and cultural institutions, most of them located in and around Berlin such as Stiftung Preußischer Kulturbesitz, Stiftung Preußische Schlösser und Gärten Berlin-Brandenburg, various museums in Berlin and Bonn (the former seat of government), and Stiftung Weimarer Klassik, cf. <http://www.bundesregierung.de/Bundesregierung/Beauftragte-fuer-Kultur-und-Me-9332/Kulturfoerderung.htm> (22.10.05).

markets. Technological progress pertains to live performing and visual arts less than to the production of other goods, including industrially-produced cultural goods such as books, music and film CDs and DVDs. As a consequence the relative price of these art forms (e.g., plays, opera, concerts, unique sculptures, etc.) tends to rise, thereby reducing the demand for them through a substitution effect. At the same time, technological progress brings about increases in per capita incomes and in average educational attainment, both of which tend to increase the demand for arts and culture through an income effect. Obviously, income and substitution effects work in the same direction for industrially-produced cultural goods such as music CDs, but have opposing directions for unique art such as theatrical plays.<sup>79</sup> The net effect is undetermined a priori; it depends to a substantial extent on how governments react to an increase of relative prices. The art forms affected by rising costs are those that enjoy the largest subsidies as a share of production costs.

Technological advances in the Internet and in reproduction technologies for digital cultural products – music, films, and digital books – have led to copyright infringements and declining sales. At this point we can only surmise all the possible effects that this will have in the future. Average movie production budgets may fall as it may become increasingly difficult to recoup them as record sales decline. The complementarity between live shows and record sales and movie visits and video sales will weaken with the consequence of rising ticket prices. For less developed countries, development will also increase the share of monetized transactions which will change the cultural and artistic production from a community and leisure activity into a for-profit activity; increased division of labor will lead to the emergence of full-time artists.

For the regional culture we observe two countervailing effects – a positive income effect that tends to increase cultural expenditures and a substitution effect that tends to divert expenditures away from live performing arts due to the relative price effect. Those forms of culture and art that experience productivity gains and are not subject to Baumol's cost disease exhibit strong scale economies (recorded music and movies, digital art) and thus tend to be present on a national or even international level. Thus even though we cannot know a priori what the net effect of the substitution and income effects will be on local performing arts, there is a strong presumption that the *share* of cultural goods produced by cultural industries (in constant prices) will increase (as income and substitution effects work in the same direction). This will tend to reduce the influence of truly *local* culture.

<sup>79</sup> With rising incomes the opportunity costs of visiting the arts goes up as well, thus establishing a countervailing force. In a longer-term perspective, however, development has not only meant increasing per capita levels, but also increased leisure time thus making cultural consumption easier.

## 4. Location of arts and cultural industries

### 4.1. Delineation of the cultural industries

Before considering factors affecting the location of the arts and cultural industries it is important to consider how these industries are delineated. The cultural or creative industries can be defined as industries that mass-produce goods and services with sufficient artistic content to be considered creative and culturally significant.<sup>80</sup> The essential features are industrial-scale production combined with cultural content [Towse (2003)]. The guiding principle that is increasingly adopted for defining the cultural industries is their reliance upon copyright law to protect the creative or cultural content, i.e. their intellectual property. Thus, “industries protected by copyright” have become virtually synonymous with the cultural or creative industries and as such they have been measured for their contribution to the GDP in a number of countries. A typical list consists of advertising, architecture, the art market, crafts, design, fashion, film, the music industry, performing arts, publishing, software, toys and games, television and radio, and videos. In most developed countries in which the size of the cultural industries has been measured, their share of the GDP is around 5 percent [Towse (2003)]. In Denmark their share has been calculated to 5.3 percent of GDP, in Great Britain 8 percent, in Sweden 9 percent, in the USA 7.8 percent, in New Zealand 3.1 percent of the GDP and in Norway 3.5 percent [Haraldsen et al. (2004)].

Cultural industries have attracted special attention from a development perspective because the cultural sector in itself – especially the cultural industries – is characterized as a growth sector. The growth is caused by increasing demand for experiences, culture, events and leisure-time activities in the course of development [Pine and Gilmore (1999); Wahlström (2002)]. The growth rate of the “creative economy” (usually defined as cultural industries plus the creative and performing arts) averages about 5 percent per annum and is therefore higher than that of “traditional” manufacturing industries [Towse (2001, 2003)].<sup>81</sup> It is therefore believed that the creative industries offer good employment prospects due to their potential for growth. Besides, many cultural industries are export industries selling their products outside the local area. The cultural industries are therefore described as important producers of substance in “the new economy”, where technological improvements, computers and the Internet play an important role.

The measurement of cultural employment and earnings in this sector is, however, so fraught with difficulties that such claims are hard to verify. Thus, these figures present many problems. The official statistics are often too aggregated and do not allow for including the relevant parts of different trades. For instance, it is questionable whether

<sup>80</sup> See further Caves (2002) and Chapter 17 in this volume.

<sup>81</sup> The Norwegian study [Haraldsen et al. (2004)] has found growth in the period 1996–2001 only in what is called “performing art” or the creative core. This might be due to the Norwegian study not including computer software which is included in most other studies, and which is characterized by a substantial growth. This illustrates the definitional problems affecting these kinds of studies quite clearly.

the whole travel trade or computer software should be included. Delineations are quite different in different countries, and the classification of different areas varies considerably. Moreover, cultural industries comprise a very wide spectrum from commercial businesses to subsidized cultural institutions and non-profit organizations, and there are also many different kinds of culture with different structures, economy and development potential.

In the following sections we will trace some of the guiding principles for the location of arts and culture, depending on the economic characteristics of the institutions and individuals that comprise the cultural sector.

#### *4.2. Agglomeration economies and the location of arts and culture*

Economic activities cluster because distance, or rather the absence of it, matters. Two relevant proximities that matter and often coincide can be distinguished: proximity to market and proximity to other producers, either competitors or producers of intermediate inputs.<sup>82</sup> These reflect positive externalities in consumption or production at the local level. In other words, agglomeration economies can be due to economies of scale and scope either in production or in consumption (or both).

##### *4.2.1. Proximity to market*

Proximity to potential customers matters to cultural institutions as distance is a major determinant of people's decision to visit the institution; it also affects existence and option values.<sup>83</sup> This is relevant to live performing arts such as theaters, ballets, opera houses and symphony orchestras, and also to museums, galleries, etc. Due to indivisibilities such institutions tend to locate only in larger cities: for instance, a symphony orchestra has a minimum size which must be supported by a large enough audience to operate profitably. The same holds true for ballet companies and opera houses, and to a lesser extent for theaters or museums. Thus smaller towns should tend to have smaller and/or fewer museums and theaters. In countries that subsidize the live performing arts heavily, however, this economic rationale may not apply to the same extent. because of political interventions. For instance, in Germany there are many more orchestras per inhabitant than in many other European countries and they are much more scattered, a result of the political fragmentation that occurred in the 17th and 18th centuries. Local courts saw orchestras as prestige objects and supported them [Schulze and Rose (1998)]. This tradition has remained and even today excellent symphony orchestras are found in the German hinterland. This is clearly not the case in a country such as the US where local public support is much smaller.

<sup>82</sup> Agglomeration phenomena are central to the 'new economic geography' literature; for surveys see, for instance, Fujita, Krugman and Venables (1999), Neary (2001) and Fujita and Thisse (2002).

<sup>83</sup> See, for example, Schulze and Ursprung (2000).



If live performing arts, museums, galleries, etc. locate where their customers are, agglomeration of these cultural and artistic institutions will follow the more general agglomeration of economic activity and of people. If we take into account that larger towns and cities tend to have larger per capita incomes and better educational attainment than the hinterland, agglomeration of cultural and artistic institutions should be even more pronounced in these locations. Of course, cultural tourists are attracted more easily the larger the supply of cultural institutions and activities. That constitutes a second externality in consumption and thus a reason for cultural institutions to cluster. This effect may show even within a city where museums locate in a museums district (Vienna) or a museum island (Berlin).<sup>84</sup>

#### 4.2.2. *Economies of scale and scope in production*

While there is a demand-side explanation of agglomeration due to consumers' positive travel costs (including opportunity costs of time), there is a supply-side explanation of agglomeration as well, arising from economies of scale and scope either at the firm level or the local industries level. The concept of agglomeration economies has been defined by Kaldor (1970):

... nothing else but the existence of increasing returns to scale – using that term in the broadest sense – in processing activities. These are not just the economies of large-scale production, commonly considered, but the cumulative advantages accruing from the growth of industry itself – the development of skill and know-how; the opportunities for easy communication of ideas and experience; the opportunity of ever-increasing differentiation of processes and of specialization in human activities (p. 340).

If economies of scale or scope occur at the firm level, they lead to larger firms, but not necessarily to agglomeration of different firms. For instance, there are obvious economies of scope for a symphony orchestra in combination with an opera house or a ballet company. Economies of scale occur due to significant fixed costs. At one level these economies of scale apply to all live performing arts; since a large part of the fixed cost is maintenance and capital costs of the building and costs for technical and artistic personnel, an additional show increases overall costs only slightly. At a different level, closely-related products can be offered without much additional cost. Theaters, for instance, can have a large stage and a studio stage where they perform more experimental plays for smaller audiences.

<sup>84</sup> Cultural districts are a prime example of economies of scale in consumption. Joint location of cultural institutions may attract larger audiences as they reduce travel costs (monetary and opportunity costs) to each cultural institution. A critical mass of cultural activities may be necessary to make the visit only worthwhile. Some of the cultural districts may also have economies of scope in production. See, for example, Scott (2000), Brooks and Kushner (2001); cf. Chapter 31 by Santagata in this volume on cultural districts.

Clustering of different firms is caused by economies of scale or scope external to the firm. These agglomerations have three principal causes: labor-market economies, scale economies in the production of intermediate inputs, and communication economies on the local level [O'Sullivan (1993)]. They all refer to essential inputs for the production by cultural industries.

#### 4.2.3. Input markets

Labor market economies of scale and scope occur when firms tap a common pool of artists which are hired only for short periods of time, then to be released into that pool again to be hired by other firms. This is typical of the London market for studio musicians; a further example is the movie industry where actors, extras and stuntmen are hired only for the duration of one film or even only for the scenes they play in. After one engagement they are available for new engagements. The larger the pool of specialized artistic labor, the easier it is for a firm to find the specific qualities they look for in the artists. In turn, the more production activities there are in a given area, the stronger is the incentive for actors to move into this area – in other words agglomeration is reinforcing. Los Angeles and New York are such agglomeration areas; Los Angeles is the center for actors and directors working in the movie industry in the US, while New York has become the point of gravity for live theater and dance. In 1990, 11 percent of all performing artists lived in New York and 13.2 percent lived in Los Angeles, while the relative shares of the population were only 3.4 and 3.6 percent, respectively [Heilbrun and Gray (2001, Chapter 15)].

The above line of argument applies not only to artistic labor, but also to highly specialized inputs. Again, Hollywood serves as an example as the movie industry hires a multitude of highly specialized inputs that range from special effects animation to post-production.<sup>85</sup> As their services are required only for short periods, an economically viable production mode requires that these inputs are hired by many firms. Producers of specialized inputs cluster around the producer of the final product, but because of short-lived contracts and the high degree of specialization, agglomeration results. Thus these characteristics of the production process, vertical disintegration and product innovation have led to the need for specialized inputs, a demand served at a relatively low level of product standardization by a cluster of creative and dynamic small-sized firms. Social capability and interactive learning are the basic factors of the local expansion. Such an agglomeration creates many kinds of external benefits and synergies for the firms [Scott (2002); Garreau (1992)].

#### 4.2.4. Creativity spillovers and the location of artists

Unlike performing artists, many other types of artists do not have to be close to their consumers, nor do they rely on intermediate inputs in order to create their artistic product.

<sup>85</sup> See further Chapter 19 by de Vany in this volume.

Thus they do not face “technological” incentives for agglomeration like live performing arts, museums, or the movie industry or studio music do. Among these artists are writers, painters, composers, even sculptors. After completion of a long creative process they contact their intermediaries – literary agents, galleries – or their buyers or publishers directly and stock up on the few inputs they need. In principle they could reside in solitude and some actually have done so for extended periods of time. However many of them have opted to live close to their peers or actually with them. For instance, Montmartre in Paris used to be an artist quarter, Worpswede, near the city of Bremen, has been an artist colony in an isolated spot where painters have worked side by side. There are many more examples.<sup>86</sup> The reason is not a technological one, but rather one of mutual inspiration, in other words “creative spillovers” have led to a clustering of artists.

An institutionalized form of creativity spillovers occurs in art schools and colleges and other modes of art education which attract young artists, create an artistic atmosphere and thus provide incentives for artists to stay after their formal education has been completed. One prime example is the Bauhaus in Dessau, another one is Ubud in Bali/Indonesia, where young artists from all over Indonesia go in order to be educated. These creativity spillovers are of course not limited to painters, writers and sculptors but apply in principle to all art forms, thereby reinforcing existing incentives to cluster.

#### *4.3. The particular role of cities*

Culture may be an important element in a comprehensive urban development strategy. The literature on “the creative city”<sup>87</sup> emphasizes the importance of the arts and culture for the economic development of cities. A critical element is the relationship of urban centers and suburban regions and the hinterland; the balanced growth hypothesis states that the development of the urban center is necessary for suburban and regional development. This hypothesis has been subject to academic discussion. Especially in the US, there has been debate as to whether city and suburban growth are substitutes or complements.<sup>88</sup> Some have argued that suburbs are no longer dependent on central cities, while others have argued that central cities and their suburban areas remain closely interconnected.

What do we know about “balanced growth”, and to what degree is the vitality of a central city critical for the future of the whole region? This question is interesting from the perspective of arts and culture in urban and regional development because if arts and culture play important roles in preserving and rejuvenating central cities and core urban neighborhoods, they can be essential for the long-run health of the entire metro area and larger region. Ihlanfeldt (1995) identifies some major “sources of interdependence” that in his view link the economies of central cities and their surrounding suburbs:

<sup>86</sup> Of course large cities provide proximity to fellow artists without making this “agglomeration” visible.

<sup>87</sup> See, for example, Landry (2000) and Landry and Bianchini (1995).

<sup>88</sup> See, for example, Voith (1992, 1993), Ihlanfeldt (1995), and Glaeser (1998).

- Outsiders' perceptions of the region, which are influenced by conditions prevailing within the central city;
- Location-specific and historically unique amenities in the central cities, which are valued throughout the entire region;
- The related notion that central cities provide a "sense of place" that is highly valued not only by central city residents, but by others living in the metro area who "identify" in some way with those things that make a city a unique place to live;
- Fiscal interdependence between a central city and the rest of the region; and
- The "traditional" concept that central cities offer unique "agglomeration economies" that provide a specialized role for the central city.

Ihlanfeldt concludes that there is no empirical evidence on the quantitative significance of "perceived image" or "sense of place" effects. He argues that the best way to measure the importance of unique central-city amenities would be via an "inter-area hedonic wage or housing equation". However, he does find some empirical support for the existence of agglomeration economies; this finding is at least consistent with the possibility that central cities make an important contribution to growth, and are perhaps even engines of growth, because their compact development allows for particularly strong agglomeration economies.

As Seaman (2001) points out, ambiguity about the direction of causality is a critical limitation to our current understanding of the role of central cities in regional development, and greatly limits our ability to draw strong inferences from "simple" tests of the balanced growth hypothesis such as studies of correlations between central city and suburban economic and demographic variables. Therefore the proposition remains plausible, but only weakly empirically confirmed.

## **5. Concluding remarks: Regional cultural and economic development – a complex simultaneity**

In this chapter we have reviewed the literature on cultural and regional development. We have demonstrated that there is a complex simultaneity between cultural and economic development. It has become clear that local culture will change in the course of economic and technological development. Demand changes not only because of increased income and educational attainment, but also due to technological developments that create new art forms and open up possibilities for transmission of culture. These changes in supply conditions are complemented by differentially lower productivity gains in the live performing arts thus triggering relative price increases. How demand (and public support) will react to this price increase remains an open question; the rising income will increase demand but at the same time alternative art forms that are not subject to stagnating productivity will become more attractive.

We have seen that art and culture can play a significant role in urban and regional development, even if we focus only on the level of overall economic activity. The contribution of culture to overall development is even larger if we apply a broader concept

of development. While short-run spending impacts can be measured relatively easily, longer-run effects on location choice of mobile people and enterprises as well as on overall creativity of a region are much harder to identify and to quantify. Here theoretical concepts remain somewhat vague and empirical assessments tentative. But assessments of the economic effects of culture, whether short-term or long-term, should neither serve as a justification for, nor as evidence against, certain cultural policies. It is not the sole goal of regional cultural policy to promote economic growth in a narrow sense, nor is it even the dominant goal.

Nevertheless, from a policy perspective the effect of culture on economic development is important in the formulation of cultural policy. While it is possible at the theoretical level to localize different “transmission channels” – ways in which culture can contribute to economic development – the matter becomes very complex if is transferred to a concrete policy situation. The degree of success in a given case is dependent on a series of factors and how these factors are mixed. Among the relevant questions are the following:

- *What kinds of cultural assets are “added” into the geographical area?* The effect of additional provision of culture on urban and regional development depends of course on the cultural asset that is “added” to the current cultural environment. A festival, a theater, a museum, facilities for creative artists, a “cultural industry”, etc. will have different impact on tourism, location decisions of firms, local residents, etc. For example, a festival or a large scale museum may be important for tourism, while libraries, etc. will cater only to local residents. Cultural industries, like the production of recorded music, film, TV and videos, are mainly export industries from the regional perspective, thus making them important for employment and economic development in the local area, but of minor importance for the local residents and for tourists in the area. Obviously, even within a category of cultural asset its effect depends on the degree of uniqueness of the asset under consideration. Current research does not tell us very much about these differential effects of cultural institutions and industries.
- *What is the current level of art and culture in the area?* The effect of a specific new cultural asset depends on the environment in which it is placed, in particular on the existing art and culture in the area. We refer here to synergy effects on the supply side which are responsible for agglomeration phenomena and which depend on the structure and level of cultural activities in place. We refer also to demand effects. In particular a “critical mass” of cultural artifacts needs to be in place to attract locals and tourists from outside the area.
- *What is the state of the existing environment (location factors such as infrastructure, quality of labor, natural amenities, etc.)?* The entire environment is important for the success of a cultural institution; relevant factors include infrastructure (transport, telecommunication and informational infrastructure), tourist infrastructure (hotels and restaurants), nature and climate. Their relative importance depends on the transmission channel – if we focus on cultural tourism, it will be hotels, restaurants, infrastructure, nature, climate, etc. that are important; if we focus on

culture as a factor for attracting new residents, other factors affecting people's localization choices must be considered such as schools, houses, job opportunities, tolerance, etc. When we look at arts and culture as location factors for enterprises, it will be tax rates, infrastructure, labor availability and costs, etc. that are relevant.

In all cases the socio-economic profile of the residents will make a difference.

There is no simple formula for success. Cultural development needs to be embedded into an overall urban and regional development strategy<sup>89</sup> that takes account of the existing situation for each region and emphasizes complementarities between the different areas of regional development.<sup>90</sup> Eventually, we need a better understanding of the dynamic process of regional and cultural development based on historical analyses which would allow us to assess the incremental contribution of additional artists and arts institutions to the urban and regional economy.

Cultural economics has produced a realm of knowledge about the influence of cultural development on economic development and conversely on how economic progress shapes the cultural landscape. Yet we are still missing a convincing theory that is able to portray the complex dynamics and interdependence of cultural and economic development. Such a theory would have particular relevance to development at urban and regional levels.

## References

- Abele, H., Bauer, H. (1984). *Die Bundestheater in der Österreichischen Wirtschaft*. Wirtschaftsuniversität, Wien.
- Adamowicz, V., Boxall, P. (2001). "Future directions of stated choice methods for environment valuation". Paper prepared for Choice Experiments: A New Approach to Environmental Valuation April 10, 2001, London, England, <http://www.irs.aber.ac.uk/mec/CEconfpaper/UK%20Workshop%20paper.PDF>.
- Adamowicz, V., Boxall, P., Williams, M., Louviere, J. (1998). "Stated preference approaches for measuring passive use values: Choice experiments and contingent valuation". *American Journal of Agricultural Economics* 80 (1), 0002–9092.
- Alberini, A., Riganti, P., Longo, A. (2003). "Can people value the aesthetic and use services of urban sites? Evidence from a survey of Belfast residents". *Journal of Cultural Economics* 27, 193–213.
- Bartik, T.J. (1991). *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.
- Bateman, I.J., Carson, R.T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Özdemiroglu, E., Pearce, P.W., Sugden, R., Swansson, J. (2002). *Economic Valuation with Stated Preference Techniques: A Manual*. Edward Elgar, Cheltenham.
- Baumol, W. (1996). "Children of performing arts, the economic dilemma: The climbing costs of health care and education". *Journal of Cultural Economics* 20 (3), 183–206.
- Baumol, H., Baumol, W. (1985). "The future of the theater and the cost disease of the arts". In: Hendon, M.A., Richardson, J., Hendon, W. (Eds.), *Bach and the Box: The Impact of Television on the Live Arts*. Association for Cultural Economics, Akron.

<sup>89</sup> See further, for example, Evans (2001), Bianchini and Parkinson (1993) and Mercer (2002).

<sup>90</sup> See Schuster (1988) and Clarke and Kahn (1988).

- Baumol, W.J., Bowen, W.G. (1966). *Performing Arts: The Economic Dilemma*. Twentieth Century Fund, New York.
- Bianchini, F., Parkinson, M. (Eds.) (1993). *Cultural Policy and Urban Regeneration: The West European Experience*. Manchester University Press, Manchester.
- Bille, T., Hjorth-Andersen, C., Gregersen, M. (2003). "A tale of tables – on public cultural expenses". In: Duelund, P. (Ed.), *The Nordic Cultural Model*. Nordisk Kultur Institut, Copenhagen.
- Bille, T., Fridberg, T., Storgaard, S., Wulff, E. (2005). *Danskernes kultur – og fritidsaktiviteter – med udviklingslinier tilbage til 1964*. AKF Forlaget, Copenhagen.
- Bille Hansen, T. (1991). *Det Kgl. Teater – et kulturøkonomisk studie*. AKF Forlaget, Copenhagen.
- Bille Hansen, T. (1995). "Measuring the value of culture". *Journal of Cultural Policy* 1 (2), 309–322.
- Bille Hansen, T. (1997). "The willingness-to-pay for the Royal Theater in Copenhagen as a public good". *Journal of Cultural Economics* 21 (1), 1–28.
- Bille Hansen, T. (1998). "The Danes value the Royal Theatre in Copenhagen". *Heritage, the Environment and the Arts: Pricing the Priceless* 6 (3), 38–66. Hume Papers on Public Policy, Edinburgh University Press.
- Bille Hansen, T. (2002). "A contingent valuation study of the Royal Theater in Copenhagen". In: Navrud, S., Ready, R.C. (Eds.), *Valuing Cultural Heritage. Applying Environment Valuation Techniques to Historic Buildings, Monuments and Artifacts*. Edward Elgar, Cheltenham.
- Bishop, R.-C., Heberlein, T. (1986). "Does contingent valuation work?" In: Cummings, R., Brookshire, D., Schulze, W. (Eds.), *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*. Rowman and Allanheld, Totowa, NJ, pp. 123–147.
- Bonet, L. (2003). "Cultural tourism". In: Towse, R. (Ed.), *Handbook of Cultural Economics*. Edward Elgar, Cheltenham.
- Boxall, P., Englin, J., Adamowicz, V. (2002). "The contribution of aboriginal rock paintings to wilderness recreation values in North America". In: Navrud, S., Ready, R. (Eds.), *Valuing Cultural Heritage: Applying Environmental Valuation Techniques to Historic Buildings, Monuments and Artifacts*. Edward Elgar, Cheltenham.
- Boxall, P., Adamowicz, W., Swait, J., Williams, M., Louviere, J. (1996). "A comparison of stated preference methods for environmental valuation". *Ecological Economics* 18, 243–253.
- Braden, J.B., Kolstad, C.D. (1991). *Measuring the Demand for Environmental Quality*. North-Holland, Amsterdam.
- Brandt, A., Klodt, T., Kramer, J., Ertel, R., Schasse, U., Revilla Diez, J. (2001). *Regionalwirtschaftliche Aspekte der Expo 2000 – Eine Schlussbilanz, Gutachten im Auftrag des Niedersächsischen Ministeriums für Wirtschaft, Technologie und Verkehr und des Kommunalverbandes Großraum Hannover*. Norddeutsche Landesbank Girozentrale, Niedersächsisches Institut für Wirtschaftsforschung e.V., Universität Hannover, Hannover.
- Brooks, A.C., Kushner, R.J. (2001). "Cultural districts and urban development". *International Journal of Arts and Management* 3, 4–15.
- Carlsson, F., Martinsson, P. (2001). "Do hypothetical and actual marginal willingness to pay differ in choice experiments? Application to the valuation of the environment". *Journal of Environmental Economics and Management* 41, 179–192.
- Carson, R.T., Flores, N.E., Meade, N.F. (2001). "Contingent valuation: Controversies and evidence". *Environmental and Resource Economics* 19 (2), 173–210.
- Caves, R. (2002). *Creative Industries: Contracts Between Art and Commerce*. Harvard University Press, Cambridge, MA.
- Christoffersen, H., Lyk-Jensen, P. (1994). *Nogle økonomiske effekter af Brandts Klædefabrik i Odense*. AKF Forlaget, Copenhagen.
- Clark, D.E., Kahn, J.R. (1988). "The social benefits of urban cultural amenities". *Journal of Regional Science* 28, 363–377.
- Cooper, C. (1997). "Show them the money!" *Calculating the Economic Impact of America's Orchestras*. American Symphony Orchestras League, Government Affairs Office, New York.
- Cooper, C., Fletcher, J., Gilbert, D., Wanhill, S. (1993). *Tourism: Principles and Practice*. Longman, Harlow.

- Cowen, T. (1996). "Why I do not believe in the cost-disease. Comment on Baumol". *Journal of Cultural Economics* 20 (3), 207–214.
- Cwi, D., Lyall, K. (1977). *Economic Impact of Arts and Cultural Institutions: A Model of Assessment and a Case Study in Baltimore*. National Endowment for the Arts, Washington, DC.
- Department of Culture, Media and Sports (1998). *Creative Industries Mapping Document*, London.
- Department of Culture, Media and Sports (2001). *Creative Industries Mapping Document*, London.
- Dewenter, R., Westermann, M. (2005). "Cinema demand in Germany". *Journal of Cultural Economics* 29 (3), 213–231.
- Diamond, P.A., Hausman, J.A. (1994). "Contingent valuation: Is some number better than no number?" *Journal of Economic Perspectives* 8 (4), 45–69.
- Easterly, W. (2001). *The Elusive Quest for Growth*. MIT Press, Cambridge, MA.
- Evans, G. (2001). *Cultural Planning. An Urban Renaissance?* Routledge, London.
- Felton, M.V. (1992). "On the assumed inelasticity of demand for the performing arts". *Journal of Cultural Economics* 16 (1), 1–12.
- Fernandes-Blanco, V., Banos Piño, J. (1997). "Cinema demand in Spain: A cointegration analysis". *Journal of Cultural Economics* 21 (1), 57–75.
- Florida, R. (2002a). *The Rise of the Creative Class – and How It's Transforming Work, Leisure, Community and Everyday Life*. Basic Books, New York.
- Florida, R. (2002b). "Bohemia and economic geography". *Journal of Economic Geography* 2, 55–71.
- Ford Foundation (1974). "The Finances of the Performing Arts, vol. II: A Survey of the Characteristics and Attitudes of Audiences for Theater, Opera, Symphony, and Ballet in 12 U.S. Cities". Ford Foundation, New York.
- Frank, B. (2004). "Art price research for fakes and imitations". In: Mossetto, G., Vecco, M. (Eds.), *The Economics of Copying and Counterfeiting*. FrancoAngeli, Milano, pp. 119–131.
- Freeman III, A.M. (1993). *The Measurement of Environmental and Resource Values. Theory and Methods. Resources for the Future*, Washington, DC.
- Frey, B.S. (1996). "Has Baumol's cost disease disappeared in the performing arts?" *Ricerche Economiche* 50, 173–182.
- Frey, B.S. (1997). "Evaluating cultural property: The economic approach". *International Journal of Cultural Property* 6 (2), 231–246.
- Frey, B.S. (2003). "Public support". In: Towse, R. (Ed.), *Handbook of Cultural Economics*. Edward Elgar, Cheltenham, pp. 389–398.
- Frey, B.S., Pommerehne, W.W. (1989). *Muses and Markets: Explorations in the Economics of Arts*. Basil Blackwell, Oxford.
- Frey, B.S., Pommerehne, W.W. (1995). "Public expenditure on the arts and direct democracy: The use of referenda in Switzerland". *Cultural Policy* 2 (1), 55–65.
- Fujita, M., Krugman, P., Venables, A. (1999). *The Spatial Economy*. The MIT Press, Cambridge, MA.
- Fujita, M., Thisse, J.-F. (2002). *Economics of Agglomeration*. Cambridge University Press, Cambridge.
- Gapinski, J.H. (1984). "The economics of performing Shakespeare". *American Economic Review* 74 (3), 458–466.
- Gapinski, J.H. (1986). "The lively arts as substitutes for the lively arts". *American Economic Review* 76 (2), 20–25.
- Garreau, J. (1992). *Edge City*. Anchor Books, New York.
- Gazel, R., Schwer, R.K. (1997). "Beyond rock and roll: The economic impact of the Grateful Dead on a local economy". *Journal of Cultural Economics* 21 (1), 41–55.
- Getzner, M. (2004). "Voter preferences and public support for the performing arts: A case-study for Austria". *Empirica* 31 (1), 27–42.
- Gibson, L., Stevenson, D. (2004). "Urban space and the uses of culture". *International Journal of Cultural Policy* 10 (1), 1–5.
- Glaeser, E.L. (1998). "Are cities dying?" *Journal of Economic Perspectives* 12 (2), 139–160.
- Goldstone, L. (2003). "Cultural statistics". In: Towse, R. (Ed.), *Handbook of Cultural Economics*. Edward Elgar, Cheltenham, pp. 177–182.



- Gomez, M.V. (1998). "Reflective images: The case of urban regeneration in Glasgow and Bilbao". *International Journal of Urban and Regional Research* 22 (1), 106–122.
- Graves, F.L. (1987). "The role of economics in evaluating public investments in the arts". In: Hendon, W.S., Hillman-Chartrand, H., Horowitz, H. (Eds.), *Paying for the Arts*. Association for Cultural Economics, University of Akron.
- Gray, C. (2003). "Participation". In: Towse, R. (Ed.), *Handbook of Cultural Economics*. Edward Elgar, Cheltenham, pp. 356–365.
- Hanemann, W.M. (1994). "Valuing the environment through contingent valuation". *Journal of Economic Perspectives* 8 (4), 19–34.
- Haraldsen, T., Flygind, S.K., Overvag, K., Power, D. (2004). Kartlegging av kulturnæringene i Norge – økonomisk betydning, vekst – og utviklingspotensial, Østlandsforskning, ØF-rapport No. 10.
- Heilbrun, J. (2001). "A correction to Hilda and William Baumol's paper 'The future of the theater and the cost disease of the arts' ". *Journal of Cultural Economics* 25 (2), 149–150.
- Heilbrun, J. (2003). "Baumol's cost disease". In: Towse, R. (Ed.), *Handbook of Cultural Economics*. Edward Elgar, Cheltenham.
- Heilbrun, J., Gray, C.M. (2001). *The Economics of Art and Culture*. Cambridge University Press, Cambridge.
- Hummel, M., Berger, M. (1988). *Die Volkswirtschaftliche Bedeutung von Kunst und Kultur*. Duncker & Humboldt, Berlin.
- Ihlanfeldt, K. (1995). "The importance of the central city to the regional and national economy: A review of the arguments and empirical evidence". *Cityscape: A Journal of Policy, Development and Research* 1 (2), 125–150.
- Johnson, P., Thomas, B. (1992). *Tourism, Museums and the Local Economy: The Economic Impact of the North of England Open Air Museums at Beamish*. Avebury, Aldershot.
- Kaldor, N. (1970). "The case for regional policies". *Scottish Journal of Political Economy* 17, 337–347.
- Kindleberger, C., Herrick, B. (1977). *Economic Development*. McGraw–Hill, London.
- Kotkin, J. (2000). *The New Geography; How the Digital Revolution is Reshaping the American Landscape*. Random House, New York.
- Krebs, S., Pommerehne, W. (1995). "Politico-economic interactions of German public performing arts institutions". *Journal of Cultural Economics* 19 (1), 17–32.
- Krueger, A.B. (2005). "The economics of real superstars: The market for rock concerts in the material world". *Journal of Labor Economics* 23 (1), 1–30.
- Lancaster, K. (1966). "Change and innovation in the technology of consumption". *American Economic Review* 56, 14–23.
- Landry, C. (2000). *The Creative City. A Toolkit for Urban Innovators*. Commedia, Earthscan Publications Ltd., London.
- Landry, C., Bianchini, F. (1995). *The Creative City*. Demos, London.
- Lévy-Garboua, L., Montmarquette, C. (2003). "Demand". In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham, pp. 201–213.
- Lewis, G., Seaman, B.A. (2004). "Sexual orientation and demand for the arts". *Social Science Quarterly* 85, 523–538.
- Martin, F. (1994). "Determining the size of museum subsidies". *Journal of Cultural Economics* 18 (4), 255–270.
- Mercer, C. (2002). *Towards Cultural Citizenship: Tools for Cultural Policy and Development*. The Bank of Sweden Tercentenary Foundation and Gidlunds Förlag, Stockholm.
- Mitchell, C. (1989). "The arts and employment: The impact of three Canadian theater companies". *Journal of Cultural Economics* 13 (2), 69–81.
- Mitchell, R.C., Carson, R.T. (1989). *Using Surveys to Value Public Goods. The Contingent Valuation Method*. Resources for the Future, Washington, DC.
- Mitchell, C., Wall, G. (1989). "The arts and employment: A case study of the Stratford Festival". *Growth and Change* 20 (4), 31–40.
- Morrison, W.G., West, E.G. (1986). "Subsidies for the performing arts: Evidence on voters' preference". *Journal of Behavioral Economics* 15, 57–72.

- Myerscough, J. (1988a). *The Economic Importance of the Arts in Britain*. Policy Studies Institute, London.
- Myerscough, J. (1988b). *The Economic Importance of the Arts in Glasgow*. Policy Studies Institute, London.
- Navrud, S., Ready, R. (Eds.) (2002). *Valuing Cultural Heritage. Applying Environmental Valuation Techniques to Historic Buildings, Monuments and Artifacts*. Edward Elgar, Cheltenham.
- Navrud, S., Strand, J. (2002). "Social costs and benefits of preserving and restoring the Nidaros Cathedral". In: Navrud, S., Ready, J. (Eds.), *Valuing Cultural Heritage. Applying Environmental Valuation Techniques to Historic Buildings, Monuments and Artifacts*. Edward Elgar, Cheltenham.
- Neary, P. (2001). "Of hype and hyperbolas: Introducing the new economic geography". *Journal of Economic Literature* 39 (2), 536–561.
- Noonan, D.S. (2003). "Contingent valuation and cultural resources: A meta-analytic review of the literature". *Journal of Cultural Economics* 27 (3–4), 159–176.
- O'Hagan, J. (1989). *The Economic and Social Contribution of the Wexford Opera Festival*. Trinity College, Dublin.
- O'Hagan, J. (1992). "The Wexford Opera Festival: A Case for Public Funding?" In: Towse, R., Khakee, A. (Eds.), *Cultural Economics*. Springer-Verlag, Berlin.
- O'Hagan, J. (1998). *The State and the Arts*. Edward Elgar, Cheltenham.
- O'Sullivan, A. (1993). *Urban Economics*. Irwin, Homewood, IL.
- Ottaviano, G., Peri, G. (2005). "Cities and cultures". *Journal of Urban Economics* 58, 304–337.
- Palmer/Rae Associates (2004). "European cities and capitals of culture". Study Prepared for the European Commission, Brussels, August 2004, <http://www.palmer-rae.com>.
- Peacock, A. (1991). "Economics, cultural values and cultural policies". *Journal of Cultural Economics* 15 (2), 1–18.
- Perino, G., Schulze, G. (2005). "Competition, cultural autonomy and global governance: The audio-visual sector in Germany". In: Guerrieri, P., Iapadre, P.L., Koopmann, G. (Eds.), *Competition, Cultural Diversity, and Global Governance: The Case of Audiovisual Services*. Edward Elgar, Cheltenham, pp. 52–95.
- Pine, B.J., Gilmore, J.H. (1999). *The Experience Economy – Work is Theater and Every Business a Stage*. Harvard Business School Press, Boston, MA.
- Plaza, B. (1999). "The Guggenheim–Bilbao museum effect: A reply to Maria V. Gomez' reflective images: The case of urban regeneration in Glasgow and Bilbao". *International Journal of Urban and Regional Research* 23 (3), 589–592.
- Plaza, B. (2000). "Evaluation of the influence of a large cultural artifact in the attraction of tourism. The Guggenheim Museum Bilbao case". *Urban Affairs Review* 36 (2), 264–274.
- Pommerehne, W. (1982). "Steuern, Staatsausgaben und Stimmbürgerverhalten". *Jahrbücher für Nationalökonomie und Statistik* 197 (5), 438–462;
- Schneider, F., Pommerehne, W.W., "Private demand for public subsidies to the arts: A study in voting and expenditure theory". In: Hendon, W.S., Shanahan, J.L. (Eds.), *Economics of Cultural Decisions*. Abt Books, Cambridge, MA, 1982. English version.
- Pommerehne, W., Kirchgässner, G. (1987). "The impact of television on the demand for cinema and theater performances". In: Grant, N.K., et al. (Eds.), *Economic Efficiency and the Performing Arts*. Association for Cultural Economics, Akron, pp. 44–61.
- Poor, J., Smith, J. (2004). "Travel cost analysis of a cultural heritage site: The case of historic St. Mary's City of Maryland". *Journal of Cultural Economics* 28 (3), 217–229.
- Portney, P.R. (1994). "The contingent valuation debate: Why economists should care". *Journal of Economic Perspectives* 8 (4), 3–17.
- Radich, A.J. (1987). *Economic Impacts of the Arts – A Sourcebook*. National Conference of State Legislature, Denver.
- Radich, A.J. (1993). *Twenty Years of Economic Impact Studies in the Arts: A Review*. National Endowment for the Arts, Washington, DC.
- Rob, R., Waldfogel, J. (2004). "Piracy on the high C's: Music downloading, sales displacement, and social welfare in a sample of college students". NBER Working Paper No. 10874.
- Rosen, S. (1974). "Hedonic prices and implicit markets: Product differentiation in pure competition". *Journal of Political Economy* 82 (1), 34–55.

- Rubernowits, S., Rubernowits, U. (1990). "Kultur som attraction – vilken roll spelar kulturen för valet av bostadsort". ERU-rapport 65, Stockholm.
- Rushton, M. (2005). "Support for earmarked public spending on culture: Evidence from a referendum in metropolitan Detroit". *Public Budgeting and Finance*. In press.
- Ryan, M., Gerard, K. (2003). "Using discrete choice experiments to value health care programmes: Current practice and future reflections". *Applied Health Economics and Health Policy* 2 (1), 55–64.
- Santagata, W., Signorello, G. (2000). "Contingent valuation of a cultural public good and policy design: The case of 'Napoli Musei Aperti' ". *Journal of Cultural Economics* 24 (3), 181–204.
- Santagata, W., Signorello, G. (2002). "Individual preferences and allocation mechanisms for a cultural public good: Napoli Musei Aperti". In: Navrud, S., Ready, R. (Eds.), *Valuing Cultural Heritage. Applying Environmental Valuation Techniques to Historic Buildings, Monuments and Artifacts*. Edward Elgar, Cheltenham.
- Scanlon, R., Longley, R. (1984). "The arts as an industry; their economic importance to the New York–New Jersey metropolitan region". In: Hendon, W.S., Grant, N.K., Shaw, D.V. (Eds.), *The Economics of Cultural Industries*. Association for Cultural Economics, University of Akron.
- Schulze, G. (1999). "International trade in art". *Journal of Cultural Economics* 23 (1), 109–136.
- Schulze, G. (2003). "International Trade". In: Towse, R. (Ed.), *Handbook of Cultural Economic*. Edward Elgar, Cheltenham.
- Schulze, G., Rose, A. (1998). "Public orchestra funding in Germany: An empirical investigation". *Journal of Cultural Economics* 22 (4), 227–247.
- Schulze, G., Ursprung, H. (2000). "La Donna e mobile – or is she?: Voter preferences and public support for the performing arts". *Public Choice* 102, 131–149.
- Schuster, J.M. (Ed.) (1988). "The arts and urban development: Four case studies". Background material for the British American Arts Association conference in Glasgow, October 1988, Massachusetts Institute of Technology, Mimeo.
- Schuster, J.M. (1995). "Two urban festivals: La Mercè and First Night". *Planning Practice and Research* 10 (2), 173–187.
- Schuster, J.M. (2002). "Sub-national cultural policy – where the action is: Mapping state cultural policy in the United States". *International Journal of Cultural Policy* 8 (2), 181–196.
- Scott, J.A. (2000). *The Cultural Economy of Cities*. Sage, London.
- Scott, A.J. (2002). "A new map of Hollywood: The production and distribution of American motion pictures". *Regional Studies* 36 (9), 957–975.
- Seaman, B. (1987). "Arts economics impact studies: A fashionable excess". In: *Economic Impact of the Arts: A Sourcebook*. National Conference of State Legislatures, Washington, DC, pp. 44–75.
- Seaman, B. (2001). "Beyond economic impact". In: Morris, V.B., Pankratz, D.B. (Eds.), *Arts in the New Millennium*. Greenwood Publishing Group, Westport, CT.
- Seaman, B. (2002). "CVM vs. economic impact: Substitutes or complements?" Paper presented at the 12th International Conference on Cultural Economic, Rotterdam.
- Seaman, B. (2003). "The economic impact of the arts". In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham.
- Smith, T. (1998). "The addiction to culture". Paper presented on the Biannual Conference of the Association for Cultural Economics International in Barcelona, June. Also part of "Two essays on the economics of the arts: The demand for culture and the occupational mobility of artists". Unpublished Doctoral Dissertation, University of Chicago, September, 1998.
- Sommers, P., Carlson, D. (2000). "The new economy in metropolitan Seattle: High tech firm location decisions within the metropolitan landscape". Report for the Brookings Institution Center on Urban and Metropolitan Affairs, Washington, DC.
- Stabler, M. (1996). "Are heritage conservation and tourism incompatible? An economic valuation of their role in urban regeneration: Policy implications". In: Robinson, M., et al. (Eds.), *Managing Cultural Resources for the Tourists*. University of Northumbria, Newcastle.
- Stanley, D., et al. (2000). "Win, place or show; gauging the economic success of the Renoir and Barnes art exhibits". *Journal of Cultural Economics* 24 (3), 243–255.

- Stern, M.J. (1999). "Is all the World Philadelphia?: A multi-city study of arts and cultural organizations, diversity and urban revitalization". Working Paper, School of Social Work, University of Pennsylvania.
- Stigler, G., Becker, G. (1977). "De gustibus non est disputandum". *American Economic Review* 67, 76–90.
- Throsby, D. (1997). "Sustainability and culture: Some theoretical issues". *International Journal of Cultural Policy* 4, 7–20.
- Throsby, D. (2001). "Economics and Culture". Cambridge University Press, Cambridge.
- Throsby, D. (2003). "Cultural sustainability". In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham.
- Throsby, D., Withers, G. (1983). "Measuring the demand for the arts as a public good: Theory and empirical results". In: Hendon, W.S., Shanahan, J.L. (Eds.), *Economics of Cultural Decisions*. Abt Books, Cambridge.
- Towse, R. (Ed.) (1997). *Baumol's Cost Disease: The Arts and Other Victims*. Edward Elgar, Cheltenham.
- Towse, R. (2001). *Creativity, Incentive and Reward*. Edward Elgar, Cheltenham.
- Towse, R. (2003). "Cultural industries". In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham.
- Van Puffelen, F. (1987). "The economic impact of the arts in Amsterdam". In: Hendon, W.S., Hillman-Chartrand, H., Horowitz, H. (Eds.), *Paying for the Arts*. Association for Cultural Economics, University of Akron.
- Van Puffelen, F. (1996). "Abuses of conventional impact studies of the arts". *Cultural Policy* 2 (2), 241–254.
- Vaughan, D.R. (1979). "Does a festival pay?" In: Shanahan, J.L., Hendon, W.S., MacDonald, A.J. (Eds.), *Economic Policy for the Arts*. Abt Books, Cambridge, MA, pp. 319–331.
- Vidarte, J.I. (2000). "Culture, renewal and development". In: Wolfensohn, J., et al. (Eds.), *Culture Counts: Financing, Resources and the Economics of Culture in Sustainable Development*. World Bank, Washington, DC.
- Voith, R.P. (1992). "City and suburban growth: Substitutes or complements?" *Business Review*, Federal Reserve Bank of Philadelphia, pp. 21–33.
- Voith, R.P. (1993). "Does city income growth increase suburban income growth, house value appreciation and population growth?" Working Paper No. 93-27, Federal Reserve Bank of Philadelphia.
- Wahlström, B. (2002). "Guide til upplevelsessamhället: Från musik & museer till sushi & spa". SNS Förlag, Stockholm.
- Whitehead, J., Finney, S. (2003). "Willingness to pay for submerged maritime cultural resources". *Journal of Cultural Economics* 27 (3–4), 231–240.
- Wiesand, A. (1995). "Musiktheater und Konzerte: Mehr Rückhalt in der Bevölkerung. Aus den Ergebnissen des 5. Kulturbarmeters". *Das Orchester* 6/95, 2–14.
- Wildman, S., Siwek, S. (1988). *International Trade in Films and Television Programs*. Ballinger, Cambridge, MA.
- Zentner, A. (2003). "Measuring the effect of music downloads on music purchases". University of Chicago, Mimeo.

## CULTURAL DISTRICTS AND THEIR ROLE IN DEVELOPED AND DEVELOPING COUNTRIES\*

WALTER SANTAGATA

*University of Turin, Italy*

### Contents

Abstract	1102
Keywords	1102
1. Introduction	1103
2. The theory of cultural districts	1104
2.1. Marshallian localized industries and endogenous economic growth	1104
2.2. Culture-based goods as “idiosyncratic”	1105
3. Industrial cultural districts and clusters	1107
3.1. In theory	1107
3.2. In practice: The Los Angeles motion picture complex	1109
3.3. In practice: The Caltagirone (Sicily) pottery district	1110
4. Institutional cultural districts	1111
4.1. Institutional setting: The collective intellectual property rights	1111
4.2. Modeling behavior in the establishment of collective intellectual property rights	1113
4.3. The institutional cultural district: The case of the Langhe, Piedmont	1115
5. Policy issues for cultural districts in developed and developing countries	1116
6. Conclusion	1118
References	1118

\* An earlier version of this chapter was presented at the 11th International Conference on Cultural Economics in Minneapolis, USA, 2000. Support for this project was provided by PRIN 2002–2004, Italian Ministry of University and Research. This chapter develops many ideas presented in Santagata (2002). Special acknowledgments to Tiziana Cuccia, whose research on the Caltagirone cultural district [Cuccia and Santagata (2004)] extensively illustrate one case study discussed in the chapter, to David Throsby for his valuable encouragement and to two referees, Marcel Canoy and Alan Peacock, for their stimulating comments.

*Handbook of the Economics of Art and Culture, Volume 1*

*Edited by Victor A. Ginsburgh and David Throsby*

*Copyright © 2006 Elsevier B.V. All rights reserved*

*DOI: 10.1016/S1574-0676(06)01031-3*

**Abstract**

The aim of this chapter is to analyze the economic properties and the institutions governing the start-up and evolution of cultural districts. Cultural districts are a good example of economic development based on localized firms and local culture. The first part of the chapter (Sections 1–2) reviews the relationships between culture, viewed as an idiosyncratic good, and the Marshallian theory of industrial districts. Sections 3 and 4 of the paper present a discussion of two models of cultural districts: the *industrial cultural district* (mainly based on positive externalities, localized culture, and traditions in “arts and crafts”), and the *institutional cultural district* (mainly based on property rights assignment and symbolic values). Policy issues are analyzed in Section 5, with particular reference to the applicability of the two models of cultural districts in developed and developing countries.

**Keywords**

cultural districts, intellectual property rights, economic development, material culture

*JEL classification:* Z1, K, O1

## 1. Introduction

Culture-based goods and services are identified by Scott (2000, p. 3) as being “marketable outputs whose competitive qualities depend on the fact that they function at least in part as personal ornaments, modes of social display, forms of entertainment and distraction, or sources of information and self-awareness, i.e. as artifacts whose symbolic value to the consumer is high relative to their practical purposes”. Some such goods are produced in specific and identifiable local areas; well-known international examples include Hollywood, the Spoleto Festival, the vineyards of Langhe in Piedmont, the fashion and textile district in Prato near Florence and the pottery district of Caltagirone in Sicily. These areas may be called cultural districts. They occur in both the industrialized and the developing world. In the latter case, there exist many localized agglomerations of micro and small enterprises producing culture-based goods such as jewellery, apparel, textiles, craftware, services of cultural heritage, etc.; these are potential cultural districts whose progress may benefit from the type of economic policies instituted in the well-known international cases.

In theoretical terms, a cultural district arises at the confluence of two phenomena: that of localization, as first identified by Marshall, and that of the peculiar or “idiosyncratic” nature of culture. When these two essential factors are brought together into a receptive economic environment, the basic conditions for a potential cultural district are satisfied.<sup>1</sup>

The aim of this chapter is to analyze the economic properties and institutions governing the start-up and evolution of potential cultural districts. The first section introduces the theory of cultural districts, whose basic elements as noted above are the localized production of culture-based goods and their idiosyncratic cultural nature. The following sections present a discussion of two main models of cultural districts: the *industrial cultural district* (mainly based on positive externalities, localized culture, and traditions in “arts and crafts”) and the *institutional cultural district* (mainly based on property rights assignment, symbolic values and cooperative behavior). Some policy issues relevant to both developing and developed countries are discussed in the final section.

<sup>1</sup> The use of the term “cultural districts” in North America refers mainly to urban concentrations of cultural activity which are often the focus of policies for the revitalization of a city area through the localization of performing arts services, leisure-time industries and cultural industries; these districts are not analyzed in this chapter (see Chapter 30 by Bille and Schulze in this volume). Other important examples of cultural districts not analyzed here are those that surround systems of cultural heritage involving networks of museums, temples, and archaeological remains; examples include local museum networks in France, Egyptian tombs (Old Pyramids, Saqqara tombs) or Khmer temples in Cambodia [Santagata (2002)].

## 2. The theory of cultural districts

### 2.1. Marshallian localized industries and endogenous economic growth

The first component of a theory of cultural districts derives from the work of Alfred Marshall, who was the first to focus attention on localized industry. The basic elements of his theory are outlined in his *Principles of Economics*, book IV, chapter X: “The concentration of specialized industries in particular localities” [Marshall (1890)]<sup>2</sup> and in *Industry and Trade* [Marshall (1919)].

Marshall divides the economies arising from an increase in the scale of production into *external economies*, those “dependent on the general development of the industry” [Marshall (8th ed., 1930, p. 266)], and *internal economies*, those “dependent on the resources of the individual houses of business engaged in it, on their organization and the efficiency of their management” (ibid). This distinction is crucial, because while the internal economies are essential for the development of the great factory, the external economies “can often be secured by the concentration of many small businesses of a similar character in particular localities” (ibid). In other words, industrial districts in Marshall’s analysis are normally based on external economies of agglomeration.

On the production side Marshall lists several advantages of the industrial agglomeration:

- free diffusion of ideas;
- free diffusion of information: “. . . the mysteries of the trade become no mysteries; but are as it were in the air” [Marshall (8th ed., 1930, p. 271)];
- prompt dissemination of innovations in machinery and in processes and in general organization;
- development of subsidiary trades in the surrounding areas;
- reductions in unit costs led by the introduction of highly specialized machinery and increased use of equipment; and
- creation of a constant market for skilled workers.

On the consumption side, Marshall observes that consumers are willing to travel to distant districts if they can find what they want easily and cheaply. Thanks to the relevance of economies of agglomeration, industrial districts are recognized as places where average costs of production and trade are decreasing, just as they are decreasing in the great factory thanks to internal economies of scale. Even if some drawbacks to the concentration of industry are acknowledged, such as the risk of dependence of the local economy on a single industry, Marshall’s original concept of external economies was profound and rich with potential for development.

At the end of the nineteenth century when Marshall was writing, economists were aware of the fact that the evolution of transport and the widening of the labor market would render industrial districts increasingly less competitive. The shift to mass production – whether favored or not by industrial policies of the central government – was one

<sup>2</sup> Quotations in the following paragraphs are from the 8th edition, 1930.



of the factors that ushered in the decline of some old districts such as Sheffield (UK) or Saint Etienne (France). Nevertheless, the crisis of the Fordist model of mass production that took place at the end of the twentieth century led to renewed interest in the industrial district as a good example of flexible organization and endogenous growth based on external agglomeration economies and small creative enterprises [Becattini (1987, 1989); Bagnasco and Sabel (1995); Pyke, Becattini and Sengenber (1990, 1992)].

As can be learnt from recent Italian and international experience, the presence of small firms is essential to a district, but the presence of many small firms does not in itself form a district. One of the most meaningful characteristics of a district is the interdependency of its firms: in this type of “industrial atmosphere”, frequent contact favors the exchange of specialized inputs, and continuous and repeated transactions cause information to circulate. Within the districts it is easier to find contractors to verify the quality of goods and services and to sign standardized contracts. The social habitat of the industrial districts tends to be made up of large families and of firms where the whole household are employed. Tacit knowledge, mutual trust and the accumulation of social capital are other pervasive traits of local society and culture. In this sense the Marshallian notion of “industrial atmosphere” [Marshall (1919)] implies an interdisciplinary approach involving economic, social and cultural factors. Moreover, when most of the economic and human resources are local, the economic process becomes endogenous. Of course, such evolution requires continual adaptation, but the strong advantages in providing technological innovation, sharing information, differentiating products, regulating the market and fostering cultural links have been shown to be reliable guarantees of sustainable growth [Piore and Sabel (1984)].

Thus it can be suggested that Marshall’s theory of localized industry can be given a new lease of life in contemporary economic analysis. In particular, we can apply the theory to the production of some culture-based goods which create a path to regional economic development through the growth of micro and small-sized firms which are intensely integrated within the territory and in the local community. In this sense Italian industrial districts, such as those producing glass in Murano (Venice) or woven goods and apparel in Prato or Biella, constitute an ideal model for the localized production of culture-based goods.

## 2.2. Culture-based goods as “idiosyncratic”

The second essential component of a theory of cultural districts derives from the nature of culture as a particular phenomenon, endowing culture-based goods with peculiar or “idiosyncratic” characteristics. Several such characteristics can be identified in the specific context of localized production.

First, some local cultures have originated from longstanding social and institutional structures within communities, with their specific content (reflected, for example, in the types of culture-based goods produced) often depending on local resources both tangible (mines, clay grounds, climate, water, etc.) and intangible (patronage of a court, universities, cultural centers, monasteries, etc.) [Marshall (1890); Harris (1977)]. In

other words, the local presence of appropriate natural and institutional resources together with given social structures may be at the origin of an idiosyncratic culture, peculiar to a community and transmissible between generations.

Second, culture has two profound anthropological roots: time and space. The production of a culture is indissolubly linked to a place, or in a social sense to a community and its history. As a result, certain culture-based goods have both space-specific and time-specific characteristics. Note that the more time-specific and space-specific a commodity is, the less the market mechanism is able to efficiently regulate its production and consumption. The more specialized and peculiar a good is, the less likely it is that competitive equilibrium will be optimal [Salais and Storper (1993)].

Third, although UNESCO conventions distinguish between tangible and intangible heritage, it remains a fact that so-called “material cultural heritage”, in which the intangible essence cannot be disentangled from the tangible, is not regarded as a specific and autonomous concept [Jalil Moreno, Santagata and Tabassum (2005)]. Yet some culture-based goods and services (functional objects, functional artifacts of material culture) supplied by industrial cultural districts and local communities are made up of both tangible and intangible factors, and this co-existence and union of traditional knowledge, skills and corporeality are their very substance. Examples of such production drawn from the developing world include the following:

- In Sigchos, Ecuador, a number of craftsmen produce pottery, woven goods and clothing reproducing old forms, designs and traditional colors.
- In Alepp, Syria, micro and small firms gathered in small areas of the town produce green olive soap according to a three-thousand-year-old tradition. The technology employed is ancient, the product highly standardized.
- In Lucknow, a city in Uttar Pradesh, many micro firms and creative traditional artisans, mostly women, develop the art of chikan embroidery. The pastel muslins traced by gossamer embroidery look like the ethereal raiments of fairies. “If you wear chikan, you are wearing history.”

What do these cases have in common? They all exhibit the existence of a local culture, which is the source of identity and creativity. In each case, there is a strong link between present-day production and the culture of the place.

A fourth idiosyncratic characteristic of localized cultural production has to do with creativity – its generational aspects and its role in capital accumulation [Santagata (2004)]. In one sense creativity per se is the original and specific product of a generation. In painting, in industrial design, in the motion picture industry, in the fashion market, even in the forms and decorations of pottery objects, waves of creativity marked by a generation period may be identified. The generation waves are visible in the idiosyncratic character of culture-based goods: the couple time/space has a formidable strength in creating the image and the reputation of a single generation. What was the essential nature of the industrial design produced in Milan during the 1950s? It is not re-producible tout court in the year 2005, because the ideas, the culture, the communication technologies, the marketing practices and the style have all changed. Thus, every generation has its own identity, its own pace, and its own creative vein. Fur-

thermore, culture-based goods translate creativity into valuable economic goods and services, providing competitive advantage through the accumulation of cultural capital. Cultural capital is “. . . an asset that embodies, stores and provides cultural value in addition to whatever economic value it may possess” [Throsby (2001, p. 46)]; it gives rise to a flow of services, consumed or used along with other inputs to produce further goods, having both an economic and a cultural value. Through these means creativity lends a further idiosyncratic characteristic to culture-based goods and services.

Finally, culture-based goods are idiosyncratic because tacit knowledge [Polanyi (1958); Polanyi and Prosch (1975)] is needed for their creation, production, distribution and also because personal knowledge relies on past personal idiosyncratic experience. Individual ability, tastes, lifestyles, social institutions and industrial organizations cannot be learned by mechanical transmission: personal and collective stories count.

### 3. Industrial cultural districts and clusters

When localized industries and idiosyncratic cultures as discussed above are brought together, the minimal conditions for an industrial cultural district are met. In this section, the main traits of an industrial cultural district will be examined first in theory, and then in practice with reference to two specific examples: the motion picture complex in Los Angeles and the pottery district in Caltagirone (Sicily).

#### 3.1. In theory

The principal features of an industrial cultural district can be inferred from the formula that led to the international success of the micro, small, and medium sized enterprises of the “Third Italy” in the 1960s and 1970s. These industrial cultural districts belong to the endogenous growth models based on the presence of small firms [Becattini (1989); Storper and Harrison (1991)] and the existence of basic social and cultural conditions [Bagnasco (1988)], and include a number of districts producing culture-based goods such as Sassuolo (decorated and designed tiles and pottery), Biella and Prato (prestigious wool clothes and apparel), Vicenza and Arezzo (jewellery) and Belluno (eyeglass frames). The characteristic preconditions for building cultural districts such as these are the following:

- a local community that is cohesive in its cultural traditions and in the accumulation of technical knowledge and social capital (trust and cooperation);
- a dense interaction between the cities and the surrounding country, the source of most of the active manpower working in the district’s firms;
- a significant development of increasing returns to scale and increasing returns to scope;
- accumulation of savings and the presence of strongly entrepreneurial cooperative local banking;

- a bent towards open international markets;
- public financial support along the entire chain of the creation of value;
- a high rate of birth of new firms, often of household size, as a result of social capability and interactive learning; and
- the ability to be district- and cluster-minded and to produce positive externalities in the field of design, technological innovation, managerial organization, the creation of new products, labor market flexibility and commercial distribution.

In economic terms, these characteristics mean that within an industrial cultural district the costs of the use of the market are lower than anywhere else because of the intense creation of positive externalities, tacit knowledge, the high rate of innovation, easy networking and the cost-free diffusion of information. In regard to positive externalities, two types can be identified. First is the so-called “atelier effect”; more individuals are trained in the local cultural profession than are required to meet the labor demand of the district, hence creating a resource for new entrepreneurial initiatives. Second, industrial cultural districts accelerate the rate of birth of new creative products and new processes of product differentiation. As a result, buyers are likely to enter commercial centers where they can find a vast amount of product. For instance, in the textile and apparel district of Carpi (Italy), 700 firms employ about 18,000 workers. They are therefore able to present about 100,000 models in their pattern-books each season.

Turning to questions of technology, we note that two varieties of cultural districts may be defined according to the nature of the technology involved. On the one hand, sometimes the technology is highly-developed and sophisticated, as in the textile or fashion industries. The more essential technology is to a firm, the longer or more developed the value chain. The production process is articulated into many hierarchically connected phases, and the *final firm* has the important role of organizing the whole production process. A final firm is the last element of the value chain, and is strategically significant because of its proximity to consumer preferences and choices. In this context the final firm may have incentives to buy the contractors' firms and to integrate vertically. The trend is toward a reduction in the number of firms and toward a change in the district's industrial structure.

On the other hand, some cultural districts rely on quite elementary technology; glass, jewellery or pottery production are examples. The less significant the technology, as is the case in pottery, rugs, glass, soap and perfume production, the shorter the value chain and the less the final firm per se counts. In this case each firm organizes its production and competes in sales with other local firms. Cooperation with other firms located at other points in the chain is more concentrated. Creativity thus counts more. In this context structural evolution is less evident. The growth of a firm is more strongly tied to waves of creativity, to its ability to develop collective intellectual property rights, and to expansion toward collective distributive structures. In such an environment the number of small and medium sized enterprises may increase.

### 3.2. In practice: The Los Angeles motion picture complex

The motion picture industry in the Los Angeles County is a special case of industrial cultural districts internalized in a great metropolis [Garreau (1992); Scott (2000, 2005)]. Its relevance is acknowledged worldwide "... because unlike many other case study industrial districts (Silicon Valley, Orange County, or Boston's Route 128, for example), its outputs trade on a pure cognitive register. For this reason alone, Hollywood is one of the most arresting examples of the burgeoning cultural products agglomerations that are on the rise all over the world today" [Scott (2002, p. 972)]. Furthermore, Hollywood movie production [Scott (2005); Chapter 19 by de Vany in this volume] has become a remarkable example of a well established and flexible network economy, emerging from the crisis of the studio system based on the old Fordist production processes [Christopherson and Storper (1986); Storper (1989); Scott (2005)]. Vertical disintegration and product innovation have led to the need for specialized inputs served at a relative low level of product standardization by a cluster of creative and dynamic small-sized firms (see Table 1).

The movie industry in Los Angeles also illustrates the two types of positive externality noted in the previous section. First, the atelier effect is evidenced by the fact that movie production requires the cooperation of a variety of professionals, including designers, painters, writers, photographers, craftsmen, stylists, musicians, multimedia experts, artists, wardrobe designers, tour managers and Shumpeterian entrepreneurs. When all these skills are mobilized, the sector becomes the source of a cohort of highly qualified workers. Given that around 60 percent of the labor force is freelance, an immense workshop results in which every worker is the potential creator of a new firm. Second, in addition to its ability to train new workers, the motion picture industry creates new tie-in and licensed products: books, videos and paraphernalia, thus opening a large merchandising market. This outcome is an example of the "creative product differentiation" externality.

The motion picture industrial complex in Hollywood could be perceived as a mature industry, yet the rate of birth of firms there is remarkable for such an old district: 188 new establishments appeared each year between 1993 and 1997 (with a yearly growth rate of 16.6 percent). This confirms the endogenous dynamic of this cultural district as

Table 1  
Los Angeles motion picture and video industry

Year	Number of employees	% of total employees	Number of establishments	% of total establishments	Number of establishments 1-19	% of total motion picture establishments
1993	183,783	5.2	5634	2.6	N/A	N/A
1997	188,465	5.3	6573	3.0	5984	91.0

Source: Bureau of the Census, 1998, *County Business Patterns for Los Angeles, CA*.

well as its potential capability for allowing the cohabitation of a few major studios with a large number of small sized independent firms. This is a peculiar characteristic of the Hollywood district and anticipates in practice the expected evolution of other typical clusters of culture-based firms.

### *3.3. In practice: The Caltagirone (Sicily) pottery district*

Caltagirone has been renowned since ancient times for its pleasing pottery, the elegance of its traditional design and the creativity of its craftsmen. Earth, fire and creativity, the three basic inputs of pottery production, have been furnished from time immemorial by the mines of clay, the woods and the cultural spirit enveloping the town. The pottery district of Caltagirone represents an example, among many others, of an industrial cultural district using very low technologies [Cuccia and Santagata (2004)].

This cultural district is small. It holds no more than 150 studios and laboratories, each of which corresponds in dimension to an extended family. The average size of each studio is small: not more than 180 square meters. The local community is intensely linked to the pottery production. In Caltagirone there is an historical Museum of Pottery, a regional training school for pottery handicraft, and a strong tourism industry. Two guilds organize the institutional links with the local authorities, important channels for public subsidies and for the forthcoming institution of a collective trademark.

The main structural differences with the Los Angeles case concern three aspects: the intensity of the technical division of labor, the rate of technological innovation and the range of products. The division of labor in Caltagirone is weak and the growth of technological innovation is stagnant. In fact, when the structure of production of a cultural good is elementary, such as with pottery or glass, the whole filière is performed in a number of small laboratories under the direction of lead craftsmen whose skill, flexibility and creativity are the fundamental inputs of firms which sell the same limited line of objects. Consequently, in this kind of industrial cultural district many small production units share consistent positive externalities and social practices developed within the district, but they are strongly competitive with each other. Moreover, each unit tends towards full internal integration of the creative, productive and distributive phases of the value chain. In particular, each studio tends to have a showroom in town or abroad, and a virtual site on the Internet. Finally, given the relatively limited division of labor, this kind of industrial cultural district does not develop specialized branches or new industrial activities. This in turn makes the industry more vulnerable to unexpected shocks on the demand side.

Nevertheless, two new factors seem to be affecting this scenario. Firstly, the accumulation of reputation may induce the localization on the site of new industrial activities from abroad such as industrial tile and pottery production, whose firms are interested in associating the renowned name of Caltagirone to their trademark. Secondly, web offerings – for example, the designing of a website – are usually created cooperatively, thus sharing positive externalities among a wide group of firms.

#### 4. Institutional cultural districts

The second type of cultural district to be discussed in this chapter is the *institutional cultural district*. Its essential characteristic is its grounding in formal institutions that allocate property rights and trademarks to a restricted area of production. These rights take on the meaning of *community or collective property rights*. In this sense, they legally protect the cultural capital of a community in a given area. Their protection concerns the intellectual and intangible components of the culture embedded in the goods and services produced. These rights are normally established through the introduction of a collective geographical trademark that only the local producers can exploit. Following the definition of the World Intellectual Property Organization [WIPO and ITC (2003)] we will refer to them as *appellation of origin* and *geographical indication*.

The content of the goods produced in these districts is strictly connected to the local civilization and its *savoir vivre*. Furthermore, the economic advancement of these products is naturally correlated with the local culture: the more their image and symbolic icon is identified with local customs and cultural behaviors, the more they seduce consumers in a cultural lock-in and the more their production is fostered. In this case, the importance of culture is all-inclusive, mobilizing the aesthetic, technological, anthropological and historical content of the district.

Once more, the Italian experience is emblematic. In regions such as Piedmont–Langhe and Tuscany–Chianti, the economic growth of well-circumscribed areas shows one decisive factor: the approval of legislation assigning collective property rights to the products of material culture, namely of local tradition in the eno-gastronomic sector. However, as we will see analyzing the case of the wine cultural district of the Piedmont–Langhe, to trace a positive clear-cut correlation between the establishment of intellectual property rights and the economic success of a localized industry is complex and non linear, due to the co-evolution of many causes. Nevertheless a strong indirect test exists: in recent years a host of other potential or historic districts have established collective property rights to consolidate a localized economic growth.

In the following sections we consider the institutional setting within which collective property rights are established, propose a behavioral model, and discuss a case-study example of an institutional cultural district.

##### 4.1. Institutional setting: The collective intellectual property rights

When the products are well identified, a trademark is the usual way to protect against piracy and unfair competition; this right, typical of the market for reputation [Landes and Posner (1987)], is individual in its essence. On the other hand, collective property rights are normative rules governing and protecting all the registered producers in a given place; individuals accept a supra-individual authority and the trademark becomes a local public good. So the main reason for establishing it as a collective institution is

similar to that which is at the origin of civil government: there is an increase in efficiency limiting absolute individual freedom and free-riding behavior when this may be done in a consensual setting [Coleman (1993)]. Two types of intellectual property rights may thus be distinguished: individual and firm-based or collective and community-based. Those that are communitarian or collective will be especially helpful for the start-up of a potential cultural district, for example, the *appellation of origin* (AO). This is usually the name of a village or place, and is assigned to a product whose characteristics are deeply rooted in the local social and cultural environment or territory. The entitlement to use the label "AO" offers a means of protecting the traditional practices of a specific place. The AO gives rise to an exclusive right, namely a monopolistic power which is shared among the producers located in the same protected area. Other important collective intellectual property rights are the *collective trademark* and the *geographical indication*.

The assignment of this sort of property right yields a mixed set of incentives and problems. In regard to incentives, these property rights create a monopolistic privilege through product differentiation; at the same time they allow an increase of prices and of yields, contributing to a substantial accumulation of capital, beyond that allowed by the premium price. Furthermore, they generate incentives so that producers invest in reputation and high quality of products selected through a long cultural tradition. They also lead to better control of the productive and distributive process, with an increase in the quality of the products.

However, improving the content and availability of information through labels of origin may have some drawbacks. Here we consider three potential failures.

The first, as public goods theory asserts, is the risk of free-riding behavior. The AO has the function of signaling the quality corresponding at least to the minimum standards required by the rules of registration. Sharing this signal and keeping constant the level of reputation implies a significant collective investment in spreading information and enhancing product quality. Low-quality producers can free-ride in order to get the benefits of the collective trademark, but in the long run, as will be seen below, their non-cooperative behavior will reduce the positive effects deriving from the assignment of the collective right.

The second failure is concerned with moral hazard that may arise if periodic inspections on the quality are not carefully carried out. In this case a local producer could decide to attach the collective sign to products of lower quality, confident that consumers will not appreciate the difference. This strategy will produce a lowering of the average quality of the product until more experience and information eventually reverses consumer behavior.

The third problem is the *exit option* for the better quality producers. If the relative quality of individual products increases, there could be an impulse toward the exit from the collective sign because the use of it would signal a quality that is, over a significant threshold, lower than that of the better quality firms. Thus the process of exit and the cumulative decrease in quality may develop a new reverse move toward adoption of individual rather than collective trademarks.



4.2. Modeling behavior in the establishment of collective intellectual property rights

In this section we investigate the strategies of individual producers with regard to the adoption of a collective trademark [Cuccia and Santagata (2004); Cellini, Cuccia and Santagata (2005)], using artistic ceramics as an example. The generalization to other cultural districts is straightforward.

Let us assume that in the market for pottery there are  $N$  individual firms corresponding to  $N$  craftsmen. It is also assumed that in the market there are two kinds of craftsmen:

- A number  $N^A$  of *art-oriented* craftsmen, producing high quality commodities. They invest in culture and care about the artistic nature of their production.
- A number  $N^M$  of *mass-production-oriented* craftsmen, producing standardized and low-quality/low-priced commodities. They do not care about the art of ceramics, but are only interested in market results. They do not respect the minimum quality standards required of members registered to the collective trademark.

To simplify the interaction among actors, let  $C^A$  stand for the agent representative of the art-oriented craftsmen and  $C^M$  the agent representative of the mass-production-oriented or low-quality craftsmen. Consider a game in which each craftsman can choose between two strategies: “to join” and “not to join” the collective trademark. A lack of interest in the collective trademark and hence its non-existence represents the status quo. We assume that the payoff of the art-oriented craftsman is  $\pi^A$ , while the payoff of the market-oriented craftsman is  $\pi^M$ ; the subscripts 0 and  $T$  ( $\pi_0^A, \pi_0^M, \pi_T^A, \pi_T^M$ ) denote the status quo and the adoption of the collective trademark, respectively.

The payoffs of the players in a static non-cooperative game with simultaneous strategies and perfect information are represented by the payoff matrix shown in Table 2. In the payoff matrix  $\varepsilon > 0$  denotes an externality resulting from the decision of the players about whether or not to join in the collective trademark. If both join in the trademark club,  $C^A$  will bear a negative externality because of the co-presence of the low-quality producer downgrading the collective reputation; at the same time  $C^M$  will receive a positive externality from the presence in the collective trademark of an art-oriented craftsman with excellent reputation. Payoff  $e$  denotes the result for the art-oriented craftsman in the case in which he/she does not join the trademark club while the market-oriented craftsman joins. No a priori assumption is made about the size of  $e$ .

Table 2  
Payoff matrix

To join or not to join a collective trademark		$C^A$	
		To join	Not to join
$C^M$	To join	$(\pi_T^M + \varepsilon), (\pi_T^A - \varepsilon)$	$\pi_T^M, e$
	Not to join	$(\pi_0^M + \varepsilon), \pi_T^A$	$\pi_0^M, \pi_0^A$

If player  $C^A$  decides to join the collective trademark club, his/her payoffs are:  $\pi_T^A - \varepsilon$  if  $C^M$  joins, or  $\pi_T^A$  if player  $C^M$  does not join. With his/her presence  $C^M$  generates therefore a negative externality ( $-\varepsilon$ ), since he/she decreases the average quality level that the consumer attaches to the collective label. If on the other hand  $C^A$  decides not to join, his/her payoffs are:  $\pi_0^A$  if  $C^M$  does not join (this payoff can be interpreted as the normalized level of profit in a market without collective trademark); and  $e$  if  $C^M$  joins.

As far as the player  $C^M$  is concerned, if he/she decides to join, his/her payoffs are:  $\pi_T^M + \varepsilon$  if  $C^A$  joins (in this case,  $C^M$  enjoys at zero personal costs the positive externality ( $\varepsilon$ ) generated from the adhesion of  $C^A$  to the label, since  $C^A$  is supposed to increase the average quality attached to the collective trademark); and  $\pi_T^M$  if  $C^A$  does not join. If  $C^M$  decides not to join, he/she gets a payoff  $\pi_0^M$  if  $C^A$  does not join; and  $\pi_0^M + \varepsilon$  if  $C^A$  joins, since, even if  $C^M$  cannot carry the trademark, he/she may benefit from the positive externalities such as those stemming from the great renown of a ceramic district like Caltagirone.

Considering the payoff matrix, the representative player  $C^M$  has one dominant strategy: *to join* the collective trademark. In fact, for whichever choice of  $C^A$ ,  $C^M$  will judge it convenient to join, that is  $\pi_T^M + \varepsilon$  and  $\pi_T^M$  are respectively greater of  $\pi_0^M + \varepsilon$  and  $\pi_0^M$ . Thus if entry/registering costs are not high, the market-oriented low-quality craftsmen will without doubt join.

As far as the representative player  $C^A$  is concerned, two cases may occur:

- If  $\pi_T^A - \varepsilon \geq e$ , the strategy *to join* is dominant also for him/her and the game has a Nash equilibrium in dominant strategies (*to join–to join*).  $C^M$  increases his/her payoff with respect to the status quo, while the change in the payoff for  $C^A$  will be positive or negative according to whether  $\pi_T^A - \varepsilon$  is larger or smaller than  $\pi_0^A$ .
- If  $\pi_T^A - \varepsilon < e$ , no dominant strategy exists for  $C^A$ , and the game's Nash equilibrium is *to join–not to join*. In this case  $C^M$  benefits from an increase in his/her payoff with respect to the status quo, while  $C^A$  will have a positive or negative change in his/her payoff according to whether  $e$  is larger or smaller than  $\pi_0^A$ .

The above discussion makes it clear that the assignment of collective intellectual property rights such as collective trademarks can work only under specific restrictions.<sup>3</sup> In particular the art-oriented agent is very cautious about joining a collective trademark, and his/her decision depends on the order of magnitude of the negative externality involved in his/her interaction with the market-oriented agent. Local economic policies aimed at increasing the quality of the production of market-oriented agents are the main instrument to reduce that negative externality and induce the art-oriented agent to join the collective trademark. This elementary condition is a prerequisite for the start-up of any institutional cultural district.

<sup>3</sup> For further discussion see Cuccia and Santagata (2004), Cellini, Cuccia and Santagata (2005).

Table 3  
Differentials in GDP deflator and DOC wines prices 1986–1998

Wine DOC	% growth rate of GDP deflator 1986–1992	% growth rate of wine prices: 1986–1992	Differentials: 1985–1992	% growth rate of GDP deflator 1993–1998	% growth rate of wine prices: 1993–1998	Differentials: 1993–1998
Barolo	45.22	86.2	41.0	20.66	146.2	125.54
Barbaresco	45.22	147.29	102.07	20.66	187.6	166.94

Source: Derived from Borrione (2000).

#### 4.3. The institutional cultural district: The case of the Langhe, Piedmont

The Barolo and Barbaresco wines produced in the Langhe district in Piedmont in Northern Italy were granted AO rights in 1992.<sup>4</sup> The basic economic effect of this assignment of property rights was to set off a massive process of capital accumulation. Table 3 illustrates the extensive redistribution of income that followed the assignment and enforcement of these property rights. Comparing the differentials between the GDP deflator and wine prices in the two periods before and after 1992, it can be noted that the price lead of the Barolo DOC over the general national price index was three times greater in the years 1993–1998 than in the years 1985–1992. The same impressive lead holds for Barbaresco, suggesting that during the phase of more effective enforcement of the property rights the economic environment has been at least as good as the natural one (good weather conditions).

Effects of the Langhe cultural district on industrial production processes in the region include high levels of innovation amongst small-scale enterprises, a more professional labor force, improved infrastructure and expanded exports. In cultural terms, the establishment of the district has resulted in enrichment of local quality of life through fairs and festivals linked to local cultural products and traditions, restoration of cultural heritage, use of the landscape as an economic resource, cultural dissemination through museums, cultural centers and wine cellars, development of the tourist industry, etc. To illustrate these cultural effects further, it can be noted that approximately one new cultural event was established per year in the 24 municipalities of the Langhe from 1980 to 1992; from 1992 to 2000, after the promulgation of the DOC Act, the frequency of the events permanently created increased to four/five units per year. This means that about 50 new permanent public performances have been established within eight years. So in spring 2001, for example, more than 100 shows, guided visits to castles, and public performances were available to cultural visitors to the area. Finally, in 2003 the Langhe

<sup>4</sup> DOC Act, February 10, 1992, No. 164. The legislation on wine property rights dates back in Italy since 1924, but the DOC Act No. 194, 1992, introduces for the first time a complete and enforceable protection.

region and Barolo and Barbaresco wines received the maximum score of “100” from the severe and established American review *Wine Spectator*, who titled the winter issue: “Piedmont Perfection”. Never was such a rating assigned before.

Analogous arguments may be made for the institutional cultural districts that are based on the economic exploitation of artistic and popular traditions, such as in the fields of music, arts and craft (glasses, perfumes, soaps), the figurative and plastic arts, and designed goods [Grefe (2002)].

## 5. Policy issues for cultural districts in developed and developing countries

As with all social phenomena, any definition of cultural districts and clusters remains in some sense imperfect, because it is progressively evolving along with the continuous transformations of local society and its economic structure. The two different types of cultural district that we have considered are obviously complementary and compatible. Taken together they show the potential for new paths of economic development. Nevertheless, even though the two models can be found all over the world, it is helpful to introduce a simplification. *Industrial* cultural districts can be considered an economic structure mainly operating in the developed countries, like Italy, France and the United Kingdom. *Institutional* cultural districts can be considered in perspective as an economic formula especially suitable for developing countries.

The logic behind the latter proposition is as follows. The foundation of an industrial cultural district is subject to two constraints: first, it must be superimposed onto an appropriate and adequate socio-economic structure, and second, it is a process of very long and often socially painful incubation. These constraints are likely to militate against the establishment of a traditional industrial cultural district in a developing country. Nevertheless it is not uncommon in developing countries to find clusters of small firms producing culture-based goods within an organizational and industrial frame which is something *less* than a cultural district. There are local economic forces, one or more pioneers, local and external demands, labor skills, learning effects, but what is commonly lacking is an incentive system leading the main economic actors of such a place towards more efficient ways of investing, trading, communicating and marketing their products. In short *good institutions* and *good governance* are lacking. Thus the problem becomes how to convey certain efficient institutions such as intellectual property rights into existing potential industrial clusters. One possible solution is to focus on the function of collective property rights in promoting market-oriented incentives sustaining local development. Thus the institutional model could be seen as the more appropriate strategy for establishing cultural districts in the developing world.

Thus provisional conclusion of this analysis is that there is a *policy design dilemma* concerning the start-up of a cultural district. The diffuse entrepreneurial atmosphere that accompanies the evolution of a cultural industrial district does not have an explicit starting point. Its pattern is historic-evolutionist. Institutional re-engineering is not possible.

In terms of public policy, this means that the sequence of random and unforeseeable events leading to a district's achieving a critical mass is, in practice, irreproducible. The localized social and economic environment cannot be constructed *ex ante*. What can be done, mostly in developing countries, is to influence some of the mechanisms governing a potential cultural district according to rational plans to modify individual economic incentives. This is the case for institutional cultural districts. Here local institutions may be capable of transforming a long and spontaneous process into a real and accomplished economic phenomenon. Collective property rights can serve as a vehicle for setting quality standards, increasing prices, generating income, and raising the rate of competitiveness of the potential district.

In principle the assignment of collective intellectual property rights has a double function. On the one hand, these rights are a safeguard against the illegal copying of a good, idea or logo. On the other hand, they introduce rules, standards and mechanisms for business development to a geographical area and to a community or association of producers; setting standards about the quality of the products implies the development of continuous activities of cooperation, marketing and monitoring among the local artisans and producers. Now while the first function is particularly relevant for developed countries, where the main problem is to safeguard the products of cultural industrial districts from copying, the second function is strategic for the economic start-up of micro and small firms in developing countries where the potential for local agglomerations must become effective.

The two types of cultural districts employ the same economic actors: artisans and entrepreneurs producing culture-based goods and services. Localized micro and small enterprises led by craftsmen and small entrepreneurs are common all over the world. However, due to the higher levels of technology employed, vertical and horizontal integration of firms is typical in developed countries, whereas the agglomeration of many identical micro firms prevails in developing countries. In both cases, due to the essential characteristics described above, the production function of cultural goods and services is influenced by the positive externalities deriving from a common location.

Finally, we draw attention to the evolution of cultural districts over time. We expect this to be positive in many respects: job creation, income generation, improved institutional capacity, international openness and the intergenerational transmission of creativity. However there may also be some drawbacks. As shown in the game model for establishing collective property rights discussed above, the choice of joining to collective property rights depends both on individual incentives and on collective behavior. If joining is possible, exit is possible as well. This means that if intellectual property rights are permanently assigned, the individual adhesion may be temporary, because of the exit option. The main function of a collective property right is to signal the average quality of the collective trademark, so those producers who make higher quality products are induced to take the exit option, those with lower quality stay outside. In other words there is a dynamic in the functioning of these rights which has a bottom line: under a certain quality threshold, signaling of collective intellectual property rights offers no advantage (not even to free-riders) and is thus not chosen. The exit option seems

to be the phenomenon mostly present in developed countries, whereas in developing countries the low-quality scenario may be a deterrent to the assignment of these rights.

In developing countries other possible dynamic disadvantages can be linked to the success of cultural networks. First of all, these countries often lack local enforced regulation on the use of land and landscape. Excess demand creates congestion costs leading to deterioration of tourist attractions: too many visitors, too many extraneous installations – from the construction of hotels to the privatization of public property. Second, external forces such as powerful multinational companies can destroy the local culture through new localizations for production often unrelated to the local culture, unfair competition and exploitation of the local reputation. Globalization may endanger indigenous cultures and has already led to the destruction of some. These cultural invasions can be limited only by developing policy instruments that protect local culture and related economic activities.

## 6. Conclusion

The main aim of this chapter has been to contribute to the economic theory of institutions and culture by providing some analytical suggestions concerning the use of culture as an engine for economic development. This explains the emphasis on the start-up of localized economies. In this sense the distinction between industrial cultural districts and institutional cultural districts serves as a fundamental divide to show how collective intellectual property rights work to assist the evolution of a potential district into an effective one. The institutional cultural district formula is being regarded with increasing interest by developing countries, where many potential cultural districts and networks already exist. International organizations such as the World Trade Organization, UNESCO and the World Intellectual Property Organization are becoming more and more aware of the significant role of collective property rights in assisting and promoting such cultural districts.

## References

- Bagnasco, A. (1988). *La Costruzione Sociale del Mercato*. Il Mulino, Bologna.
- Bagnasco, A., Sabel, C.H. (Eds.) (1995). *Small and Medium-Size Enterprises*. Pinter, London.
- Becattini, G. (Ed.) (1987). *Mercato e Forze Locali: Il Distretto Industriale*. Il Mulino, Bologna.
- Becattini, G. (1989). "Sectors and/or districts: Some remarks on the conceptual foundations of industrial economics". In: Goodman, E., Bamford, J., Saynor, P. (Eds.), *Small Firms and Industrial Districts in Italy*. Routledge, London, pp. 123–135.
- Borrione, P. (2000). "Analisi economica ed istituzionale di un distretto culturale: I diritti di proprietà e il caso delle Langhe". Mimeo, Turin.
- Cellini, R., Cuccia, T., Santagata, W. (2005). "Policy decisions on collective property rights in cultural districts: A positive model". *Rivista di Politica Economica* 95, 1–15.
- Christopherson, S., Storper, M. (1986). "The city as studio, the world as back-lot: The impact of vertical disintegration on the location of the motion-picture industry". *Environment and Planning D: Society and Space* 4, 305–320.

- Coleman, J.S. (1993). "The role of rights in a theory of social action". *Journal of Institutional and Theoretical Economics* 149, 213–232.
- Cuccia, T., Santagata, W. (2004). "Collective property rights and sustainable development: The case of the pottery cultural district in Caltagirone, Sicily". In: Colombatto, E. (Ed.), *Companion to Property Rights Economics*. Elgar, Cheltenham, pp. 473–488.
- Garreau, J. (1992). *Edge City*. Anchor Books, New York.
- Greffé, X. (2002). *Arts and Artists from an Economic Perspective*. UNESCO and *Economica*, Paris.
- Harris, M. (1977). *Cannibal and Kings. The Origins of Culture*. Random House, New York.
- Jalil Moreno, Y., Santagata, W., Tabassum, A. (2005). "Material cultural heritage, cultural diversity, and sustainable development". Working Papers 2005, Elba Center, University of Turin.
- Landes, W.M., Posner, R.A. (1987). "Trademark law: An economic perspective". *Journal of Law and Economics* 30, 265–309.
- Marshall, A. (1890). *Principles of Economics*. MacMillan, London, eight ed. 1930.
- Marshall, A. (1919). *Industry and Trade*. MacMillan, London.
- Piore, M., Sabel, C. (1984). *The Second Industrial Divide: Possibilities for Prosperity*. Basic Books, New York.
- Polanyi, M. (1958). *Personal Knowledge*. The University of Chicago Press, Chicago.
- Polanyi, M., Prosch, H. (1975). *Meanings*. The University of Chicago Press, Chicago.
- Pyke, F., Becattini, G., Sengenber, W. (1990). *Industrial Districts and Inter-Firm Co-operation in Italy*. International Institute for Labour Studies, Geneva.
- Pyke, F., Becattini, G., Sengenber, W. (1992). *Industrial Districts and Local Economic Regeneration*. International Institute for Labour Studies, Geneva.
- Salais, R., Storper, M. (1993). *Les Mondes de Production*. Ed. le l'Ecole des Hautes Études en Sciences Sociales, Paris.
- Santagata, W. (2002). "Cultural districts, property rights and sustainable economic growth". *International Journal of Urban and Regional Research* 26, 181–204.
- Santagata, W. (2004). "Creativity, fashion and market behaviour". In: Power, D., Scott, A.J. (Eds.), *Cultural Industries and the Production of Culture*. Routledge, London, pp. 75–90.
- Scott, A.J. (2000). *The Cultural Economy of Cities*. Sage, London.
- Scott, A.J. (2002). "A new map of Hollywood: The production and distribution of American motion pictures". *Regional Studies* 36, 957–975.
- Scott, A.J. (2005). *Hollywood. The Place, The Industry*. Princeton University Press, Princeton.
- Storper, M. (1989). "The transition to flexible specialisation in the US film industry: External economies, the division of labour, and the crossing of industrial divides". *Cambridge Journal of Economics* 13, 273–305.
- Storper, M., Harrison, B. (1991). "Flexibility, hierarchy and regional development: The changing structure of industrial production system and their forms of governance in the 1990s". *Research Policy* 20, 407–422.
- Throsby, D. (2001). *Economics and Culture*. Cambridge University Press, Cambridge.
- WIPO, ITC (2003). *Marketing Crafts and Visual Arts: The Role of Intellectual Property*. ITC/WIPO, Geneva.

## THE ARTS AND ECONOMIC POLICY\*

ALAN PEACOCK

*The David Hume Institute, Edinburgh, UK*

*and*

*The Association of Cultural Economics International*

### Contents

Abstract	1124
Keywords	1124
1. Introduction	1125
2. The changing analytical scenario	1125
3. The market as allocator of resources	1129
4. Cultural services and the market	1130
5. The rationale for state support	1132
6. Economics and policy implementation	1135
7. Conclusion: A possible future agenda	1138
References	1139

\* I am particularly grateful to the tolerant discussant of this chapter, Rick van der Ploeg, and, of course, to the editors of this volume for their forbearance.



**Abstract**

Consumer sovereignty is taken to be the aim of policy by which the nature and form of public intervention is identified, both in order to influence the provision of and also the demand for cultural services. Such services are defined by enumeration and include creative arts, performing arts and heritage. In practice, some such services, such as museums and galleries, are publicly provided and paid for, and others, such as artists, theatre, opera and ballet companies and broadcasting receive substantial financial support in the form of grants and tax relief. The interesting question is why such support, coupled with regulatory measures to control the provision and sale of historical artefacts, is found in the arts, whereas in other forms of productive activity, such support is increasingly reduced, as instanced in privatisation measures. The answer seems to lie in scepticism by governments about the ability of consumers to choose for themselves the cultural services that they wish to enjoy and therefore the strong influence of producer interests on government policy exploiting the argument that peer group assessment is the sole guarantee of quality. The consequences of this situation are then explored with reference to both allocative and productive efficiency. Suggestions for policy changes are given seriatim and a view is taken as to how such changes might be influenced by possible future economic trends.

**Keywords**

welfare functions, consumer sovereignty, creative, performing arts and heritage, rationale for state support, allocation and productive efficiency, policy implementation, globalization of culture

*JEL classification:* Z1, HO, H4

## 1. Introduction

This chapter tackles the difficult question of how to formulate government policies concerning the arts and culture and the problems arising in implementing them, using a simple methodology. First of all, an account is given of the nature of the analytical issues that have arisen and how they have changed in the light of developments in economic analysis since the author first considered the subject [see Peacock (1969)]. Secondly, a schematic presentation is given of the relation between the various ‘actors’ in the policy process. Thirdly, the fundamental question as to why government intervention to influence the arts and cultural activities generally is considered in some detail, before, fourthly, showing how economic analysis may aid the process of appraising actual policies.

## 2. The changing analytical scenario

The author’s first attempt at deriving policy criteria and associated policy measures for public intervention in order to support the arts was written when Paretian welfare economics was taken to be the appropriate paradigm for policy appraisal [see Peacock (1969)], although he indicated then that this paradigm had certain limitations. This article is still referred to from time to time, but much has happened in economics to make a reappraisal of this approach in order to take account of both the changes that the author feels required to make in the paradigm itself and in the light of what we now know about cultural markets and government attempts to influence them. Attempts to bring policy into a suitable normative framework frequently do not take these changes sufficiently into account.

The first important change has been the widening of the concept of economic behaviour which is embodied in the standard welfare analysis. This is particularly evident in the case of the individual utility function. Models are now commonly presented which elaborate the arguments in that function in various directions. The function is expanded to take account of the utility derived from spending on goods and services which benefit others, to allow for the satisfaction derived from one’s choice of occupation and from the esteem of one’s peer group. Some brave economists have faced the challenge offered by these hypotheses in order to test empirically their relative significance.<sup>1</sup> In the case of the arts, particular attention has been paid to the analysis of the motivation of creative artists, and, as we shall observe later, this has a bearing on the methods that are likely to be considered effective, if it is decided that they might receive some form of public support.

<sup>1</sup> Cultural economists should note in particular the masterly summary of these issues by their confrère Bruno Frey (1997) and the bearing it has on research into the behaviour of those in charge of public institutions offering cultural services, as in his examination of the motivation of those who administer museums and galleries [see also Frey (2002)].

The authors's original article paid considerable attention to arguments based on the 'publicness' characteristics of the arts and which form the basis for claims for public subsidy. The second major change affecting these arguments concerns the growing problem about how to retain the assumption of consumer sovereignty as an initial value judgment, when the responsibility for the size and structure of cultural activities fall partly on the government, once it has been agreed that the market, in some sense, fails as the guarantor of optimal provision of cultural goods.

One approach is simply to abandon the assumption altogether. This is most clearly stated in the famous Musgravian discussion of 'merit wants' [see, for example, [Musgrave \(1987\)](#)]. Different definitions of the term abound, but embodied in all of them is the idea that members of society develop preferences in common – and Musgrave quotes the important cultural example of maintenance of historical sites – which means that individuals accept a societal view of what should be done to fulfil them even though it does not fully accord with their individual assessment of their value. When such goods come to be identified, goods and services with publicness characteristics are frequently found to be important examples. In consequence, such goods have to be publicly financed but whether they require to be publicly provided is a separate matter. The interests of consumers are then no longer synonymous with consumer sovereignty and, as Musgrave argues, the latter has to be replaced by some alternative norm.

The idea of merit wants is now firmly embedded in the cultural economic literature though those who accept it, perhaps without full realisation of its implications, often make the proviso that state action is not necessarily to be entirely divorced from the articulation of individual preferences. Those public officials who are able to take unilateral action to supply merit wants cannot afford in practice to regard potential beneficiaries simply as passive reactors. Hence a useful role for the cultural economist is to devise methods for testing such reactions through suitably devised survey methods.<sup>2</sup> However, the merit wants approach raises the question as to the terms and conditions that individuals might wish to attach to the giving up of their right to choose. There are clearly circumstances in which the surrender of consumer sovereignty cannot be implemented without the consent of the governed in a democratic state. Accordingly, an alternative approach to the study of cultural policies is provided by public choice theory in which criteria are laid down for political decisions so that they simulate if not replicate a market system, as in Wicksell's attempt to prescribe the political system which would produce 'approximate unanimity' [for a full analysis of this approach, see [Peacock \(1998a\)](#)]. This approach is the forbear of studies in cultural economics which discuss methods by which the preferences of consumers are not only revealed but are operational in the sense that, within the limitations on obtaining such unanimity, consumers themselves determine the general nature of what governments are to do in their name.<sup>3</sup>

<sup>2</sup> For a comprehensive account of such methods coupled with a useful bibliography, see [Cuccia \(2003\)](#).

<sup>3</sup> For the exploration of the link between voter preferences and decisions governing public intervention to promote the performing arts, see [Schulze and Ursprung \(2000\)](#).

Before examining how far these recent analytical developments affect the presentation of the problems of formulating and implementing cultural policies, attention has to be focused on a third major change in economic analysis. The large segment of resource use controlled by government through its own spending and regulation of private allocation of resources invites the speculation that economic analysis is not suitably adjusted to examine situations where transactions regarding resource use cannot be completed through the market mechanism. It is commonly asserted by arts pundits that the main concern of economics is that of studying markets which turns them into apologists for profit-taking entrepreneurs,<sup>4</sup> thus ruling them out of consideration when it comes to analysing institutions such as museums and galleries, state broadcasting systems etc. which are managed by professionals with high ideals of public service. The economists' riposte is that non-market situations do not produce some moral transformation, though they create a different set of conditions under which individuals maximise their utility, notably the instruction to maximize profits. Moreover, in cases where public services financed out of taxation are monopolies, they allow the possibility of discretionary behaviour. In other words, where such services entail some form of exchange relationship, standard forms of economic analysis can be employed in order to arrive at conclusions about the outcome of negotiations between the parties<sup>5</sup> with appropriate empirical tests devised.

Examples now abound in the field of the arts and culture where transactions within government and between government and private institutions are involved, particularly where, as in the case of specific subsidies to individual arts companies, some form of contractual relationship has to be instituted. Bargaining theory, coupled with principal-agent analysis have become part of the standard equipment of cultural economists. This last change has altered the emphasis in the policy interests of economists and in their desire to have governments pay attention to their efforts, namely in endeavouring to measure the efficiency of government spending on the arts and how far economists can help to devise control methods which act as a substitute to market forces in promoting efficiency itself.

The implications of this changing scenario for further analysis can be encapsulated in a simple diagram. See [Figure 1](#), lifted from [Peacock \(1997, p. 15\)](#). In the lower part a conventional depiction of the welfare approach would recognize only one decision-making process by government which is some unitary being acting as an 'ethical observer' in order to implement welfare rules (e.g., devised from Paretian criteria) in the perceived interests of the community. Such a process implies the existence of publicness in the provision of goods and services, as is commonly argued in the case of cultural activities. If the amount and composition of cultural goods and services requires such

<sup>4</sup> And not only by them. I recently came across the assertion by Edward Wilson, described by the *London Times* as 'the greatest living writer on science' that economists are congenially placed on the boards of commercial companies, whereas environmentalists are associated with non-profit making foundations [see Wilson (2003)].

<sup>5</sup> For a penetrating analysis of the economics involved in such relationships, see particularly [Ricketts \(2002\)](#).

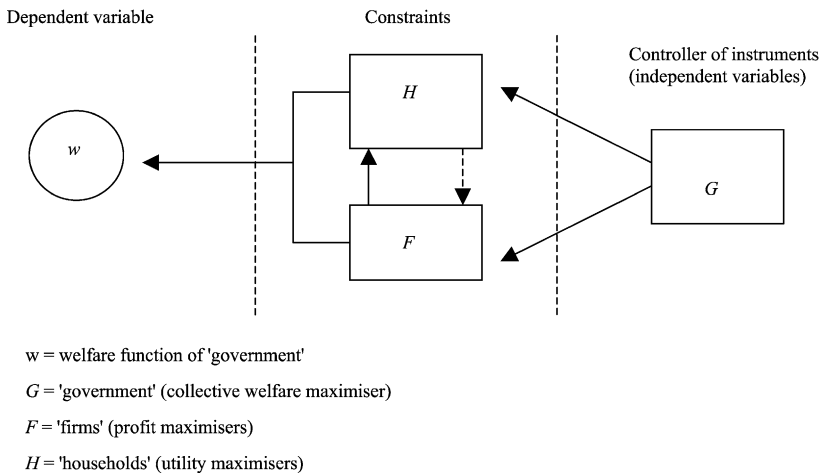
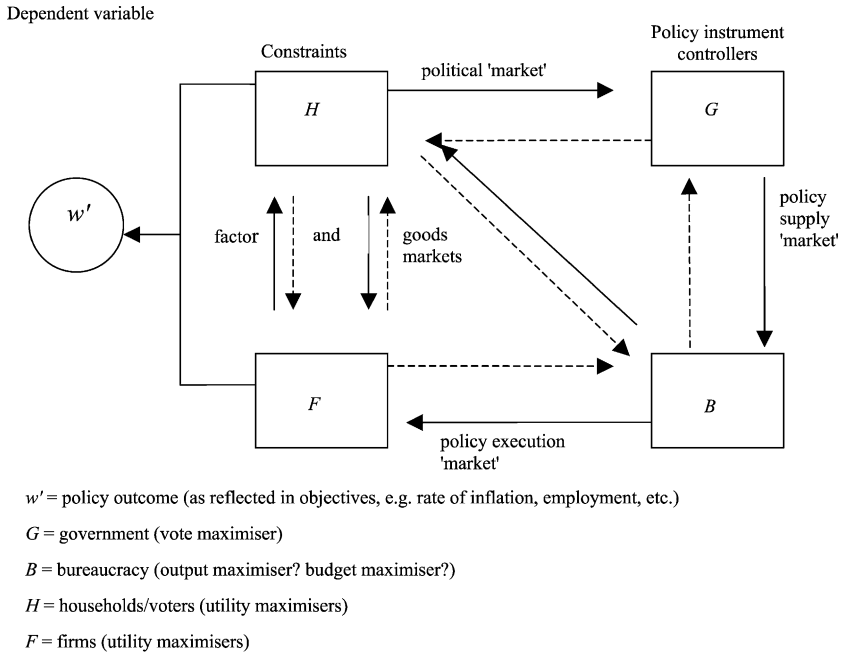


Figure 1. Scope of public choice analysis.

intervention, then the rules commonly require that the tastes and preferences of individuals are taken into account, suggesting, as in the famous Samuelson (1958) presentation of the public goods problem, some method of preference revelation, other than through voting systems. The devising of such methods is clearly of prime concern to cultural economists who accept conventional welfare economics.

Turning to the upper part of the diagram, if it is recognized that preference revelation requires public participation in the decision-making process, then, in addition to a transactions system directly between the ‘unitary being’ and the largely passive adjusting economic units, two other such systems, or ‘markets’, must be embodied in the policy paradigm. These are:

- (i) a political system for election of governments; and
- (ii) a ‘policy supply’ system recognizing the transactions process between politicians and bureaucrats.

This extends the scope of discussion of arts policy beyond defining the extent to which intervention is required on welfare grounds to the possible violation of the welfare criteria by the practicalities of implementing policy. The question as to whether these are areas in which the economists have anything useful to say, which might be denied by experts in political science and public administration, is easily answered by reference to the rapid development of public choice analysis, though some degree of persuasion still seems necessary in order to extend its scope to the world of culture and the arts. The identification of this ‘transactions chain’ brings to the fore the opportunities for discretionary behaviour and the bargaining nature of transactions between government and bureaucracy and the latter with ‘clients’ such as performing arts of companies giving ample scope for such fashionable areas of economics as principal-agent analysis.<sup>6</sup>

In Section 2 below, the policy model commonly devised to indicate the kind of policy measures used to regulate the economy is presented. This raises the question why the arts should be treated differently from other sectors of the economy, but within broadly the same normative framework – Section 3. Sections 4 and 5 contrast the policy conclusions drawn in Section 3 with actual practice. Section 6 speculates on the future role of economists in cultural policies. A short hortatory conclusion follows.

### **3. The market as allocator of resources**

Mixed economies with governments supporting the arts display definite principles of policy action in regard to intervention with the markets for goods and services. Inevitably a short account of such principles is superficial but a necessary prelude to

<sup>6</sup> The diagram might be developed, but perhaps become unnecessarily complicated, by the introduction of other instruments of political participation well known to observers of the arts field, such as pressure groups, including trade unions, NGOs and suchlike organisations who lobby not only governments but also bureaucracies. Some of these seek some special status in advising on policy matters. Much more might be said about the design of the system of political participation, in keeping with the normative rules commonly accepted in Constitutional Economics. Passing mention of some of these complications appear later in this discourse. The hatched lines are added as reminders of feedback effects which emanate from such complications.

consider why the arts should be treated in an exceptional way. The basic principle is that of Adam Smith that the sole object of production is consumption. The consumer's interest is paramount and the consumer is the best judge of his own interests – though this does not preclude consumers, but with their consent, assigning to others choices to be made on their behalf. In a free market system, the consumer can exercise some control through direct payment for such assignment. As we shall observe later, with the provision of cultural services which are not priced, notably those provided by museums and galleries, such control can only be, at best, indirect.

This principle is best invoked by choices made through markets which are characterised by competition. A primary element in government policy is to foster competition. How this is best done is an open question, but the tendency has been in recent times to place emphasis on the competitive process as a dynamic one in which the state's function is to remove barriers to entry into markets. A particular feature of current thinking is the recognition that the government itself may be a barrier to entry, leading to wide discussion of the merits of privatisation of nationalised concerns now subjected to competition, and such measures as competitive bidding for government contracts.

A policy implication of importance for cultural services is that positive intervention in the form of subsidies or grants to firms would have to be strictly limited. There could be a case made out for government general support for high risk projects embodying innovations where patent protection would not be sufficient or for risk-sharing where export performance is affected by political uncertainties. Otherwise, employment and production subsidies to specific firms should be frowned upon, as they discriminate between actual or potential competitors and may cause X inefficiency. Both effects are against the interest of consumers.

This philosophy of economic policy may also accommodate recognition of situations where not all costs incurred by firms are borne by them and not all benefits that they provide are paid for by their customers. The recognition of externalities of production and consumption may make some kind of a case for discriminatory taxes and subsidies, but one which must be compared in actual situations with alternative courses of action such as voluntary negotiation between the parties concerned and regulatory measures. The chosen course of action has to minimize reduction in the operation of market forces as the major guide to the allocation of resources.

#### **4. Cultural services and the market**

There is a marked contrast between the common relationship between government and private firms and that currently found between government and producers of cultural goods and services. It is this contrast which gives rise to a discussion of the rationale of the extensive financial help given to cultural organizations.

First, the creation of cultural works. Painters and composers are typically self-employed. They may co-operate with each other in common facilities, such as ateliers or music studios, and in protection of their rights in the sale and hire of their works, as

in the case of composers' performance rights and the promotion of *droit de suite* provisions for painters. However, it is rare that they attempt to achieve economies of scale or scope by joint production of single works, for independent control and therefore identification with a particular work is necessary for the creative artist concerned with his reputation with his peers. This consideration has a bearing on how the artist lives. The emphasis in her utility function on satisfying herself and her peers may be so strong that she will only take other paid employment as a last resort. Reliance has to be placed on outside support through patronage, long-term investment by galleries or music publishers, or, given the present concern, grants from public bodies. Success in achieving this kind of support is limited because, as is well documented, creative artists do in fact rely heavily on alternative employment, notably teaching.<sup>7</sup>

Second, the performance and presentation of works. The nearest equivalent to the market economy in this area is the commercial art market for the sale of pictures, though once again public bodies may have an influence through their purchase of contemporary art. The position is very different in the case of the performing arts. Using the narrow definition of culture adopted in this contribution, the striking feature is the organization of concerts, opera and drama as non-profit-making ventures. These can take two forms. The first is as private sector companies with charitable status which improves the opportunities of obtaining finance from the public sector either by direct subsidy or indirectly through donations attracting tax relief. The second is publicly operated concerns, such as state orchestras and opera companies, whose affairs are directed by public officials. It is relevant to mention that, in the first case, the orchestras of drama companies may consist of self-employed musicians and actors, whereas in the second, they would tend to be public employees.<sup>8</sup>

Third, the preservation and presentation of historical artefacts. The ownership of the assets regarded as of historical and therefore educational and cultural importance is a very complicated issue. It is sufficient to note that in the case of moveable assets housed in museums and galleries, a large proportion are owned by public bodies which do not require that their managers operate commercially. On the contrary, in most countries viewing days are set aside when entry is free. Direct public funding rather than entrance fees then becomes the major source of finance. The growth in interest in the past as a leisure pursuit has had the result in several countries in a remarkable growth in private bodies offering similar services but still often relying on public funding. In the case of historical sites, those owned by the public sector will normally be managed by government departments or agencies which may raise some revenue from charging admittance, but the major source of revenue will be the government purse. The public

<sup>7</sup> See particularly, Benhamou (2003) on labour market analysis of the supply and demand for creative artists, and also the sceptical view of support for artists expressed by Abbing (2003). Both articles contain useful bibliographies. An early attempt to examine the economic position of composers is to be found in Peacock and Weir (1975) who made one of the first known attempts to quantify their alternative sources of income.

<sup>8</sup> Not to forget the important profession of singer. The locus classicus of analysis of the singer's economic position is found in Towse (1993).



interest extends, however, to the preservation and presentation of artefacts which are privately owned, leading to government measures to induce or even force owners to preserve artefacts and to arrange for their public display.

Fourth, the growing influence of the media. Access to information and enlightenment on cultural matters is increasingly dependent on television and radio. There are now few barriers to the commercial operation of broadcasting, given the alternative forms of transmission now available. Many countries, however, try to exercise control over the content of programmes and actively encourage the use of broadcasting as an educational medium for better appreciation of cultural goods, known as 'public service broadcasting'.<sup>9</sup> Various measures, not necessarily mutually exclusive, may be used, including regulation of commercial broadcasting, but the most striking illustration of the setting aside of market forces appears in publicly owned and operated broadcasting services financed by some form of taxation, either from the general government budget or by some form of hypothecated revenue. In addition to the differences in motivation, in managerial structure and in ownership found in cultural services, as compared with market-oriented businesses, the link between customer and supplier in the financing of the arts in many countries is, to say the least, indirect. Of course, purchasers of cultural services do pay for them directly to the extent that suppliers rely on the box-office, turnstile receipts and the like. However, reliance on government grants brings into being a series of institutions beginning with the raising of revenue to finance them through to the allocation of funding to the relevant government department, and the governmental agencies who assess applications for grants and administer their disbursement. The position is further complicated by the use of tax relief as a method of support for cultural activities, requiring rules governing eligibility.

Therefore, alongside the question of the rationale of direct subsidies to cultural services, and their operation as government agencies, there is the further question of the consequences of this indirect method of finance for the efficient provision of the services themselves.

## 5. The rationale for state support<sup>10</sup>

Given the amount and pattern of state support indicated in Section 2 above, its rationale must be derived either from very different value premises from those adopted by economists or because there is particularly strong evidence of market failure. What appears to be the case is that even conventional welfare economics provides only a weak case for state intervention on the scale commonly found in mature economies. However, it

<sup>9</sup> The place of public service broadcasting has been extensively discussed in the UK in relation to the future of the BBC. See [Office of Communications \(2004\)](#) and [Peacock \(2004\)](#).

<sup>10</sup> This section draws on previous contributions of the author, notably [Peacock \(1998b, 2003\)](#).

is important to examine this case in some detail, if only to increase awareness of some of the ambiguities present in the interpretation of the market failure concept.

The argument may be developed as follows:

(i) *Spillover benefits to consumers.* Individual consumers may attach an option value to the arts, even though they personally do not attend arts events or view historical artefacts, notably in the form of the prestige conferred on a country or community from their existence. This argument is given a more subtle twist by assuming inter-dependent utility functions, so that individual satisfaction may be derived from the fact that others, notably their children or friends, may enjoy cultural events. To derive a case for public subsidy from this argument requires a number of assumptions. The first is that recognition of an option value does not automatically result in voluntary payments to prevent under-provision of culture. The second is that preventing under-provision on option value grounds by public expenditure presupposes that, at the margin, other possible ways by which the same end could be achieved are inferior, e.g., prestige value of international sports teams, medical research. The third is that direct subsidies to individual cultural entities is superior to other forms of subsidization.

In common with welfare analysis in general, a position has to be taken about the welfare of future generations who have no say in current decisions which affect their interests. Rather than guess or prescribe the nature of those interests, I prefer to argue that in operational terms, it is the utility that present generations derive from conferring benefits on future generations that is relevant. Put in a rather negative way, I may be uneasy at the thought that future generations, notably my immediate descendants, would disapprove of any actions of mine that would deny them access to art forms that they would enjoy. This unease might be particularly acute in the case of historical artefacts that, once destroyed, cannot be re-created. The amount and composition of artefacts to be preserved and maintained is not likely to accord with what present generations are voluntarily prepared to pay for viewing them and having them preserved, assuming that their ownership does not prevent them from doing so in the first place. This is particularly the case where charging for viewing them is impossible or prohibitively costly, as with a beautiful town square or country castle. Here again, a general case for public support is presented but it does not specify the amount and form of intervention in the market. The way in which the argument is usually developed certainly does not support the common practice by which cultural services, notably heritage, are not only produced by public agencies but are largely financed by direct grants from government, given the economist's own welfare propositions. It may be that economists on pragmatic grounds would support several of the existing practices, without having to concede that cultural policy should be driven by aims which do not accord with the consumer sovereignty. For example, granted the case for preservation of artefacts of cultural importance in order to take account of the welfare of future generations, the common practice of assigning this task to art historians, archaeologists and the like may be widely supported, given the public's problems of acquiring sufficient information on the provenance of such artefacts. At the same time, it must be realised that expert opinion represents no more

than informed judgment.<sup>11</sup> There may also be practical reasons why economists would not question too closely why heritage services should be provided by state and regional governments, even though it would be more in keeping with welfare philosophy if they were owned and operate by private concerns, albeit subject to government regulation. Apart from the fact that public buildings are frequently of historical importance in their own right, there is clear evidence that when private individuals wish to provide future generations benefits in the form of access to works of art, they may prefer to give or bequeath their property to public institutions perceived as more likely to fulfil their wishes. But the argument so far does not support any case for direct subsidies to either public or private concerns offering cultural services.

(ii) *Quality of choices.* Traditional welfare analysis finds it convenient to assume that in any comparisons of states of well-being, tastes and preferences remain constant. This neglects the strong possibility that individuals will derive utility from investing in knowledge which will improve their satisfaction, the arts being given as a prominent example. Hence the paradox that given tastes and preferences may embody a desire to change them. The Classical economists, notably Smith and Hume, argued that such investment was to be encouraged. As Hume put it, the arts “draw off the mind from the hurry or business and interest; cherish reflection; dispose to tranquillity; and produce an agreeable melancholy, which, of all dispositions of the mind, is the best suited to love and friendship” [Hume (1742)]. However, Hume at least did not believe that investment in taste would be open to all to enjoy, and it is only in recent times that the externalities derived from improvements in the quality of choices have been discerned as positive. The ‘soothing of the savage breast’ by exposure to the arts has been claimed to divert youth away from the excitements of criminality. Ultimately, the question is whether investment in quality of choices requires public support. The strongest argument for doing so is that individuals, particularly the young, under-estimate the benefits from such investment because it is only in retrospect that its benefits become apparent. The general argument may be accepted, but noticeably absent from its presentation is how much and what form state support should take. It implies some form of tied grant to individuals, such as a voucher system<sup>12</sup> recognizing that individuals should be free to choose the shape of the investment package. It does not offer support for subsidies to specific forms of cultural production.

(iii) *Spillover benefits to other producers.* It is claimed that expenditure on the creative and performing arts benefits other producers through the creation of a cultural ambience attracting skilled factors of production and, as in the case of tourism, acts as a loss-leader in attracting business from which industries will benefit. One must be clear at the outset about the connection between this spillover effect and the interests of the final consumer. If employment is at a higher level that would otherwise be the case,

<sup>11</sup> Naturally, experts find this difficult to accept. However, the point is conceded by no less an authority than Gombrich (1979), one of the most distinguished art historians of the 20th century.

<sup>12</sup> Experiments with voucher systems indicate important difficulties in implementation, quite apart from bureaucratic resistance to loss of control of the pattern of expenditure. See Peacock (1993).

then consumers as factors of production have more purchasing power – the income effect. If arts output is higher than it would otherwise be without some form of subsidy, then domestic consumers of arts services may benefit from exploitation of economies of scale or from the availability of arts events that otherwise would not take place. There has been much enthusiasm for this argument, and heavily-funded studies of the impact effect of arts expenditure in particular cities or regions. However, within a country, the extra benefits may largely be pecuniary ones, because any attempt by the use of public funding to realise spillover benefits in one area has to take account of the consequential budgetary effects, such as the reduction in overall incomes from the extra taxation or reduction in other forms of government expenditure. Pushing the argument any further would involve the introduction of interpersonal and interregional comparisons of utility. The spillover benefits could be of national importance, say in small economies with a comparatively large tourist industry. However, for any real benefits to arise it would need to be demonstrated that extra inputs of cultural goods would be the most efficient way of producing the desired results. In any case, if the benefits are perceived by producers themselves to be of material consequence, there is a presumption that they might themselves negotiate private agreements to support cultural loss-leaders.

Clearly, there is little correspondence between the forms of support derived from welfare economics incorporating consumer sovereignty as the main policy objective and the forms that actually exist. It would be wrong to conclude that this is incontrovertible evidence that prevailing systems are paternalistic. The public may be willing to assign to others choices made on their behalf because they believe that there are cadres of experts who have better insight into cultural benefits. However, that is not to concede that this superior knowledge of benefits is anything more than a set of informed value judgments and that those who make them should necessarily be in charge of cultural services in a managerial capacity. Moreover, the scope of the operation of this form of trusteeship will vary from country to country, for it is not to be assumed that public support is perceived as being the sole source of revenue even of publicly owned institutions.<sup>13</sup>

## **6. Economics and policy implementation**

The important presupposition in much of public choice analysis is that any attempt to rectify market failure may be frustrated by government failure. So far as the allocation of resources is concerned, interest attaches to the means available for the beneficiaries from public expenditure on culture to express their preferences through political mechanisms. The associated issue of considerable importance is how far the policies actually decided upon are carried out efficiently, meaning the minimization of resource use for specified levels of performance. Both matters have been extensively investigated by economists.

<sup>13</sup> For further pursuit of these questions by economists, see Johnson and Thomas (1998). The recognition of economic issues by museum directors is rare, but see Thomson (2002).

The first issue raises the question of the efficiency with which far any political system can fully express the preferences of voters. In practice one is unlikely to find, given the relatively small proportion of public budgets spent on arts, that there is strong and persistent political pressure to make it an issue which will decide the fate of national governments. However, the Swiss case is interesting for, occasionally, referenda on cultural matters have been demanded, accepted and voted upon within the cantons, this being a case where the degree of devolution of political decision-making facilities more informed debate about the issues and encourages political participation.<sup>14</sup>

Any democratic government has at least to pay lip service to public reactions to arts policies conducted in its name. The degree to which substitutes for preference revelation leading to *government* reactions are prevalent varies enormously. At the very least, public expenditure projections and audited accounts will reveal information, normally of a fairly general character, on arts financing. Commonwealth countries follow the practice of appointing trustees or some such persons who are not politicians or public officials to boards of management of state museums and galleries, but they are there to put their expertise at the disposal of the public but not as their representatives. Some protection against such appointments becoming sinecures of arts pressure groups is afforded by limiting the period of tenure. However, none of these arrangements implies that governments use an exact measure of the extent to which publicly-funded private or state providers of cultural services are implementing some careful calculation of the removal of market failure as the economist would understand this term.

The economist's skills in examining the allocation issue reveal the logical contradictions that beset arts policies, which is not exactly a way of incurring popularity with public authorities. Once policies are laid down, both in respect of principle and practical implementation, her role becomes more acceptable or at least tolerated. This is because what she has on offer is in tune with public expenditure controls in general in which economists have traditionally played a major part as advisers to Ministries of Finance. Ministries responsible for the funding of cultural institutions will be under pressure to prove that targets are clearly identified and expenditure to achieve them are minimized.

The problems of matching policy aims with this objective provides a useful example of the explanatory power of the dynamics of contracting, using principal-agent analysis. An example from arts funding may be useful here. Imagine a public funding body (pfb) for the arts contemplating the funding of a national opera company. Implicitly, such a body must first of all decide on the funding aims expressed in performance indicators, such as a given number of opera performances a year and perhaps a proviso that a given number of individual operas must be staged with one of them a new commission.

The first problem for the pfb will be to draw up a budget for the specified number of years of the 'contract', taking into account the proportion of revenue that the successful bidder is expected to raise from non-public sources. One initial strategy in an attempt to

<sup>14</sup> Cf. Frey (2002) and Schulze and Ursprung (2000).

minimize cost would be to put the contract out to tender, but alternative sources of supply may be precluded by institutional constraints, particularly if some extant national opera company is an incumbent or even run as a public institution, as with national museums and galleries. If, therefore, there is only one 'agent' offering a service, a bargaining situation is present at the outset.

The next stage would be the drawing up of a 'contract'. The devil is then firmly in the detail. The asymmetric information problem arises immediately if the sole source of information is the incumbent company or cultural institution. The more detail in the contract, the more bargaining takes place over such matters as the delivery dates for specific performances, built-in guarantees of quality of production, the contractual arrangements with singers, orchestral players, chorus etc. and administrative staff. There is an obvious trade-off between relying on trust and avoiding detailed specification – which might suit both parties – and political repercussions if rules governing public expenditure estimates are broken or bent in the process.

The logic of such a contractual arrangement requires that a monitoring system is built into the contract. The asymmetric information problem points towards periodic progress reports, and early warning given when there is a departure from its terms. This is not necessarily entirely onerous to the grantee who then has an opportunity to obtain approval for changes in the contract which may be reasonable enough in the light of circumstances beyond her control. A systematic and continuous dialogue between pfb and grantee is implied, with the former represented in some capacity on the board of management.

In commercial relationships, the non-fulfilment of a contract may be the occasion for compensation payable by the agent to the principal, though the initial contract may only specify the means by which compensation may be sought, e.g., by independent arbitration or by court action. This contingency may be covered by the possibility of insurance against loss by both parties. Such a protection for the pfb, and therefore ultimately the taxpayer, may be very limited. This is clearly demonstrated in the case of defining quality of performance within the context of a regime using performance indicators. A watertight definition of quality cannot be specified in the contract, and the way is open to extended legal argument with settlement only possible after the performance themselves have taken place.

This attempt to spell out some of the implications of a dynamic model of subsidizing arts organizations can only be the starting point for varying its assumptions. The author as a one-time chair of an Arts Council has found it useful as a *vade mecum*, but in his experience the 'noise' produced by the need for openness in public support for the arts makes it necessary to take into account further features in the bargaining process. Thus, as in most forms of public grant-giving, dissatisfied players may endeavour to alter the rules of the game, notably by coalitions with those outside the process as with the use of political feedback. Politicians are particularly sensitive to arguments that place them in the position of being regarded as cultural philistines, with the result that expenditure

on the arts receives political attention well beyond what might be expected from what is normally a small proportion of government funding.<sup>15</sup>

## 7. Conclusion: A possible future agenda

The economic environment in which arts companies and Governments that support them changes rapidly and is bound to affect the nature of the economic advice that might be useful in policies. A continuing feature of economic change in major economies has been the combination of fairly steady economic growth in which technological improvements have played a major part. These have affected, and are likely to continue to have a major influence on both the supply of and demand for cultural experiences associated with the arts; but it is the interaction of supply and demand factors which is forcibly demonstrated in the process of change. This remarkable growth in Western and newly industrialised economies has been accompanied by an increased demand for leisure and accordingly in the demand for cultural experiences. This growth has been associated with an equally remarkable increase in longevity, and also better health, so that the concentration in demand is particularly noticeable in older age groups. One would have thought that the accompanying technical changes which has affected arts production, notably rapid information flow and CD and video presentation forecasted to turn our living rooms into theatres and cinemas, would have counteracted any income effects and emptied the living theatres and concert halls, historic sites and museums and galleries. However, such changes seemed to have operated as a form of advertisement for extensive travel both to produce remarkable international mobility in cultural productions, including international exhibitions of paintings and sculpture, and in tourists anxious to enjoy cultural events in their country of origin. The growth in the concentration of artistic and cultural events generally into festivals covering a limited time span have reduced the costs of attendance to those for whom the opportunity cost of leisure may be high, notably well-off business persons. In short, we are viewing a “Globalization of Culture”.

Of course, international trade in cultural activities is nothing new, but its growing intensity means that policy has to be re-orientated. In this respect, it may reasonably be claimed that at least the technical role of the economist as an adviser to governmental bodies is likely to be increased. Already this is apparent in three areas. The first is in the impact of the growing concentration in the media, notably television, in international corporations which calls for advice on its effects on national competition policies and on its implications for public service broadcasting designed to support quality production. The second is closely related to the first, namely the impact of technology on intellectual property rights, including performance rights, which have increased the

<sup>15</sup> I have explored the bargaining framework between grant-giving bodies and performing arts companies in more detail in Peacock (1993, 2004).

demands for international action.<sup>16</sup> The third is the identification of heritage as an important source of income from cultural tourism. This last area has brought to the fore the recognition of historical artefacts as an important capital resource of developing countries. Governments have sought advice from economists on methods of evaluation of its cultural capital stock as a guide to deciding priorities in preservation. Now that the World Bank has decided that 'sustainable' environment embraces artefacts as well as physical resources, aid programmes now may include schemes for conservation. This obvious way in which economists will extend their influence in the world of cultural policy is through the study of the impact of archaeological remains on the ever-growing curiosity of tourists searching for 'pastures new'.<sup>17</sup>

Keynes looked forward to the days when economists would act and be regarded rather like dentists, more concerned with the immediate realities of improving the human condition than with impressing the public with the profundities of their statements about the good life. One suspects that Keynes's hope may be the way that cultural economics will develop in relation to its policy relevance. That will be all to the good, but the author still regards it as essential that economists will retain a watching brief on those who claim that their expertise entitles them to pride of place in policy decisions. If we do not continue to demonstrate that their judgments of value are arbitrary, then we must not be surprised if they continue to invent the economics for themselves.

## References

- Abbing, H. (2003). "Support for artists". In: Towse, R. (Ed.), *Handbook of Cultural Economics*. Edward Elgar, Cheltenham, UK, pp. 437–444.
- Benhamou, F. (2003). "Artists' labour markets". In: Towse, R. (Ed.), *Handbook of Cultural Economics*. Edward Elgar, Cheltenham, UK, pp. 69–76.
- Cuccia, T. (2003). "Contingent valuation". In: Towse, R. (Ed.), *Handbook of Cultural Economics*. Edward Elgar, Cheltenham, UK, pp. 119–130.
- Frey, B. (1997). *Not Just for the Money: An Economic Theory of Personal Motivation*. Edward Elgar, Cheltenham, UK.
- Frey, B. (2002). *Arts and Economics*, second ed. Springer, Berlin. Chapter 8.
- Gombrich, E. (1979). "Art history and the social sciences". In: Gombrich, E. (Ed.), *Ideals and Idols: Essays on Values in History and Art*. Phaidon, London.
- Hume, D. (1742). "Of delicacy in taste and passion". In: Hume, D. (Ed.), *Essays, Moral, Political and Literary*. Liberty Fund, Indianapolis.
- Johnson, P., Thomas, B. (1998). "The economics of museums: A research perspective". *Journal of Cultural Economics* 22, 75–85.
- Krowitz, E. (2003). "The battle for the past". *Culture without Context*, Newsletter of the Illicit Antiquities Research Centre, MacDonald Institute of Archeological Research, Cambridge, UK, pp. 24–35.
- Musgrave, R.A. (1987). "Merit goods". In: *The New Palgrave Dictionary of Economics*, vol. 3. Macmillan, London, pp. 452–453.

<sup>16</sup> For detailed analysis, see Towse (2001).

<sup>17</sup> On the economics of sustainable heritage in an international context, see Throsby (1997). For an interesting debate on the use of markets to control international trade in historical artefacts, see Krowitz (2003).



- Office of Communications (2004). *Ofcom Review of Public Service Television Broadcasting*. Ofcom, London.
- Peacock, A. (1969). "Welfare economics and public subsidies to the arts". *Manchester School of Economic and Social Studies* 37, 323–335.
- Peacock, A. (1993). *Paying the Piper: Culture, Music and Money*. Edinburgh University Press, Edinburgh. Chapter 7.
- Peacock, A. (1997). *Public Choice Analysis in Historical Perspective*. Cambridge University Press, Cambridge, UK, pp. 3–20.
- Peacock, A. (1998a). "Wicksell and public choice". In: Newman, P. (Ed.), *The New Palgrave Dictionary of Economics and Law*. Macmillan, London.
- Peacock, A. (1998b). "Subsidization and promotion of the arts". In: Giersch, H. (Ed.), *Merits and Limits of Markets*. Springer, Berlin, pp. 185–210.
- Peacock, A. (2003). "Public financing of the arts in England". In: Miles, D., Myles, G., Preston, I. (Eds.), *The Economics of Public Spending*. Oxford University Press, Oxford, pp. 341–372.
- Peacock, A. (2004). *Public Service Broadcasting and the Future of the BBC*. Institute of Economic Affairs, London.
- Peacock, A., Weir, R. (1975). *The Composer in the Market Place*. Faber Music, London.
- Ricketts, M. (2002). *The Economics of Business Enterprise*. Edward Elgar, Cheltenham, UK. Chapters 11 and 13.
- Samuelson, P. (1958). "Aspects of public expenditure theories". *Review of Economics and Statistics* 40, 332–338.
- Schulze, G.G., Ursprung, H.W. (2000). "La Donna e Mobile – or is she? Voter preferences and public support for the performing arts". *Public Choice* 102, 131–149.
- Thomson, K. (2002). *Treasures in Store: Museum Collections and Paradoxes*. Oxford University Press, Oxford.
- Throsby, D. (1997). "Seven questions in the economics of cultural heritage". In: Hutter, M., Rizzo, I. (Eds.), *Economic Perspectives on Cultural Heritage*. Macmillan, London, pp. 13–30.
- Towse, R. (1993). *Singers in the Market Place*. Clarendon, Oxford.
- Towse, R. (2001). *Creativity, Incentive and Reward*. Edward Elgar, Cheltenham, UK.
- Wilson, E. (2003). *The Future of Life*. Abacus, London, pp. 24–40.

## CULTURE IN INTERNATIONAL TRADE

KEITH ACHESON

*Department of Economics, Carleton University, Canada*

CHRISTOPHER MAULE

*Department of Economics and Norman Paterson School of International Affairs, Carleton University, Canada*

### Contents

Abstract	1142
Keywords	1143
1. Introduction	1144
2. International cultural economics	1144
3. Guiding principles for international policy	1147
3.1. Differing takes	1147
3.2. Case for protectionism	1148
3.3. Case for openness	1149
3.4. National cultures	1150
4. International dimensions of visual and performance arts	1151
4.1. Mobility of artists and diffusion of creative processes and ideas	1151
4.2. International trade in art and artifacts	1152
4.3. Touring exhibitions, multinational museums and networking	1153
5. International dimensions of the cultural industries	1154
5.1. Production	1154
5.2. Distribution	1155
5.2.1. Content	1155
5.2.2. Uncertainty, information challenges and high promotional budgets	1156
5.2.3. Film and television	1157
5.2.4. Books and magazines	1157
5.2.5. Music	1158
5.3. Exploiting successfully promoted content across media	1159
5.4. Multinationals and antitrust issues	1160
6. International aspects common to the arts and cultural industries	1162
6.1. The new versus the old	1162

6.2. Openness and creativity	1163
6.3. Trade in complementary equipment, goods, and services	1165
6.4. Heritage tourism and festivals	1166
6.5. Versioning content for language markets	1166
7. The governance of international trade and investment	1168
7.1. The GATT, GATS and the WTO	1168
7.2. Other agreements	1170
7.2.1. WIPO	1171
7.2.2. UNESCO	1171
7.2.3. Regional agreements	1172
7.3. An international agreement on cultural diversity	1174
8. Measuring trade	1175
8.1. Measurement problems	1175
8.2. Data sources	1177
9. Conclusions and suggestions for further research	1179
References	1180

## Abstract

Cultural allegiances whether inherited, imposed or chosen, affect economic activity. Many of these cultural layers – ethnic background, religion, language, ideological orientation, and artistic interests – spill over national boundaries. Cultural ideas travel the world along many routes from the Silk Road to modern electronic networks. Historically, peripatetic artists, composers and writers have responded to shifting patronage and market opportunities. More recently, firms in the cultural industries develop and produce content and distribute it as widely as the market will bear. Visual and performing arts and the cultural industries have both common and distinct international economic dimensions. In trade agreements, countries voluntarily limit their policy options in return for restrictions on the choices of the other member countries. Arguments for protection versus openness for cultural activities are more complex and nuanced than for other economic sectors because of a wide range of views on how international cultural policy affects individuals and the national culture. The inclusion of GATS and TRIPS in the WTO made the WTO a more important influence on international cultural policy than its GATT predecessor. UNESCO continues to play a complementary role. The Florence agreement (1950) encourages the free flow of cultural products and a convention addresses illicit trade in cultural property, a heritage issue. Currently, UNESCO is the focus of efforts to create a rules-based convention to protect and promote the diversity of cultural expressions, which is designed to either separate international cultural policy governance from the WTO or strengthen the bargaining position of cultural industry interests in WTO negotiations. These discussions take place in circumstances where there are serious shortcomings in the measurement of trade in cultural goods and services.

**Keywords**

international trade, cultural goods, trade agreements, arts and cultural industries, GATT, GATS, WTO, UNESCO, WIPO

*JEL classification:* Z1, F, F12, F13, F21, F23, L4, L82, L83, D80, D23

## 1. Introduction

The audience for spontaneous moments of creativity, a capella harmonies or a vivid image made in conversation, is multiplied many fold if these moments are recorded and shared. Recording and recalling have developed from drawing on the memories of troubadours and visual artists through the circulation of books, photographs, and sound recordings to the vast capacity of the Internet. In the process, content has evolved from the spontaneous to the crafted. Networks of relationships – market, cooperative, and contractual – coordinate the production and distribution of cultural products and services. Increasingly the networks are international in scope. These developments complement as well as compete with cultural activities that are local and relatively unaffected by technological change.

A possible root of “troubadour” is the Latin *turbare* – to disturb. From the beginning, the transmission of ideas and images not only increased the recipients’ pleasure and knowledge but also disturbed the relations among individuals in the importing societies and between them and their governments. The tension between those who favor stability and those who are more open to outside stimuli continues to make domestic and international governance of the exchange of cultural products contentious subjects of debate. In this chapter, we discuss international cultural economics, give a stylized account of the debate on international cultural policy, describe and analyze the international economic features of the visual and performing arts (theater, opera, ballet, musical concerts, art, etc.) and the cultural industries (publishing, cinema, broadcasting, recorded music, etc.), assess the existing governance of international trade and investment, and conclude with comments on the available statistics.<sup>1</sup>

## 2. International cultural economics

International cultural economics combines something old, international economics, with something new, cultural economics. The first ingredient, international economics, analyses the coordination by the price system of individuals (identified by their unchanging preferences) with non-labor resources and given technological opportunities. This stylized world is partitioned into countries with governments responsible for policies within their borders. Countries are usually distinguished by their factor endowments and degree of technical sophistication and not by their histories, cultures, climates, and ethnic make-up. The basic international trade model has been augmented by refinements borrowed from other areas of economics such as industrial organization or game theory. Within the basic or refined versions, researchers address how the mix of policies adopted by different countries affect trade, capital flows, movement of

<sup>1</sup> Space limits treatment of a broader definition of cultural activities that might include sports, advertising, crafts and cuisine.

people among countries, growth rates of national income, and the welfare of individuals in different countries. At their simplest international trade models are analytically challenging; at their most complex, they appear very simplistic compared to reality.

Governments negotiate and sign international agreements. The range of feasible international policy initiatives is more limited than that of domestic policy because of the absence of a world government with authority to pass international laws and establish a complementary adjudication and enforcement system. The impact of different regional or international agreements is explored in international trade theory by tracing the effect of a change in a small set of parameters, each representing a complex set of measures; for example, a single  $t$  may represent a complicated tariff schedule and set of non-tariff barriers to trade. Some international agreements are declarative of “proper” behavior. The hope is that they encourage adhering governments to conform to the embedded maxims and support the development of sympathetic norms. Traditional international economics has little to say about them. Other agreements, in particular regional and international trade agreements, constrain the policies of their members. They typically rely on a self-contained dispute settlement mechanism (DSM) that adjudicates disputes and can penalize non-compliance by a measured denial of the agreement’s benefits. The viability of trade agreements attests to the disciplining potential of denying the gains from international trade. Their social contentiousness reflects the uneven impact of trade liberalization among groups within each country and “gaming” by member countries to improve their share of the increased wealth. The governance of international trade is less “thick” than that of domestic trade because it depends on non-coercive agreements. This difference means that trade between countries within a free trade zone, for example, differs from trade within a single country.

The second ingredient is cultural economics, the unique characteristics of which are addressed by other chapters in this book. Like international economics, it has borrowed from other areas of economics.<sup>2</sup> Cultural economics also wrestles with issues that other fields of economics have largely ignored.<sup>3</sup> Not surprisingly, it stresses the importance of culture for the development of individuals. Since there is unfortunately no accepted paradigm of how to model this influence, we outline below simply those elements that play a role in our review of culture in international trade.

We assume that all individuals belong to many cultures – groups that have boundaries, sometimes fuzzy, whose members share values and related behavioral norms. Cultures in this broad sense span a spectrum. At one end are those dependent on birth,

<sup>2</sup> Some imported ideas that have been fruitful are: the provision of services with some or all of the attributes of public goods; the characteristics of markets for differentiated goods and services; auction theory; the specific human capital requirements of a cultural career; the economic role of copyright; and the measurement of contingent demand.

<sup>3</sup> Some examples are the evolution of cultural capital – the artifacts and collective memories of the past – and its impact on economic life, the implications for the future economic viability of activities like theater and opera of their limited potential for experiencing cost-reducing technological innovations, and the influence of extreme uncertainty of demand on contracting and organizational practices.

residency or parental choice. At the other end are those loosely held together by common interest or self-appointed membership (computer hackers or people of the “left” or “right”). In between are cultures that accept only a few applicants and others with relatively open memberships. Cultures can be hierarchically or democratically organized, or have no discernible formal organizational structure. Each culture partitions the world differently. From this perspective, nationality is one of an individual’s cultures. It is of particular importance because it determines the laws and regulations governing an individual’s access to other cultures. The borders of other cultural allegiances are usually not geographical but reflective of choices or endowed characteristics of individuals. Each cultural layer establishes psychic or real costs and rewards that interact in a complex way with legal constraints and market prices to shape economic behavior. Many cultures, small or large, are transnational. Their contours affect the direction and “texture”<sup>4</sup> of international exchanges of information as well as of cultural goods and services.

From an analytical perspective, membership in a culture has an impact on demand for cultural goods and services similar to the effect of capital stock on the demand for variable inputs in the short run by the firm. If membership in a culture involves significant investments in capital specific to the culture, hysteresis delays decisions of non-members to join or of members to exit.<sup>5</sup> The hysteresis impact is strengthened if the culture penalizes apostates by, for example, shunning them socially. Variations in the degree of being locked in to different cultures blur the distinction between a short and long run. Another sea anchor on change is the process governing the evolution of codes of behavior in most cultures.

Our view of international cultural economics assumes that a number of cultural partitions join with national cultures to affect individual behavior. *Rushton (1999)* discusses the tension between individualism and communitarianism. Our individuals make decisions constrained not by one association, the community, but an overlapping set of associations, some chosen and others imposed. Individuals in a sense create themselves by their choice of discretionary cultures and by how they reconcile any dissonance among the behavioral prescriptions of their personal allegiances. Some cultural allegiances facilitate reconciliation by valuing the stories, symbols and existence of other cultures. Others preach dominance over competing views and eschew compromise. The social manifestations of how individuals alter their allegiances and provide coherence to their lives are intermediated by governments and their cultural organizations. The valence of both tolerant and uncompromising allegiances is evident in some government’s willful destruction of part of its cultural patrimony and the ensuing reaction by individuals, foundations, and government agencies of other countries.

<sup>4</sup> Cultural categories and beliefs, for example, condition an individual’s view on: what is “fair” in different contexts; obligations to kin; sympathies and antipathies towards the choices and life-style of others; and the appropriate social protocols for interactions with strangers.

<sup>5</sup> See, for example, *Dixit (1992)* and *Dixit and Pindyck (1993)* for particularly clear expositions of a difficult topic, and *Schulze (1999)* for a discussion in the context of trade.

Traditional international economics generally assumes that cultural factors are insignificant in explaining behavior at least in the long run. Methodologically, international cultural economists are searching for disciplined modes of analysis to more effectively address the role of complex cultural processes and activities. International economics has the precision of a recipe but fails to include key ingredients; cultural economics puts a creative process of self-fulfillment at the center but modeling details are works in progress.

### 3. Guiding principles for international policy

#### 3.1. Differing takes

Cultural products and services from other societies have historically been treasured as tributes to the sophistication and cosmopolitanism of those who possessed them, but also feared as threats to the integrity and stability of domestic relationships. Parents would not exert costly efforts to control the radio listening, television watching and reading habits of their children if they did not think that they were important influences on development. With regard to international trade and investment, the tension between the appeal of greater diversity and a fear of its cultural consequences is reflected in a clash between two discourses, one cultural and the other economic or industrial. An early expression of the former is found in Jean-Paul Sartre's magazine *Les Temps Modernes* of 1953.

Cinema is something very different from an industry. It is a means of expression for a collectivity. The image which a country, a society, offer of themselves. That is why a politico-economic offensive which seeks to stifle this means of expression in a given country can only be compared to the forcible means which conquerors sometimes use to deprive the vanquished of their language.<sup>6</sup>

Forty years later during the Uruguay round France's President Mitterand reiterated this position.

Creations of the spirit are not just commodities; the elements of culture are not pure business. Defending the pluralism of works of art and the freedom of the public to choose is a duty. What is at stake is the cultural identity of all our nations. It is the right of all peoples to their own culture. It is the freedom to create and choose our own images. A society which abandons to others the way of showing itself, that is to say the way of presenting itself to itself, is a society enslaved.<sup>7</sup>

<sup>6</sup> Reprinted from Jeancolas (1998, p. 53).

<sup>7</sup> Speech given at Gdansk, Poland, September 21, 1993.



Similar views have been heard from ministers responsible for culture in a number of European countries as well as in Australia, Canada and the developing world. A repeated refrain is that cultural goods and services are unlike those of other industries and should be treated differently in international trade agreements. Currently, the pursuit of cultural diversity is the rallying cry for those supporting this view.

The opposing perspective is that the cultural industries, especially those that offer entertainment to the masses, are similar to other industries and should be made subject to international trade and investment agreements. The Motion Picture Association (MPA) and in particular, Jack Valenti, its leader from the 1960s to 2004, articulated this view with similar rhetorical flourishes:

Is the culture of any European country so flimsily anchored, so shakily rooted, that European viewers must be caged and blinded else their links with their honored and distinguished past suddenly vanish like an exploding star in the heavens? Our market is totally open, absolutely free. There are no restrictions, no barriers, no quotas of any kind. What is in our marketplace, however, is the most fierce kind of competition to win the eyes and ears of those who go (to) the cinema and/or watch TV.<sup>8</sup>

Both sides are largely silent about the high arts. Neither objects in principle to public support for opera, symphonies, theater, literary publishing, museums and art galleries, nor for international agreements that protect intellectual property. The major battle is over mass entertainment that has an articulate industry lobby on each side of the debate in most countries. The US government promotes liberalization while those of most other countries defend different degrees of protectionism. The opposing positions can be summarized as follows.

### 3.2. *Case for protectionism*

The nationalist viewpoint uses the following arguments. A healthy national cultural industry is necessary to assert national sovereignty and identity. Countries with small domestic markets are overwhelmed by imports from larger markets where producers can recoup their costs of production and then dump content abroad. English speaking countries are especially vulnerable to American imports, but similar circumstances apply elsewhere, for example in the case of Spanish language productions from Mexico sold into smaller Central American countries. The actions of American producers in cooperation with the American government, especially in the audiovisual sector, constitute American imperialism and result in the lessening of cultural diversity, as national industries are either unable to compete, or respond in ways that lead to a homogenization of production.

From an economic perspective it is well known that scale economies, imperfect competition and existing distortions elsewhere in the economy can justify protective

<sup>8</sup> Hearings Before the Committee on Finance, US Senate, 102nd Congress, 1st Session, 17–18 April 1991, Washington: USGPO 1991, 150 and 151.

measures. For example, [Francois and van Ypersele \(2002\)](#) have constructed a model in which the introduction of a prohibitive tariff by “France” on international movies raises economic welfare at home and in the “United States”. In the post-tariff world Americans enjoy Hollywood films with lower production values but gain from the production of American independent films that had been “crowded out” by high production value Hollywood films. French movie-goers do not lose from the barring of high production value Hollywood films, which generate no consumer surplus for viewers either in France or the United States, and gain from the consumer surplus rise when their auteur film industry expands.

Arguments about the merits of competitive markets for promoting creative activity are met by the claim that standard economic arguments for trade liberalization do not apply. The audiovisual sector is about culture not commerce, and the interests of culture, especially national culture, should prevail in international trade agreements. In order to counteract the perverse effects of unconstrained commerce, it is desirable to negotiate and implement an international agreement to preserve national cultures in an increasingly economically and politically interdependent world.

### 3.3. Case for openness

Proponents of greater liberalization stress the contribution of open markets to the efficient production and distribution of cultural products. Reliance on market forces also frees creative people from the paternalistic attentions of politicians and bureaucrats. Cultural policies often become subject to regulatory capture and are administered with a view to the interests of producers and government officials. Consumers, like schoolchildren, are told what they should watch and read, but the statistics indicate that few are paying attention to their teachers. If the political process decides that some cultural content is undersupplied, either the state can step in as it does in the case of public broadcasting, or corrective subsidies can be offered. If anti-competitive behavior undermines the efficient working of markets, competition policy should be used.

The “such-and-such-industry-is-different” argument is the rallying cry of all special interest pleading whether its spokesperson, groomed in communication skills, represents agriculture, healthcare, education or national security. Every industry is of course different. By itself the claim provides no guidance to politicians allocating the fiscal pie. In terms of economic characteristics, many industries in a knowledge-based economy have high fixed and low marginal costs but are not granted protection. Many of the special characteristics of cultural products – the non-rivalry of content, economies of scale, and agglomeration economies – enhance the case for integrating markets. That licensing prices for content are lower in small markets than larger is not dumping but its opposite – charging as high a price as different markets will bear. Protection raises rather than lowers prices. The [Francois and van Ypersele \(2002\)](#) strategic trade policy model requires assumptions that are not descriptive of the international film industry: for example, the international industry in their model must be located in the United States, the French cannot invest in the Hollywood monopoly, price discrimination cannot be

practiced, a slight increase in the ticket price of high value Hollywood films in France transforms their dominant share of audiences to a zero share, and no independent film production occurs in the United States in the absence of a French tariff.

All content wherever it is produced contains disparate philosophical and moral messages. Each individual's judgment provides the ultimate defense against being corrupted by the banal or the hate-filled book or program. Governments already act as an agent for individuals and for children by censoring material that offends the current social consensus of decency. Since the nationality of the source does not excuse offensiveness, no discrimination for or against foreign content is warranted on this ground.

### 3.4. *National cultures*

Supporters of protectionism argue that the process of globalization and American cultural imperialism are undermining national cultures. Globalization is an imprecise term. National culture in this context is much more than the laws, institutions and regulations of the country. What constitutes the much more is difficult to define. In short, it is hard to determine what the threat is and what needs protecting from pressures from abroad. National cultures are the result of past forces, many of them international, and they continue to evolve with time. If change is important to sustain viable national cultures then protectionism may impede a desirable outcome. This debate gets tied up with internal conflicts over modernization and societal changes that occur in all countries.<sup>9</sup>

Modernization has taken place as a result of a wide range of international exchanges involving goods, people and ideas. It can be illustrated by tracing the impact of imperial expansion. As well as market exchanges, conflict and tributes imposed on those conquered have provided channels of cultural influence. The empires of Rome, Spain and Britain amongst others both dispersed and absorbed technologies, architectural styles, religious ideas, cultural artifacts, and forms of government. All of the modes of cultural interaction have been criticized. Barber (1995), Gray (1998) and Tunstall (1977) are among those currently arguing that markets destroy cultural diversity.<sup>10</sup> In the opposite camp are Nozick (1974) and Cowen (2002) who contend that market economies provide a widening range of cultural choice.

How do we evaluate these conflicting positions and decide whether there is such a creature as a national culture that can be defined, protected and promoted? In general, creative people are among the first to cross national borders in search of inspiration. Writers, painters and composers, for example, draw on the creative capital of the world to provide them with ideas. Their creativity borrows from the past, reinterprets it, and contributes to diversity over time. We consider their mobility in the next section.

<sup>9</sup> Benedict Anderson (1991) examines the concept of nationhood by tracing the roots of today's nations defined geographically but made up of disparate elements that have coalesced and fragmented over time.

<sup>10</sup> Others who review the issues are Robertson (1992), Schiller (1992) and Tomlinson (1991).

## 4. International dimensions of visual and performance arts

### 4.1. Mobility of artists and diffusion of creative processes and ideas

The high arts have a significant international dimension. From classical times, with a lull in the dark ages, wealthy or politically powerful patrons have sought out and “imported” gifted writers, composers, sculptors and painters. Their experiences varied within and across these cultural activities. For example, in his early twenties Peter Paul Rubens (1577–1640) became the court painter of the Duke of Mantua where he studied the works of Titian and other Italian masters. He visited the court of Philip IV of Spain, affecting the course of Spanish baroque art. On his return to the Spanish Netherlands, he established a highly efficient studio, based on the Italian artists’ workshops that had developed in the earlier Italian Renaissance [Welch (1997)]. Catholicism and reverence of antiquity profoundly influenced his work and his life. An international commission from Maria de’ Medici, the Queen Mother of France, to commemorate her life embroiled Rubens in French politics.

Competition among patrons in the period from the Renaissance to the mid-nineteenth century foreshadowed the current rivalry among teams in international sport. Personal service contracts were the instrument of choice then as now. For example in 1809 Beethoven received an offer to be the Kapellmeister for the King of Westphalia. Three of his Viennese patrons responded by contracting to provide him with a lifetime stipend of 4000 florins to keep him in Austria. When two of the patrons defaulted, Beethoven sued. A settlement making good on the obligations and providing some adjustment for inflation was negotiated [Baumol and Baumol (2002)]. Whereas the religious impact on Rubens’ art is obvious, the cultural influences on music are less obvious; according to Solomon (2003), Freemasonry, which championed Enlightenment and anticlerical ideas, was one of the general influences on Austrian music in general and Beethoven’s in particular.

At least since the Renaissance, the cultural audience has enjoyed “conversations” about the work and lives of internationally recognized “stars” in the traditional arts. Some artists gained notoriety, contemporaneously or retrospectively, for their intense and occasionally destructive life styles. Others like Rubens lent their fame to a social or political cause. Most were not as cosmopolitan as Rubens who wrote in a 1625 letter: “Other things being equal, I regard all the world as my country, and I believe I should be very welcome everywhere” [Magurn (1955, p. 12)]. Nevertheless, almost all stars benefited from, and later became a source of, foreign inspiration, training and on occasion support. They transferred organizational innovations and aesthetic concepts and were subject to transnational religious and philosophical influences. Modern communications and transportation have enlarged the scope of international influence for subsequent artists and expanded the international reach of their predecessors’ reputations.

#### 4.2. *International trade in art and artifacts*

International art dealers sell the works of the artists that they represent from inventory or displayed works that are on consignment. Alternatively they may be commissioned by a buyer to obtain a particular piece or on hearing that something is available shop it to their list of contacts. Their clientele is a mix of galleries, corporate and governmental buyers, and private collectors. Information on the operations of leading dealers is selective and anecdotal. Considerable information, for example, is available on the career of Joseph Duveen (1869–1939) who dominated sales of European art to wealthy American patrons for many decades.<sup>11</sup> The scale of his activities was impressive:

Fifty-five of the hundred and fifteen pictures, exclusive of American portraits, in the Mellon Collection, which is now in the National Gallery in Washington, came to Mellon through Duveen. Of the seven hundred paintings in the Kress Collection, also in the National Gallery, more than a hundred and fifty were supplied by him, and these are the finest [Behrman (1951, pp. 32–33)].

Particularly valuable art and artifacts have been sold through international auction houses and dealers for over two centuries. Electronic bidding and advanced communications have widened the potential audience at auction sales. Initiatives on the Internet have expanded the reach of auctions in distributing existing works. Web sites such as eBay have developed clever protocols for auctioning a wide array of lower-priced art, collectibles, and artifacts. A number of services track transaction prices of art.

Curbing opportunistic manipulation of the identity of the artist of a work and of the provenance of the piece being sold is important for creating a more efficient international market in art. Although a painting is physically unchanged by altering the attribution, its value is significantly altered. In mid 2002, Rubens' *Massacre of the Innocents* was sold for £49,506,650 (US\$76.73 m.), a record price for a painting sold in a British auction. Since the 18th century the painting had been attributed to Jan van den Hoecke, a relatively unknown follower of Rubens. A few weeks before its owner put the painting up for sale, a Sotheby's representative identified it as an early Rubens, which dramatically changed the bidding interest in the painting (and Sotheby's fees).

With respect to the provenance of high-valued items, buyers face a risk of a suit claiming that the item was stolen or had been illegally confiscated by a government at some point in its history. One response to reduce this risk was the development in 1991 of the Art Loss Register, an international database of stolen and lost works of art, manuscripts, books, antiques and other valuable subjects. Currently, 120 thousand items that were lost, stolen or confiscated during the Second World War or subsequently are listed. The Art Loss Register's mission is the creation of an international clearance house in which provenances can be accurately traced.

<sup>11</sup> Among his clients were Hearst, Mellon, Morgan, Rockefeller, Whitney, Huntington, Frick, Carnegie, Altman, Widener, Stotesbury and Kress.

### 4.3. Touring exhibitions, multinational museums and networking

The exhibition of foreign collections of art in other countries is an alternative to cultural tourism. The piecing together of an exhibition celebrating a common theme from a number of museums and private collections and putting it on tour has become popular [Frey and Busenhart (1996, pp. 275–280)]. For example, the organizers of “Aztecs”, which visited major European cities and Tokyo from November 2003 to April 2004 consolidated over 350 pieces from Mexico and from a wide array of international museums. Ironically the exhibition could not have been held in Mexico because the foreign-owned pieces would have been seized under Mexican law, nor could it have proceeded in its full splendor outside Mexico without the permission of the Mexican authorities to allow pieces from its museums to be included. Contributors were compensated for their participation. As part of the arrangements for the London exhibit, for example, the Royal Academy of Arts donated 10 percent of its ticket revenue to the Mexican Heritage Authority.

Highly promoted visiting blockbuster exhibits such as “Aztecs” may “crowd out” collections of host museums. For example, Cannon-Brookes (1996), after acknowledging the impact of temporary visiting exhibits on local pubs and restaurants, notes that they “consume a disproportionate amount of the resources needed by the museum to discharge adequately its core functions” (p. 271). Foreign exhibits are available at a relatively low price to the host museums because the publicity surrounding a special exhibit raises the value of the items lent from private collections and the visibility of contributing institutions [Frey and Busenhart (1996, p. 294)]. The curators of the Smithsonian’s Hirschhorn Museum and Sculpture Garden have recently debated the crowding-out issue and in response mounted Gyroscope, a show mobilizing the museum’s own collection, around three themes of artistic creativity; this is the first show since the museum’s opening in 1974 that features only its own works. The blockbuster visiting exhibits also impact on incentives for regional and idiosyncratic museums to maintain their collections. Similar issues of the survivability of local creativity in the face of international competition arise in other cultural activities.

The internationalization of some large museums allows a more economic exploitation of extensive collections that might otherwise be warehoused. For example New York’s Guggenheim has been particularly active in expanding internationally,<sup>12</sup> although the “walk” has not always measured up to the “talk”. The foreign successes include the Peggy Guggenheim collection in her former home on the Grand Canal in Venice, and a joint venture with the Guggenheim Berlin to display works from its collection and those of German artists at the Deutsche Bank’s headquarters. Since the Guggenheim Bilbao, designed by Frank Gehry, opened in 1997, it has reportedly attracted over a million tourists a year. The “dry holes” included the cancellation of an agreement with the municipal government of Rio de Janeiro to build a multi million-dollar museum

<sup>12</sup> Since 2001, selected works from the Guggenheim Museum collection can be viewed online for free.

designed by French architect Jean Nouvel, and four attempts by the Guggenheim to establish a presence in Japan.

There are also gains from museums with complementary collections cooperating to stock exhibitions. In June 2000, the Guggenheim formed a long-term content alliance with St. Petersburg's Hermitage and Vienna's Kunsthistorisches Museum, which have collections focused on different periods. A less visible but important international linkage for museums and art galleries is the professional networking of administrative and managerial professionals, whereby ideas on technical matters and management are shared informally and through more formal conferences.<sup>13</sup>

## 5. International dimensions of the cultural industries<sup>14</sup>

### 5.1. Production

The international aspects of production in the cultural industries vary according to the technological characteristics of media and the content carried. Writing is a solitary undertaking that is nourished by contacts with other writers, governed by contracts with publishers, disciplined by editors, and complemented by layouts, illustrations, and printing. In contrast, large teams produce film and television programs. Members of a Hollywood film's creative and professional crews are often from different countries; shooting may occur in different venues, and post-production work may be located in still others. Television scheduling imposes tight deadlines and sequential production of episodes, which are consequently less internationalized in their production. In recorded music the organizational challenge of coordinating music, lyrics, and performances and modifying and synchronizing tracks on a master is a less complex process than producing a film. The production teams are small. International diversity is more pronounced among records than within the production of a particular record. Musical influences flow among countries in unpredictable patterns.

The cross-fertilization of ideas in all of the cultural industries is extensive. Movies are based on plays, short stories, novels and newspaper reports. Agents, guilds, industry or activity associations, and trade publications deliver information to those working in cultural activities. This information is interpreted and disseminated originally through relatively thin networks of personal contacts. Internet chat rooms have expanded networking to include strangers from all over the world who share cultural interests. What is called national music is almost always the product of different ingredients brought about by the movement of goods, people and technology. As discussed by Cowen

<sup>13</sup> For example, at the Salzburg Seminar in 1993, Schulz (1994, p. 138) recommended the adoption of budgeting for the German museum system in which the state institutions receive 90 percent plus subsidies to induce their managers "to act as their American counterparts do, in a businesslike fashion".

<sup>14</sup> Studies that treat the economics of the cultural industries include Acheson and Maule (1999a, Part 1), Caves (2000, 2003), Cowen (1998, 2002) and Sedgwick and Pokorny (2004).

(2002, pp. 21–29), this is true for Cuba specifically as well as for Latin America and the Caribbean more generally, with influences from Africa, Europe and latterly North America. In Africa, the popular music of Zaire was influenced before World War II by immigrant workers from different parts of Africa, the West Indies and China. In the postwar period, radio broadcasts and the arrival of cruise ships introduced Cuban influences, while Greek immigrants set up recording studios.

The resulting international sharing creates tensions between creators and receivers when the content is proprietary. The fuzzy boundary between what is protected by copyright and what is in the public domain is constantly being redefined by shifts in norms, copyright law, and technology. The larger the technological shock, the greater the difficulties in resolving how law and behavioral prescriptions should respond. The Internet has been a “multi-sigma” technological shock. Although the adjustment stage has just begun, the courts of many countries are interpreting the relationship of existing law to file sharing and the responsibilities of gatekeepers to police infringement, while legislators are busy amending national laws.

## 5.2. Distribution

### 5.2.1. Content

The low incremental cost of making content available to someone else makes the market for much content extend across borders. This is surprisingly true for many niche markets in the cultural industries.<sup>15</sup> As an illustration of the interplay of international cultural influences on composition and key performers, and of international commercial responses to distribute content with a relatively small market, consider the compact disc *Esteban Salas: Un Barroco Cubano del Siglo XVIII* (Cuban Baroque Music of the 18th Century), performed by the Exaudi Choir of Cuba. Salas (1725–1803), the composer of the music, was born in Cuba and ordained as a priest in the Catholic Church late in his life; the director of the Exaudi choir, Maria Felicia Perez, was trained at the Franz Liszt Music Academy in Germany; the CD was recorded in Cuba, and manufactured and distributed by a division of a German-owned multinational, BMG Music; many of the sales within Cuba are to tourists at Exaudi concerts, souvenir counters in hotels or airports, and special dollar stores; outside of Cuba, the CD is available at web music retailers and specialized record stores.

More obviously, the international market is of critical importance to mass-market content, which is often modified to cater to different cultures and national sensibilities. The producer, the licensed distributor or censors may edit content and on some occasions replace bits of the work that is shown to audiences in a particular market. The

<sup>15</sup> For example, a Canadian company, Rhombus, making cultural programming for arts channels and for cinematic exhibition (*32 Short films about Glenn Gould and Red Violin*) is briefly discussed in Acheson and Maule (1999a, p. 125).



licensing value of a work varies across national markets in response to the potential revenue. The diminution of value when rights of a television program or film from one culture are licensed to broadcasters or exhibitors in other cultures has been labeled a cultural discount [Hoskins and Mirus (1988); Hoskins, McFadyen and Finn (1997)]. Cultural discounts differ depending on whether a program or film is animated, on its special effects, and on its genre (sports, action series, space dramas, documentaries on universal topics).

Distributors are cognizant of the impact of content decisions on revenue generated in different markets in granting advances and setting contract terms. Success abroad may provide a signal of quality and enhance acceptance at home because viewers believe that domestic programming and film are often propaganda or inferior in quality. For example, Taylor (2000) reports that Eisenstein's *Battleship Potemkin* was removed from Moscow's screens shortly after its premiere in January 1926. When it returned to Moscow's cinemas that summer, "it was sold to Soviet audiences as a foreign hit".

### 5.2.2. *Uncertainty, information challenges and high promotional budgets*

Because of the high degree of uncertainty surrounding revenues and profit of a mass-market film, book or record, knowledge of market prospects has a high payoff. Experienced distributors are often in a relatively advantageous position to assess a project's prospects, a comparative advantage reflected in their critical role in the financing through advances of mass-market films, books, or records for which they have distribution rights. Successes have to generate a sufficient surplus to make the overall rate of profit attractive, given the risk.<sup>16</sup>

Distributors of mass-market content strive to persuade potential customers that they offer a more attractive "surprise" than their competitors. The surprise occurs within a set of expectations shaped by genre, the reputation of key figures, and teasing "leaks". Their success depends on seeding a favorable informational cascade<sup>17</sup> about a prospective release over its potential market. What appear to be obscenely expensive marketing, advertising and promotion (MAP) campaigns are coordinated within a short window of time in pursuit of positive "buzz".<sup>18</sup> Distributors exploit the complementary impact of arranging appearances of artists and writers on talk shows, encouraging articles and reviews in magazines and newspapers, distributing flyers, and experimenting with the proliferating counterparts of these communication channels on the Internet to distinguish their wares.

<sup>16</sup> See De Vany and Walls (1996, 2002) for analysis of these distributions.

<sup>17</sup> See Banerjee (1992) and Bikhchandani, Hirshleifer and Welch (1998) for analyses of informational cascades.

<sup>18</sup> The proportionate expenditures on production and MAP vary. Dale (1997, p. 31) suggests that a typical Hollywood film with a production budget of \$39 m. would spend \$20 m. on US advertising and \$9–11 m. in foreign advertising.

### 5.2.3. *Film and television*

Foreign films and lower budget English-language films are usually distributed through the art-house circuits in English-speaking countries. Distribution of these films is more decentralized and less concentrated in time than is the case for large-budget Hollywood films. The majors are pursuing a larger role in the distribution and financing of these films as their popularity increases. In North America, the art houses generally schedule a wide array of films into time slots over a month or two. This scheduling structure resembles that of television more than that of Hollywood films, which have their runs extended at mainstream cinemas for as long as they fill enough seats. North American viewers obtain variety among Hollywood films by choosing among cinemas on any night of the week and among foreign films by going to the tightly scheduled offerings at the local art-house cinema.

Since the 1980s video cassettes, DVDs, and digital files stored on disk have added potentially profitable links to the chain of exhibition possibilities for films. An increasing number of niche films can be successfully launched without a cinematic release. The greater capacity of DVDs and digital files has raised the value of film libraries by permitting more economical re-release of differentiated versions combined with complementary content. The costs of international distribution have been reduced at each stage in the development of home viewing options.

Digital satellite transmission provides those who live on farms and in isolated towns the same wide array of films and television programming as city dwellers. PVRs (personal video recorders) remove scheduling impediments to viewing. High definition television signals and sets reduce the gap in quality between watching a movie at home and in a cinema. The integration of these three technologies in service packages or through video on demand promises to significantly change the international distribution and exhibition of films and television programming. On the policy side, the inadvertent or conscious spillover of satellite signals from one jurisdiction to another creates international tensions that are difficult to resolve.<sup>19</sup> The time-shifting PVRs can also make advertising less efficient, putting greater pressure on subscription prices as a source of finance.

### 5.2.4. *Books and magazines*

High transportation costs of distributing books to bookstores and of returning unsold stock to publishers and inventory costs result in different arrangements for the distribution and retailing of books. Wholesalers, for example, economize on storage and facilitate ordering by retailers of books published by small houses. A multinational publisher typically has subsidiaries or joint ventures operating in the larger centers of

<sup>19</sup> For an account of the interaction between Canada and the United States on this issue see Acheson and Maule (1999a, Chapter 12).

its main language market and licensing arrangements with independent publishers for smaller centers. The subsidiaries and joint ventures cultivate productive relations with domestic authors and license among themselves and with outside firms rights to distribute their authors' books or publish differentiated versions of their books. Like the film studios, the multinational publishers have internal divisions that compete in smaller markets.

Magazine and journal publishers sell internationally. Competitive considerations provide an incentive and current technology allows magazine publishers to practice price discrimination and to bundle their offerings. Non-commercial journal publishers do the same. The American Economic Association, for example, discriminates by whether a subscriber is a student and if not, by salary, as well as bundling its three journals in different packages. One avenue for expanding magazine sales internationally is the use of split runs, editions with common editorial content coupled with advertisements targeted to different national and regional markets. National publishers who use this format across regions in their domestic market often lobby successfully to prevent foreign publishers from competing in the national advertising market with split runs. Policies designed to deter foreign split runs have generated a number of trade disputes.<sup>20</sup>

#### 5.2.5. *Music*

After the European Commission approved a joint venture combining most of the recorded music operations of Sony and BMG in August of 2004, four companies – Universal Music (French owned), Sony BMG Music (50% owned by the Japanese and German parent companies), Warner Music (American owned), and EMI (British owned) – dominate the international part of the popular music recording industry. The production and distribution structure is typically more decentralized than that of the major film studios. Universal Music, for example, has affiliates, generally wholly owned, in most music markets of any significance. The American affiliate includes large popular music labels – Interscope and Island Def-Jam – as well as smaller specialized labels dealing with classical music and jazz. Each label of one affiliate is aligned with a counterpart label of affiliates in other markets. When a label in a market area signs an artist it becomes the repertoire holder and is responsible for the cost of recording an album and for the core marketing tools – photographs, videos, etc. The repertoire holder of a release makes all the profits earned in its home market and shares in profits earned in other areas. The affiliated labels in other countries are responsible for manufacturing to serve their area, local radio campaigns, advertising, other marketing, and contracting with retailers.

Organizational innovations around this core structure are common. For example, Interscope in the United States became interested in Tatu, a singing duo of two young

<sup>20</sup> For an account of Canadian frictions with the United States over split runs and other aspects of magazine policy see Acheson and Maule (1999a, Chapter 10, 1999b, 2000, 2001a, 2001b).

women discovered by its Russian affiliate. The two affiliates formed a joint venture to be repertoire holders for Tatu. The group reached number one in ten countries including England in January 2003. A subsequent US tour has generated some commercial success and considerable cultural dissonance.<sup>21</sup> A recording artist's contractual royalty rate from sales in different areas varies systematically across international markets reflecting revenue prospects and the costs of distribution. An artist signed by an American label of a multinational, for example, receives a US basic rate of say  $x$  percent of the price for normal retail sales. For Canadian sales by the same artist, the rate is reduced to  $0.85x$  percent. For sales in major territories, which are areas in which American records generally do well, the rate is between  $0.6x$  and  $0.75x$  percent and for sales in the rest of the world, the rate is between  $0.5x$  and  $0.6x$  percent.<sup>22</sup>

An independent recording company without a worldwide network would license a promising release to the rest of the world through an umbrella deal or on a one-by-one basis. Throsby (2002, p. 15) notes that "(i)n most cases, music from the Third World has been brought to wider attention through the activities of independent record producers, standing somewhat apart from the major transnational companies". The average proportion of world sales accounted for by locally produced music, as compared to international repertoire music, increased from 58 percent in 1991 to 68 percent in 2000 (ibid, p. 5). In interpreting this statistic, it is important to keep in mind that export markets are important to independent producers. We give an example from our city to illustrate. In a 1997 interview, Peter Burnside, the head of Pacemaker Entertainment, an Ottawa firm that specializes in reissues of the works of good but not well known Canadian artists, noted that "(b)etween 80 and 90 percent of Pacemaker's sales come from outside Canada, mostly the United States and Europe" and attributed the importance of exports to "the fact that foreigners discovered Canadian acts on their own, whereas many Canadians have a negative reaction to music that they were force-fed because of domestic-content radio-play rules".<sup>23</sup>

### 5.3. Exploiting successfully promoted content across media

If content is successful in its original medium, the economic imperative is to expand the market internationally and explore adapting the concept to other media. The original initiative in cross-media successes can be a book, a film, a play, a comic strip or a musical. MAP campaigns adjust to the characteristics of each medium and their complementarities across media. Successes are typically serialized across media. A recent example is the success of J.K. Rowling's Harry Potter stories in film. Her novels have sold more than 200 million copies worldwide and have been officially translated into 55 languages. Over two-thirds of the box-office income of the first two films released,

<sup>21</sup> Sabrina Tavernise, "A bubblegum duo sets off squeals and squirms", *New York Times*, March 4, 2003.

<sup>22</sup> See Passman (1997, Chapter 13).

<sup>23</sup> Toronto, *Globe and Mail*, December 27, 1997, C6.

*Harry Potter and the Philosopher's Stone* (2001) and *Harry Potter and the Chamber of Horrors* (2002), was generated outside of the United States. The release of the films and the more recent books were tightly coordinated with promotional campaigns. Bookstores were not allowed to sell until a prescribed date and the larger ones opened at midnight of the appointed day to meet the demand from enthusiasts. On the cultural side, fundamental Christians in the United States reacted negatively to the books and film because of the central role of witchcraft, while the US Conference of Catholic Bishops were more ambivalent in their reaction.<sup>24</sup> English professors joined the debate with, for example, *Trites* (2001, p. 484) arguing that the books represented “the type of postmodern questioning of power relations traced by such theorists as Barthes and Foucault and Lacan”.

#### 5.4. *Multinationals and antitrust issues*

The ambiguity about whether or not the business practices of media conglomerates are welfare enhancing makes the application of competition policy and sector regulation in the cultural industries particularly difficult, a challenge exacerbated by the international nature of these industries. The practices include:

- block booking and blind bidding in film;<sup>25</sup>
- most-favored-nation pricing clauses in contracts between the studios and arm's-length movie channels or pay-per-view services;
- exploiting regional playback control system codes to limit DVD competition and facilitate price discrimination by, for example, preventing American DVDs from being played on machines in Australia;
- vertical integration into exhibition, pay-per-view services, and broadcasting;
- the duration and terms of contracts between artists and record companies, payola, bundling of individual songs on a CD, and the pricing of those CDs;
- author-publisher contracting, returns policy with booksellers, and the impact of chains and e-commerce on the local bookstore in book publishing and distribution; and
- cable and satellite access, bundling and pricing as well as the concentration of station ownership in local markets in broadcasting.

In addition to the impact of concentration on traditional measures of welfare, media concentration can limit the breadth and vitality of internal discourse on issues. Political reaction to this possibility is heightened if the domination is foreign. This concern may explain both the additional measures adopted to curb concentration in media industries

<sup>24</sup> Office of Film and Broadcasting, United States Conference of Catholic Bishops, June 3, 2003.

<sup>25</sup> Sedgwick and Pokorny (2004) include updated versions of influential articles on this and other aspects of studio management.

in Europe<sup>26</sup> and in other parts of the world and their relatively strong discrimination against foreign ownership.

The need for coordinated competition policy across jurisdictions has been frequently expressed and a number of agreements have been signed. The recent agreement between Japan and the European Commission (EC), for example, calls for timely notification of actions that would affect the interests of the other party, assistance consistent with its enabling law, possibility of coordination, consideration of action by the other country if that would be more effective, and protection of confidential information exchanged.<sup>27</sup> Coordinated actions have been sporadic and the political pressure to engage in retaliatory rather than cooperative competition policy actions is strong. The Working Group on the Interaction between Trade and Competition Policy of the World Trade Organization (WTO) is currently developing a multilateral framework on competition policy that is consistent with paragraph 25 of the Doha Declaration.<sup>28</sup> In pursuit of this mandate the committee began by examining compliance mechanisms for a multilateral framework on competition policy; progressivity and flexibility clauses for such a framework; ways of providing technical assistance and capacity building to developing countries; and creating an inventory of national legislation.<sup>29</sup>

EC competition policy actions have affected proposed mergers of multinational cultural industry firms. In October of 2000, AOL Time Warner abandoned a takeover of EMI after the EC expressed concern about the resulting concentration in the European music industry and with the vertical link to AOL. In approving the joint venture between Sony Music and BMG Music in July of 2004, the European Commission noted that it and the United States Federal Trade Commission had “co-operated closely in their review of the case” (EC IP/04/959). In another case with international links, the EC in 1999 extended an exemption for another five years to UIP (United International Pictures), a joint film distribution company established by Paramount Pictures, Universal Studios, and MGM. In return UIP committed to change its business plan and extend additional support to the European film industry.<sup>30</sup> In late 2000 the US Justice Department charged Sotheby’s of conspiring with Christie’s to fix auction commission rates

<sup>26</sup> Doyle (2000) investigates whether there is an economic explanation for the recent trend to relax these constraints by examining the impact of size and vertical and horizontal linkages on the profitability of British media firms.

<sup>27</sup> Agreement between the Government of Japan and the European Community concerning Cooperation on Anticompetitive Activities, July 10, 2003.

<sup>28</sup> The paragraph instructed the Working Group on the Interaction between Trade and Competition Policy to: “focus on the clarification of: core principles, including transparency, non-discrimination and procedural fairness, and provisions on hard core cartels; modalities for voluntary cooperation; and support for progressive reinforcement of competition institutions in developing countries through capacity-building. Full account shall be taken of the needs of developing and least-developed country participants and appropriate flexibility provided to address them.” Text of Doha Ministerial Declaration, November 14, 2001.

<sup>29</sup> See Report (2003) of the Working Group on the Interaction between Trade and Competition Policy to the General Council.

<sup>30</sup> EC, *Commission Renews UIP Authorisation for Five Years*, Brussels IP/99/681, September 14, 1999.

charged to sellers (“sellers’ commissions”) over a six-year period in the United States and elsewhere. The EC mounted a similar case. Sotheby’s was fined in both cases and its chairman was jailed and fined in the US case. The two firms split the payment of \$512 million in class action damages.<sup>31</sup>

## 6. International aspects common to the arts and cultural industries

### 6.1. *The new versus the old*

In the traditional arts, the new often competes with an expanding body of existing works. Only bits and pieces of the new are absorbed into an art form’s canon and acceptance is grudging. International approval helps crack this barrier to entry. As a result the collections of the great museums and art galleries of the world reflect similar views about the composition of the world’s heritage. They compete locally with contemporary venues offering avant-garde exhibits and a varied set of popular cultural shrines.<sup>32</sup> Viewing an art installation in a contemporary art gallery, for example, differs markedly from shuffling by the ordered display of Rubens’ *Life of Maria de’ Medici* paintings [Millen and Wolf (1989)] with fellow visitors to the Louvre. A contemporary museum might feature an installation that satirized this experience: visitors might be provided headsets that describe a different set of paintings from those hanging on the wall of the installation while showing the audience’s confused response on a framed video screen placed as one of the paintings in the sequence. Physical art installations are site specific and are often destroyed when the exhibit ends. Unlike Rubens’ paintings, they generally have no direct market value outside of their impact on fees to enter the gallery and their enhancement of the artist’s reputation. Virtual art installations posted on the Web are internationally accessible at a keystroke and rapidly growing in number. Installation artists hail from all over the world. The top four on Google’s list of installation artists (ordered by “hits”) as of August 3, 2003 were a Swiss living in Germany (Urs Jaeggi), two artists from the United Kingdom (Graham Nicholls and David Hall) and an American (Kenneth Rinaldo).<sup>33</sup> All are internationally recognized.<sup>34</sup>

<sup>31</sup> In *Kruman v. Christie’s* the United States Court of Appeals for the Second Circuit re (decision of March 13, 2002) ruled that foreign customers had a right to participate in a suit based on a US antitrust violation even though their own transaction was not part of US interstate commerce. This interpretation of the Foreign Trade Antitrust Improvements Act, 15 U.S.C. would have had significant consequences for the international reach of US antitrust policy. It was overruled by the Supreme Court of the United States in *F. Hoffmann-La Roche Ltd. et al. v. Empagran, S.A. et al.* (decision of June 14, 2004). For more detail of the Sotheby’s case see *Mason (2004)*.

<sup>32</sup> Some examples from the United States of popular sites seeking wider cultural canonization are Graceland, the hall of fame of baseball in Cooperstown, the boardwalk of Venice Beach, and the Haight-Ashbury district in San Francisco.

<sup>33</sup> Google’s list is at [http://directory.google.com/Top/Arts/Visual\\_Arts/Installation\\_Art/Artists/](http://directory.google.com/Top/Arts/Visual_Arts/Installation_Art/Artists/).

<sup>34</sup> For example, Rinaldo has exhibited at The Biennale of Electronic Arts, Perth, Australia; Transmediale Berlin, Germany; ARCO Arts Festival Madrid, Spain; The OK Center for Contemporary Art, ARS ELEC-

In European and North American theater, current playwrights face stiff competition for an audience from the works of Chekhov, Ibsen, Moliere, Racine, Shakespeare and Shaw. An international array of composers from the late eighteenth and nineteenth century dominates symphony programs. Similarly, the vigor of debate about the changing composition of the canon of world literature contrasts with the persisting presence of a core set of Euro-American authors. However, the relative importance of this core has been reduced in recent years by adding many more authors from African, Asian, and indigenous cultures to the current pantheon.<sup>35</sup> The cinemas of Europe have a disproportionate percentage of American films while the repertoire of American theaters and concert halls have a disproportionate percentage of works by dead Europeans. The conflict between the new and old that Swift satirized in his *Battle of the Books, A Full and True Account of The Battle Fought Last Friday between the Ancient and the Modern Books in Saint James' Library*, has spread to other media, experienced lulls, and no peace map exists. The international dimensions of that battle reflect the asymmetric importance of different countries in different media and times.

## 6.2. Openness and creativity

In the overlapping European Renaissances, artists often developed in towns or cities with a craft-industrial orientation and a trading tradition. The guilds were important in developing an apprenticeship system. They regulated the quality of work and relations among clients, and the hierarchy of masters, journeymen, and apprentices in a workshop. In some towns the guilds were protective and ran closed shops. Burke (1999, p. 69) notes:

In Florence, however, guilds did not have so much power. The Florentine governments would not allow them to force all craftsmen to join. Some artists, like Botticelli, entered a guild only at the end of their career. As a result 'foreigners' could come and work in Florence. This more liberal policy, which exposed local tradition to stimuli from outside, may help to explain Florence's cultural lead.

Although commissions by patrons dominated in this period, a secondary market for art began to develop and art dealers started to scour the world for works on behalf of wealthy patrons. The development of a market provided artists with an alternative or supplementary mechanism to patronage for making a living.

TRONICA, Austria; The Kiasma Museum of Contemporary Art, Helsinki, Finland; The Australian Center for Photography; The Chicago Art Institute, Chicago; The Museum of Contemporary Art, Chicago; The Northern Illinois University Art Museum, Chicago; The Home Show, Seoul, Korea; V2 Dutch Electronica Arts Festival, Rotterdam, Holland; Image Du Future, Montreal, Canada; Siggraph, Los Angeles; and The Exploratorium, San Francisco. See <http://www.ylem.org/artists/krinaldo/emergent1.html>.

<sup>35</sup> For example, *The Norton Anthology of World Literature*, second edition [Lawall (2003)] comes in a set of six volumes covering different periods of time. The entries, particularly in volume 6, *The Twentieth Century*, include a large number of authors from Asia, Africa, and indigenous cultures in other continents.



Many writers, professionals, and artists are part of a number of different national infrastructures. Some artists, composers, and writers seek inspiration and training abroad as Rubens did in the sixteenth century. Beckett (1906–1989) lived most of his creative life in Paris, one of the many European cities to host Joyce (1882–1941) after his self-imposed exile from Ireland in 1904. Mendelssohn composed his Scottish and Italian symphonies after visits from his native Germany in 1829–1831. Bartok, Stravinsky, and Schönberg composed in the United States after seeking asylum there. Some writers wrote in the language of their new countries. Beckett wrote many of his works first in French. Conrad and Nabokov wrote classics in English, which was not their native tongue.

Technological developments in film and television have broadened the market for the works of writers, composers, artists, directors, cinematographers and performers and increased their potential incomes. Subsidization and a high income elasticity of demand have done the same for the performing arts. As a result, an increasing number of creative and talented people make a living in the arts and the cultural industries. As the infrastructure of professionals, performers, and technical experts develops, an area may attract more custom and an inflow of specialists from other countries. The “thickness” of the infrastructure provides producers of complex projects more options and greater flexibility to adjust to unexpected events. Examples of creative centers for film production are Hollywood for the English-language film industry, Paris for the French-language film industry [Scott (1997, 2000a, 2000b)], and Mumbai [Rajadhyaksha and Willemen (1999)] for the Indian film industry. Unfortunately, policymakers’ knowledge of how to create a growth center for a cultural activity is limited. A large number of would-be production areas have wasted public monies in ambitious schemes to become the next Hollywood, Paris or Mumbai of the – pick any direction among: South, North, East or West.

The migration of foreign actors to Hollywood is well known. Among the earliest stars were British born Charles Chaplin, Canadian Mary Pickford and Pola Negri from Poland. They were followed by many others including Greta Garbo and Ingrid Bergman (Sweden), Stewart Granger, Cary Grant and David Niven (UK), Merle Oberon and Erroll Flynn (Australia), Peter O’Toole (Ireland), Lilli Palmer (Germany), Walter Pidgeon (Canada), Anthony Quinn (Mexico), and Edward G. Robinson (Romania). Such was the presence of British stars in Hollywood that a cricket club was formed. The same is true of directors: many émigrés and refugees from Europe have been among the well-known directors to work in Hollywood, including Eli Kazan (Turkey), Mike Nichols (Germany), Roman Polanski (Poland and France), and Otto Preminger (Austria). In addition, some European directors have worked for a time in Hollywood and then returned; they include Erich Pommer, head of production at Ufa in Germany, who on his return to Europe made films for Fox and RKO and was active in promoting European coproductions as a way of creating larger markets for European productions. Alfred Hitchcock produced films in both the UK and Hollywood: in the UK he worked for Paramount and in the US for RKO, Universal and 20th Century Fox. Jean Renoir spent the war work-

ing in the United States and did not make a film in France again until 1954 [Bergstrom (1998)].

### 6.3. Trade in complementary equipment, goods, and services

Technological change has also had a direct impact on the visual and performing arts. In addition to the flow of creative ideas internationally, the visual and performing arts have quickly absorbed new or superior quality media, materials, instruments and equipment from international suppliers. Suppose a friend attending a piano concert featuring an internationally-acclaimed artist asks you to guess the nationality of the performer: the large population of virtuoso pianists makes the probability of giving a correct answer low. The probability of being right would be much higher if the friend were to ask you to name the manufacturer of the piano that would be used at the performance.<sup>36</sup> The dominant manufacturers of concert pianos have distanced themselves from their competitors through innovation and quality control; they are “stars” of a different kind. The same is true in other fields serving mass markets: in an article describing the debilitating effects of digital technology and sampling on musical creativity in pop music, Tony Scherman<sup>37</sup> wrote that “(e)very generation, of course, recoils from its successor’s new sounds. Jazz was barbaric, Elvis the low point of humanity, Dylan sold out when he plugged in that Fender”. There was no need for Scherman to explain to the reader what a Fender was. On occasion, equipment becomes a collectible: a 1956 Fender owned by British guitarist, Eric Clapton, was sold at Christie’s for US\$497,000 in mid 1999 while a violin made by Stradivari in 1700 sold at the same house for US\$1,326,000 about a year later.<sup>38</sup>

Each visual and performing art requires complementary equipment. It may be an elaborate lighting system for a staged performance or a security system at the museum for protecting the precious items in a visiting exhibit. Cultural trade encompasses all of the complementary production equipment and the technology embodied within it for shaping images and sounds to satisfy local and international audiences. The linkages to other trade are more obvious in the cultural industries than in the visual and performing arts. Movies and children’s books are linked to the sale of related merchandise. The capital goods used in the production of movies, television programs, books, and records as well as in cinemas and broadcasting studios are significant export items for a number of countries. Content is the software for extensive hardware owned by the consumer – television sets, video or DVD cameras and players, sound systems, and increasingly computers, burners, and scanners. Generally speaking the countries that are significant

<sup>36</sup> A June 5, 2003 article in *The Economist*, “Making the sound of music”, lists (in alphabetical order) Bechstein (Germany), Blüthner (Germany), Fazioli (Italy), Steinway (USA), and Yamaha (Japan) as leading manufacturers of concert grand pianos.

<sup>37</sup> Tony Scherman, “Strike the band: Pop music without musicians”, *New York Times*, February 11, 2001.

<sup>38</sup> Christie’s Review 1999–2000, 257 and 91.

players in trade for related merchandise, capital goods and complementary consumer equipment differ from those that are significant players in content trade.<sup>39</sup>

#### 6.4. *Heritage tourism and festivals*

International sites, such as Angkor Wat, Machu Picchu, the Alhambra Palace, the Imperial Palace (Beijing), the Taj Mahal, and the Hermitage continue to be cultural focal points for legions of visitors. Venice remains a unique cultural complex exceeding the sum of its considerable parts, as Mosetto (1992) has explored in a thoughtful economic study. Some unique attractions face difficult tradeoffs between the benefits of cultural tourism and the stress it imposes on fragile environments.

Scheduling regularity, varied menus of events around a theme, and complementary accommodation networks and booking services make music and performing arts festivals attractive to visitors. The festivals bundle a cultural activity with other amenities of the site. Many are associated with small towns and are a welcome counter to the cultural dominance of large urban centers. Their formats are highly differentiated. A number stress the international flavor of their productions. Some are associated with training schools for the arts.<sup>40</sup> All of those that are popular locally also attract foreign visitors, sometimes to the dismay of local patrons.<sup>41</sup> Frey (1994) and Frey and Busenhart (1996) explore the economics of such festivals in more detail. Different levels of government provide significant subsidies to festivals encouraging their proliferation and expansion. O'Hagan (1992) incorporated international aspects in assessing the subsidization of the Wexford Opera Festival.

Theme parks and sites featuring stars, animated characters, themes and related fantasies associated with a major studio attract many tourists. For the large projects, success depends on the impact on contiguous real estate values bought before the project is consummated and the success of ancillary activities like hotels and restaurants.

#### 6.5. *Versioning content for language markets*

The market for instrumental music is affected by different musical cultures but not by language distinctions. There is no translation of music from one musical tradition to another but influences are integrated. When music and words are combined distinctions proliferate. Lili Marlene is a rare example of a popular song (music written in 1938) that became in the Second World War the most popular song in English with Allied troops on the western front, and in German with the Afrika Corps. When genres travel, the

<sup>39</sup> See Section 8.2 on data sources below.

<sup>40</sup> The Festival of Arts and Music in Aldeburgh, England, for example, is associated with the Britten–Pears School for Advanced Musical Studies.

<sup>41</sup> The two largest theater festivals in Canada, the Stratford Shakespearean Festival (Stratford, Ontario) and the Shaw Festival (Niagara-on-the-Lake, Ontario) report that about 40 percent of their ticket sales are to Americans; see Michael Posner, “Theatre festivals get SARS bailout”, *The Globe and Mail*, June 21, 2003.

adaptation of the language component lags behind the adoption of the music. Mathews (2000, p. 64), for example, comments on the history of rock and roll music in Japan:

Rock in Japan used to be sung in English, and a long-running argument was waged as to whether Japanese was a suitable language for rock. From Japanese bands of the 1970s mouthing songs in English whose meanings they didn't understand, to bands like the Southern All Stars in the 1980s singing in a stilted American-sounding Japanese as if to subordinate the rhythms of the Japanese language to the Western rhythms they felt to exist in rock, to Japanese bands of today, singing rock in a Japanese that sounds more or less like Japanese, we see a distinct linguistic Japanization of rock.

Knowledge of the language of the libretto broadens the audience for opera but the impact is not sufficient to make translating librettos economically attractive. The simpler and less informative expedient of providing subtitles on a suspended screen is more common.<sup>42</sup>

The leading works of a modern playwright are typically translated. Translations of fiction and non-fiction books are more common. Permission of the copyright holder in the original language is required to publish a translation, which is protected by a separate copyright. The quality of translations varies widely. At one end of the spectrum are the almost instant translators provided by some Internet search engines.<sup>43</sup> At the other are creative responses to the original works. Susan Sontag eloquently describes the role of translations in broadening the audience for literary works:

Translation is the circulatory system of the world's literatures. Literary translation, I think, is preeminently an ethical task, and one which mirrors and duplicates the role of literature itself, which is to extend our sympathies; to educate the heart and mind; to create inwardness; to secure and deepen the awareness (with all its consequences) that other people, people different from us, really do exist. [Sontag (2003, p. 15)]

Dubbing presents an additional challenge as the translation has to be synchronized with the picture. The skills of both the translator and the actor reading for the new sound track contribute to a more seamless transition from one language to another. Some countries, like Egypt, do not allow dubbed foreign language pictures to be shown;<sup>44</sup>

<sup>42</sup> They describe what is occurring on stage in brief phrases of the local language that appear on suspended screens. In Australian film director Baz Luhrmann's version of *La Boheme*, which opened in San Francisco in the fall of 2002, the teleprompter flashed "Kapow!", "Thwack!" and "@#!&%" during a mock fight scene.

<sup>43</sup> These programs are improving rapidly but the results are crude. To illustrate their current state a computer-wise colleague entered into the Google English to Italian translator "the flesh is weak but the spirit is willing". He then entered the result into the Italian-English translator and recovered "the meat is weak person but the spirit is arranged".

<sup>44</sup> See Ghoneim (2005, n. 58). Subtitling, which dates from the 1950s, "reflects the regime's objection to the presence of a foreign language being heard on national TV" that "reminds people of the era of colonialization and what followed from all the actions of fascism". He also notes that subtitling is cheaper than dubbing.

restricting foreign films to subtitles provides domestic films more protection because of the relatively low literacy rate in Egypt. Other countries will not allow pictures to be shown unless a specified amount of dubbing occurs nationally. Spain is among the countries that require foreign-language films be dubbed. According to Lluís Bonet, a film critic for Barcelona's *La Vanguardia*, "Franco instigated the practice in the 1940s, as a way of censoring foreign films, and it continues to this day".<sup>45</sup> In some instances, dubbing may improve a film; Jake Eberts, a successful producer and head of England's Goldcrest Films in its heyday, claimed this was true for the film *Revolution*.

The only thing that saved *Revolution* was that it was spectacular to look at and some of the battle scenes were breathtaking. Also, I felt that in foreign territories it might not play too badly. The one thing I couldn't stand was Al Pacino's accent, and the mixture of accents between Pacino, Sutherland and Kinski. When the film was dubbed into Spanish, Italian or German then at least the accents would be consistent and some of the dialogue might be easier to hear. [Eberts and Illott (1990, p. 578)]

At the policy level, countries vie to be the dubbing conduit for foreign films into their language market.

## 7. The governance of international trade and investment

We return now to the way in which cultural trade is governed in international, regional and bilateral agreements, as well as the proposal for a new UNESCO-based international agreement on cultural diversity that will impinge on existing commitments made by countries.

### 7.1. The GATT, GATS and the WTO

Member countries' commitments on trade in cultural goods have added to the basic obligations and rights of the original General Agreement on Tariffs and Trade (GATT 1947). Article XX(f) provides exceptions to the GATT for the protection of national treasures of artistic, historic or archaeological value. Article IV permits countries to impose screen quotas as long as they do not discriminate among contracting parties. The interests of non-United States commercial exhibitors in gaining access to popular American movies, and diplomatic pressure from the United States government and its film industry, tempered the willingness of other governments to impose film quotas that dramatically reduced imports. With the introduction of television the United States was concerned that public broadcasters would change these circumstances. It sought assurances in 1961 from a GATT working party that Article IV would not apply to trade in

<sup>45</sup> *The Economist*, December 6, 2001.

television programming. Draft recommendations were prepared but no agreement was reached. Although Article IV explicitly sanctions film quotas, other protective instruments such as box-office taxes that finance subsidies for domestic production are illegal under the GATT (Articles III and VIII).

The special mention of film in Article IV implies that in 1947 the negotiators considered movies to be a good and not a service in international trade. The rising importance of the transmission of audiovisual content by satellite or through telecommunication systems raised the possibility that these methods of trade would also be categorized as a service and GATT would not apply. Since the reduction in transportation costs would likely expand international trade, major players in the cultural industries were concerned that a rules-based regime be developed for trade in services. Some visionaries also foresaw the significant impact that the Internet would have in blurring the boundaries among the traditional cultural industries and enhancing the flow of content among countries. It was not until the establishment of the General Agreement on Trade in Services (GATS) in 1995 that a rules-based regime was developed for trade in services. The WTO incorporated the GATT with some new revisions, the GATS, and the Trade-Related Aspects of Intellectual Property Rights Agreement (TRIPS). TRIPS affected the cultural industries by extending Berne commitments (with the exception of its moral rights obligations) to a larger set of countries and specifying legal measures and enforcement obligations. If Berne had succeeded in harmonizing copyright policies and enforcement processes, creators would have been assured of a minimum level of protection through legal action against infringement of their copyright in other Berne countries. Unfortunately, this right of private action provides no protection if the other country's legislation or adjudication system fails to comply with its Berne obligations. Under TRIPS, but not under Berne, the government of the copyright holder's country can take effective action against another member country's government for not meeting its intellectual property obligations.<sup>46</sup> Similarly, service commitments under GATS are enforceable under the WTO's dispute resolution mechanism.

The GATS commits members to transparency and giving most favored nation (MFN) treatment unless a reservation is noted. A member may offer other countries market access and national treatment. Such commitments can be total or partial for each of the four modes of GATS supply: cross-border supply where seller and buyer remain in their own countries; consumption abroad, where the buyer moves to the country of the supplier; commercial presence, where the seller sets up a foreign subsidiary in the country of the buyer; and presence of natural persons, where the supplier is temporarily in the country of the buyer to supply the service. Fewer than 25 countries including the US have so far made partial or complete commitments for audiovisual services [Messerlin, Siwek and Cocq (2004, p. 32)]. In the Uruguay Round, countries with an interest in liberalization, principally the United States, accepted this lack of commitment in return for

<sup>46</sup> There have been three copyright cases adjudicated under the WTO's dispute settlement mechanism. They are listed at [http://www.wto.org/english/tratop\\_e/dispu\\_e/dispu\\_subjects\\_index\\_e.htm-bkmk31](http://www.wto.org/english/tratop_e/dispu_e/dispu_subjects_index_e.htm-bkmk31) (accessed August 6, 2004).

gaining assent to the WTO package and a promise to reopen negotiations on audiovisual trade in five years [Cahn and Schimmel (1997, p. 297)]. The resulting sectoral initiative has since been folded into the comprehensive Doha negotiations. In the process, Brazil, Switzerland and the United States responded to a request to outline members' views about governing trade in audiovisual services.<sup>47</sup> Canada and France, on the other hand, have stated that they will make no commitments on culture in the WTO until the passage of an international agreement on cultural diversity.<sup>48</sup>

It would also be possible for a waiver to be negotiated for cultural sectors under Article IX, 3–4 of the WTO Agreement, but this is highly unlikely, as three-quarters of the entire WTO membership must approve and any damaged members would have access to the WTO's dispute settlement mechanism (DSM). Carmody (1999) supports a waiver and provides a draft of a cultural waiver that would last ten years. Under the WTO as currently constituted, GATT and GATS obligations apply to cultural goods and services, respectively. What this means, when obligations differ and the activity is a mix of service and goods components, is slowly being clarified through decisions of dispute resolution panels. It is clear that when both components are present in a product, the defendant in a dispute cannot base its policy on the least constraining agreement.

## 7.2. *Other agreements*

The WTO is not the only international agreement to make reference to culture. Some of these agreements are trade related while others deal with a range of issues including cultural rights as basic rights, preservation of cultural heritage, protection of copyright, international cultural cooperation, co-production and cultural dissemination, the interests of artists and cultural creators, and the promotion of linguistic diversity. A few of the existing instruments are considered legally binding but many constitute a soft law approach to establishing international norms.<sup>49</sup>

<sup>47</sup> These are non-binding communications or "cheap talk" in game theory parlance. See Communication from Brazil: Audiovisual Services WTO S/CSS/W/99, July 9, 2001, Communication from Switzerland: Audiovisual Services WTO S/CSS/W/74, May 4, 2001, Communication from the United States: Audiovisual Services WTO S/CSS/W/21, December 18, 2000.

<sup>48</sup> See further in Section 7.3 below. Canada's position was outlined by its Heritage Minister, see <http://www.cdc-ccd.org/Anglais/Liensenanglais/frameevents.htm> (accessed August 6, 2004). France's position is stated at <http://www.diplomatie.gouv.fr/actu/article.gb.asp?ART=32971> at para.2 (accessed February 28, 2004).

<sup>49</sup> See <http://unesdoc.unesco.org/images/0013/001307/130798e.pdf> (accessed July 28, 2004). Examples of related agreements include: The Universal Declaration of Human Rights (1948); Declaration on the Rights of Indigenous People (1994); Declaration of the Responsibilities of the Present Generations Towards Future Generations (1997); Convention on the Protection of the Underwater Cultural Heritage (2001); and Convention on the Preservation of Digital Heritage (under negotiation).

### 7.2.1. WIPO

After the formation of the WTO, national copyright laws in many countries were altered to comply and the levels of protection chosen often exceeded that required by TRIPS. In addition new treaties covering copyright and neighboring rights not explicitly addressed in TRIPS were negotiated in 1996 under the auspices of the World Intellectual Property Organization (WIPO). Unlike TRIPS they lack an effective enforcement mechanism. The *Copyright Treaty* introduced two new rights – a distribution right, which leaves the issue of exhaustion to be determined by each member, and a right of communication to the public. Members must provide adequate legal protection and effective legal remedies against the circumvention of protective technological measures adopted by copyright owners. In particular, they must provide suitable penalties against removing or altering electronic rights management information from copyrighted works or distributing works that have had this information removed or changed. The *Performance and Phonograms Treaty* of 1996 generally reinforces and extends neighboring rights to cover content posted on the Internet. It creates a public performance remuneration right of performers and record companies subject to national treatment. Performers and record companies are given reproduction, rental, distribution, and the right “of making available to the public of their performances in such a way that members of the public may access them from a place and at a time individually chosen by them”. The last of these is a right to authorize downloading on the Internet. File-sharing and CD-ripping and burning have reduced the effective protection provided recorded music, films, and audio books by copyright and by any newly introduced distribution right. Public attitudes to piracy are important and they are in flux. Expenditures on lawyers and the development of new technological fences are reducing creative budgets. Out of the reactions of the industries, the public, the courts, and policy makers some *modus vivendi* as to what will be in the commons and what will be protected on the Internet will emerge. No one knows what that world will look like.

### 7.2.2. UNESCO

The United Nations Educational, Scientific and Cultural Organization (UNESCO) administers the Universal Copyright Convention, an agreement originally supported by the United States as an alternative to the Berne Convention. With the accession to the Berne Convention by the United States, the incorporation of Berne obligations (except its moral rights provision) in the TRIPS, and the new Copyright treaty of 1996, the relative importance of the Universal Copyright Convention has declined. UNESCO also administers the Rome Convention of 1961 governing neighboring rights.

In 1950, UNESCO drafted the Florence agreement to ensure the free flow of cultural products especially books. The treaty was later amended to include audiovisual materials. Over the years a number of countries ratified the Florence agreement including the UK, France and the US. The agreement has the form of a trade agreement but it lacks an



effective enforcement mechanism. In 1970 another UNESCO agreement, the Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property, was negotiated. Cultural property is defined as treasures from the past ranging from rare collections of fauna, flora, minerals and anatomy to furniture that is over 100 years old. The parties agree that there is a moral obligation to respect not only their own cultural property but also that of other countries. Members can take measures to protect cultural property against theft, clandestine excavation, and illicit export. Importing countries have obligations to deny entry to cultural property that is not certified for export from its country of origin. Members must adopt laws disciplining their museums, libraries and archives from obtaining illicit cultural property from other countries. Cultural property cannot be appropriated and exported under compulsion by an occupying force from a member country.

### 7.2.3. Regional agreements

In the founding documents of the European Union (EU), culture was not mentioned. This has not prevented individual members and the EU from being active participants in the culture-trade debate and in formulating policy. The developments are difficult to summarize. At any time, the outcome reflects the conflicting pressures of drawing a boundary around the EU to ward off American influences, and a series of borders around individual countries in the EU that want to be both European and distinct from each other. This results in lengthy discussions over whether culture should promote unity or diversity. Verbal compromises are reached such as aiming for “unity in diversity”, but they provide little policy guidance.

An important initiative was the 1989 publication of the *Television Without Frontiers* (TWF) document. TWF aimed at creating a single television Community market by setting minimum quotas for European content but allowing additional national quotas and other measures that distort trade. A 1998 review of the progress towards a single television market concluded that the “basic principle of freedom of movement has been restricted in practice”. Other trade-related European policies are the EU cultural subsidies available to content producers and distributors and the extensive use made of coproduction treaties for films and television programs. EU trade negotiators have a mandate to act on behalf of member countries after receiving authorization from the Council of Ministers. According to the Treaty of Nice, in force since 2003 and covering sensitive service sectors that include audiovisual services, the Council of Ministers has to have unanimity from member countries to enter into and to conclude negotiations. Thus each country has a veto on conclusion for sensitive sectors. For all other sectors a qualified majority is required to conclude the negotiations.

The North American Free Trade Agreement (NAFTA), contains provisions relating to cultural trade between Canada and the United States that were previously contained in the bilateral free trade agreement between those two countries. There are three components: an exemption for specified cultural industries; a provision for retaliation; and a list of exceptions. The exemption and retaliation provisions represent a compromise that

is read differently by the two countries. Canada stresses the exemption and the US the possibility of retaliation. To date the wording has not been tested in a formal dispute, in part because neither side is certain whose interpretation would prevail. When disputes have arisen, as in the periodical case,<sup>50</sup> they have been taken to the WTO or settled in the political arena through negotiations between the two countries. Culture was treated as any other sector by the United States and Mexico. Canada made no commitments on cultural trade to Mexico.

Other trade agreements, actual and under negotiation, incorporate special provisions for culture. For example, the agreements between Canada and Chile and between Canada and Israel both include an exemption for measures taken by each country with respect to the cultural industries; no provisions for retaliation are included. In the ongoing negotiations among 34 countries for a Free Trade Agreement of the Americas (FTAA), some countries are seeking a cultural exemption.

As noted earlier, co-productions (the formation of bi-national and multinational funding and production teams) have been a typical mode of operation for the audiovisual industry. These have led to countries signing bilateral (and sometimes trilateral) co-production treaties for film and television programs that depart from MFN since they give preferences to some countries over others. Parties to a co-production treaty often benefit by having access to subsidies and to qualifying as national content in the other treaty country or countries. In addition, countries have bilateral investment treaties (BITs) for foreign investment that may contain wording specifically directed at the cultural industries. The failed negotiations for a proposed Multilateral Agreement on Investment (MAI) aspired to replace these numerous agreements. The draft agreement would have contained wording that addressed cultural concerns; a French proposal would have exempted culture.

For those favoring special treatment for culture, the cultural exception/exemption route, as discussed by Cahn and Schimmel (1997), has been the means chosen to date, but there is growing recognition that this has not achieved the desired level of protection

<sup>50</sup> In 1996, consultations failed to resolve concerns of the United States government that the following Canadian policies – a prohibitive excise tax on split-runs of magazines (editions with largely the same editorial content but advertisements focused on different markets), a prohibitive tariff measure restricting the importation of split-runs, and subsidized postal rates to domestic publications – violated Canada's WTO obligations and a complaint was referred to a WTO panel. A panel decision that was modified on appeal (see WTO, WT/DS31/R, March 14, 1997, and WTO, WT/DS31/AB/R, June 30, 1997) resulted in the Canadian excise tax, the tariff and the structure of the postal subsidy not complying with GATT. Canada repealed the tariff, changed the form of the postal subsidy, and passed *The Foreign Publishers Advertising Services Act*, which would have achieved the same goal as the excise tax while in the Canadian government's opinion complying with its WTO obligations. The United States considered the last step inadequate. A negotiated modification of the Canadian legislation followed, which provided, inter alia, for a staged increase in the limits of Canadian-oriented advertising permitted before a foreign magazine would be subject to the law and modifications to the income tax treatment of advertisements in split runs, a policy that had not been part of the original complaint. For a discussion of the complex interaction between the United States and Canada over the latter's magazine policy see Litvak and Maule (1974), Acheson and Maule (1999a, Chapter 10, 1999b, 2000, 2001a, 2001b).

in either the WTO or regional and bilateral agreements. An initiative is now underway to promote an international agreement on cultural diversity that would offset the liberalizing commitments of trade and investment agreements.

### 7.3. *An international agreement on cultural diversity*

The initiative for an international agreement on cultural diversity has emerged from discussions of a group of like-minded countries concerned over the competition facing their cultural industries and the likelihood that domestic protectionist policies would be challenged under trade agreements. Those drafting a proposed international agreement have stressed the loss of cultural diversity as motivating their actions.

While several draft agreements have been in circulation since 2000, the most recent official version, released in June 2005,<sup>51</sup> results from a UNESCO initiative to develop by the fall of 2005 a convention on protecting and promoting the diversity of cultural expressions. The draft wording reveals the difficulties faced in a number of areas, in particular the determination of what constitutes cultural diversity, as well as the rights and obligations of members of the convention, and the determination in draft Article 20 of how it will relate to commitments made by countries in other international agreements such as the WTO.

Proponents claim that cultural diversity is a *per se* value and a prerequisite for the existence and functioning of the market. They believe that openness results in a great reduction of cultural diversity, retards cultural innovations, and safeguards markets not cultures. The protection sanctioned by a cultural diversity accord would preserve local cultures and their languages. The openness of the current system is responsible for the number of spoken languages that are in danger of extinction.<sup>52</sup> Those who are not persuaded argue that cultural diversity defies effective measurement. As a result, there is no basis to judge whether cultural diversity is declining or what policies promote it. Neither protection nor openness is inherently pro or anti market. By favoring domestic commercial sources, protection does not challenge the existence of the market but redirects it and in the process typically generates fewer benefits to consumers. Cultures and the relations among them are constantly evolving. Some cultures are remarkably resilient in maintaining their distinctiveness from the ubiquitous influences of mainstream cultures.<sup>53</sup> Conceiving and implementing effective governmental measures to arrest the

<sup>51</sup> [http://www.cdc-ccd.org/Anglais/Liensanglais/nouveautes\\_eng/Texte\\_revised\\_Unesco\\_Eng.pdf](http://www.cdc-ccd.org/Anglais/Liensanglais/nouveautes_eng/Texte_revised_Unesco_Eng.pdf).

<sup>52</sup> We are indebted to a communication from Professor Santagata clarifying the concerns of those in favor of a cultural diversity instrument. For a discussion of the pros and cons of a cultural diversity instrument see Acheson and Maule (2004), Benhamou (2004), Iapadre (2004) and Van der Ploeg (2004).

<sup>53</sup> Consider a North American example with European roots. The University of Waterloo in Canada anchors an advanced engineering and scientific complex of higher education, firms, and research centers. The university is also at the center of an old and thriving Mennonite community that eschews worldliness. Old order sects within this community have preserved a dialect of German despite their immersion in an English-speaking environment. The tolerance of a modern secular economy has posed less of a threat to their preservation of language and culture than religious intolerance did in an earlier Europe.

decline of cultures (and their languages) lacking that resilience has been extremely difficult within a nation state.<sup>54</sup>

In preliminary discussions leading up to the UNESCO draft, proponents have suggested that members can self-define what their rights and obligations should be. This makes it difficult to resolve cases where the exercise of one country's rights harms another's. The convention may end up as a declaratory statement, in which case it will make little progress in resolving cultural trade disputes, other than expressing the desires of the signatory countries. An alternative motivation of the proponents may be to persuade WTO members not to make further commitments on audiovisual services in GATS negotiations, thereby maintaining the limited level of liberalization in this sector. Meanwhile trade in cultural goods will remain subject to the GATT. Any progress on such a convention will require the participation of the leading countries involved in trade in cultural products. They will need to be persuaded that a separate convention is preferable to the present commitments contained in the WTO. Aside from the US, countries such as Germany, China (Hong Kong), the Netherlands and the UK have media enterprises whose interests are in maintaining open markets. Even countries with governments of a protectionist bent, such as Australia, Canada and France have business interests that are supported by export activities and would be harmed by trade restrictions.

Progress at the international level on cultural trade negotiations for goods and services is now taking place on two tracks, in the WTO and UNESCO. At some point these must merge for progress to occur. Underlying this process is rapidly changing technology that is making it difficult to enforce any internationally agreed commitments. For those who view openness as promoting cultural diversity, as discussed above, this may be a welcome outcome.<sup>55</sup>

## 8. Measuring trade

### 8.1. Measurement problems

Culture is traded as both a good and a service. In some cases both categories can apply, as in the case of a CD carrying a musical composition. Because of the variety of items

<sup>54</sup> Consider an indigenous North American example. According to the Ethnologue website [http://www.ethnologue.com/show\\_country.asp?name=Canada](http://www.ethnologue.com/show_country.asp?name=Canada) (accessed November 30, 2004), there are at least 9 languages – Abnaki (Western), Chinook Wawa, Haida (Northern), Han, Munsee, Sechelt, Tagish, Tahltan and Tuscarora – in danger of extinction in Canada. International treaties do not restrict the Canadian government's policy responses in this area. The paralysis about what to do reflects the lack of economically viable responses.

<sup>55</sup> For the case that the proponents have failed to establish any enforceable rules see Acheson and Maule (2004) and references therein. For informative comments on that paper see Benhamou (2004), Iapadre (2004) and Van der Ploeg (2004).

traded and differences in organizational and contractual arrangements in the various cultural sectors, data collection, especially for services, often requires the use of special surveys. In this section we survey, with selective examples, the types of issues faced by those collecting and interpreting these data.

Data sources differ between reproducible items such as books, films and music, non-reproducible items such as visual art originals, and activities such as cultural tourism. Items in the first category can either be sold as physical objects, or the rights to use the content may be licensed to a number of persons. For example a book can be sold to a foreign library or person, or the rights to publish a book may be contracted with a foreign publisher. A painting may be sold to a foreign buyer by a gallery or through auction. In these cases a physical object is traded but for those collecting trade data the source of information will differ. The secondhand market is the location of much information for non-reproducible art; some may be collected from public auction houses but other transactions may not get recorded.

Other differences occur due to the location of seller and buyer. In some instances both will remain in their own countries while completing the transaction – the electronic transfer of a manuscript to a printer, for example. In others, the buyer travels to the seller such as when tourists visit museums, historic sites, exhibitions and performances, or the seller travels to the buyer in the case of a soprano or baritone contracted to perform abroad. The establishment abroad of foreign subsidiaries, such as cinemas is another means of servicing foreign markets. Each of these alternatives reflects a GATS mode of supply. Negotiators have an incentive to develop statistics to measure their importance in cultural trade.

UNESCO distinguishes among tangible, intangible and material culture. We have stressed the importance of the international spread of cultural ideas, which are intangible, but measuring their importance remains subjective and is likely to remain so. There is more promise that the measurement of trade in services (which are immaterial) will improve significantly under the stimulus of WTO negotiations. The best existing statistics, such as they are, measure trade in cultural goods, which are material.

New technology can create measurement problems. Visits by foreigners to museums are recorded as tourist expenditures in the balance of payments, i.e., as an export of services. The same museum or a gallery may be visited in virtual form on the Internet for free. If a credit card payment is made for downloading a digital print from a museum shop, it should appear as a service export but will likely not be recorded. Subsequent file sharing of print and audiovisual work in digital format between individuals in different countries for which no payment is made will also be unreported. The same is true when consumers gain free access to foreign newspaper, periodical and broadcasting websites. Executives of Torstar, publishers of Canada's large circulating daily newspaper, the *Toronto Star*, testified before the (Canadian) Senate Standing Committee on Transport and Communications that "The *New York Times* alone has more Canadians registered to its website than the *Toronto Star* has subscribers".<sup>56</sup> Much

<sup>56</sup> Torstar Corporation Brief to the (Canadian) Senate Standing Committee on Transport and Communications, February 16, 2005 at [http://www.thestar.com/static/PDF/050216\\_senate\\_brief.pdf](http://www.thestar.com/static/PDF/050216_senate_brief.pdf), p. 17. Worldwide,

of what has value and constitutes international exchanges currently escapes measurement.

Another form of measurement is the documentation of existing policies and estimates of their effects. In 2002, we participated in an IAI-HWWA conference in Hamburg in which researchers described the array of national policies impacting on audiovisual trade of seven countries: Canada, Egypt, France, Germany, India, Italy and the United Kingdom. The remit included developing a measure for the impact on audiovisual trade of the policies of each of these countries. A number of different approaches were taken to arrive at an informed guesstimate.<sup>57</sup>

## 8.2. Data sources

Cultural statistics of any kind are of generally poor quality, and this includes those recording trade. Part of the problem is deciding what activities should be covered and part in determining how to collect the statistics. UNESCO is a starting point but, as Goldstone (2003, p. 177) notes, the list of activities omitted from the World Culture Reports of 1998 and 2000 is lengthy. These reports only publish those aspects of world culture that are readily available. For 2000, the report contains two tables (pp. 348–355) showing cultural trade for about 150 countries for 1980 and 1997. There are numerous blank entries especially for countries in Africa, Asia and the former Soviet Union. South Africa and Jordan, for example, report no entry in any category. Those countries that do report include only goods – in particular books, newspapers, newsprint and periodicals, typewriters, word and data processors, music related goods, cinema and photographic equipment, radio, television and VCRs, visual arts, antiques and sporting goods. These statistics suggest that for most of the industrialized countries over half of cultural trade in 1997 was in typewriters and word and data processors.

According to Ramsdale (2001) who undertook a study for UNESCO based on the UN's Commodity Trade Statistics on trade in cultural goods, developing countries as a group experience a trade surplus. For the developing countries in 1998, cultural exports exceeded imports (US\$51.8 billion and US\$44.4 billion, respectively), whereas in the developed countries the reverse was true (US\$122.5 billion in exports and US\$169.3 billion in imports). The study also notes “the developed countries were net exporters in the categories of printed matter and literature and cinema and photography, while the developing countries were net exporters of fast moving goods like radios, televisions, sporting goods and games, and recorded music” (p. vi). While the balance of payments position varies among developing countries – Malaysia has a surplus of US\$5.7 billion, Mexico US\$5 billion and China US\$13.3 billion – as a group they have a considerable stake in further trade liberalization, contrary to the views of those advocating increased cultural

the *Economist* (June 18, 2005, p. 52) reports users of news websites have increased from 1.6 million in 2000 to 7.8 million in 2005.

<sup>57</sup> See Guerrieri, Iapadre and Koopmann (2005).

protectionism. Recorded trade among less developed countries is low but pirated trade, particularly in music in Africa, appears from unofficial accounts to be significant.

Reliable information on trade in cultural services is sparse although initiatives are underway to fill this deficiency. In its annual *Review of International Trade Statistics*, the WTO summarizes the data available for three categories of services, transportation, travel and other commercial services. Data on trade relating to culture involving audiovisual services and copyright are part of other commercial services. An analyst interested in trade in these activities would require finer distinctions than presently exist.

Government publications of trade statistics typically cover both cultural goods and services but each country survey does not necessarily include the same items. National and international industry associations typically pick from an assortment of official and other surveys to describe a cultural industry including its foreign activities. For example in Canada, The Nordicity Group in conjunction with the federal government and industry conducts an annual survey of the film and television production industry which includes a figure for the "export value" of the industry that does not conform to the concept of exports recorded for balance of payments purposes; export value in this case includes the value of foreign location-shooting in Canada based on surveys of provincial film commissions. Some of these are domestic expenditures and some are already included as tourist expenditures for meals and accommodation and give rise to double counting in terms of export revenues. The survey does not attempt to report on an equivalent "import value".

Schulze (1999, p. 114) reports on trade in categories of non-reproducible and reproducible art using the 4-digit standard international trade classification (SITC): SITC 8960 (Works of art, collectors pieces and antiques for non-reproducible art); and SITC 8983 (Gramophone records and similar sound recordings for reproducible music); and SITC 8921 (Books, pamphlets, maps and globes, printed matter for reproducible literature). These categories are not limited to art in a narrow sense nor are data available for royalties associated with this goods trade. Ginsburgh (2003) uses data collected from auction houses to analyze international art prices. Other data are found in trade journals and the reports of industry associations that tend to be national and have their own particular mandate to pursue, such as measuring the loss of revenue due to piracy, the loss of jobs due to foreign location shooting, or the screen time devoted to foreign as opposed to domestic films. In the case of piracy, much is unreported or at best "guesstimated". For motion pictures, records and music, entertainment software and books in 2002, the International Intellectual Property Alliance estimates losses of US\$6 billion assigned to over 50 different countries.

Overall the trade data landscape is bleak if the researcher is looking for reliable, consistent and comparable data. What to do? Some help is on the way for measuring services trade resulting from the implementation of the GATS and the organization of data according to its modes of supply. Commercial presence (Mode 3) describes servicing the foreign market through a domestic sale by a foreign subsidiary. It is the only GATS mode that does not have a direct balance of payments impact, although previously the foreign subsidiary may have purchased rights from the parent company

through a Mode 1 transaction. Mode 3 is particularly important because controls on foreign ownership are widespread in the cultural industries. Almost every country imposes ownership constraints on obtaining broadcast licences. In other sectors, the ownership pattern is irregular and affected by historical influences. Canada, for example, has ownership restrictions on bookstores but none on record stores or video rental outlets.

At the national level, there are attempts to establish and coordinate the activities of so-called cultural observatories in different countries to compare policies and measure activity levels. These have been talked about for a number of years and are still some way from providing useful comparisons. Meanwhile technology is changing and altering the form of the activities that need to be measured. The absence of reliable uniform official data suits some parties. When trade disputes occur, reference can be made to particular surveys and used opportunistically to support some desired position.

## **9. Conclusions and suggestions for further research**

Modifying traditional models from international economics by incorporating some economic characteristics of cultural activities and industries – the public good attributes of content, the nature of the medium, degrees of copyright protection and scale and scope economies in production and distribution, for example – is difficult. Incorporating other economic attributes such as the nature of uncertainty faced and how creative processes respond to it is even more difficult. Understanding the mutual impact among the accessibility of cultural products and services and the evolution of individual and communal culture is the most difficult. Not surprisingly given the early stage of development of international cultural economics and the challenges faced, there are many empty boxes for researchers to fill in these categories.

We began this chapter by contrasting the rigor and suspected sterility of international economics with the mushiness and potential fecundity of cultural economics. Making the latter more rigorous is important for the future role of cultural economics in the discipline. Integrating effective models of cultural influences on tastes into international trade will contribute to the relevance of international economics and enhance its effectiveness as a guide for international governance. The political debate on the impact on culture of international openness versus protection begs to be informed by clarification of the economic and cultural forces at play and disinterested analyses of the economic impact of alternative policies.

In our discussion of the international aspects of the high arts and cultural industries, brief examples illustrated the processes by which techniques, aesthetic views and organizational ideas were diffused through collaboration and creative networks. Space considerations limited the number of examples and the detail provided. The international linkages in other cultural activities or industries would have similar contours but would differ in detail because of their unique technological and production characteristics. We would have preferred to cite more detailed economic narratives and histories to illustrate the themes and we hope that future researchers will fill this near void.



New content displaces and complements the old in a complex manner across cultural activities and industries. Similarly new technologies frequently inspire creative adaptation rather than destruction of the old ways of doing things. The phonograph and radio did not destroy live concerts but rather changed their role in the world of music. The organizational responses induced by the advent of new technologies in the cultural industries deserve more academic interest. A current case in point is the changes in the international music industry that have occurred and may occur because of file swapping and CD burning.

International markets and distribution networks have been instrumental in dispersing art, prints, artifacts, rights and the ideas that influence their value around the world. Museums make decisions conditioned by the interaction of the entrance fees charged to citizens and visitors and support received from domestic and international patrons and governments. Television programming is available from around the world in bundled services that depend on a mix of fees, advertising revenues, and government financing to remain viable. Films raise revenues by exploiting demand across different “windows” over time and space. These organizational adaptations to financing imperatives are a small sample of other industrial and non-profit organizational research issues that merit further research. It is hard to believe that graduate students would not find it fulfilling to apply modern research tools to any of these areas, but the direction of research is difficult to predict. We expect to be as surprised by the directions of future research in international cultural economics as moviegoers are by what is currently playing at their local cinema.

## References

- Acheson, K., Maule, C.J. (1999a). *Much Ado about Culture*. University of Michigan Press, Ann Arbor.
- Acheson, K., Maule, C.J. (1999b). “Battlefield and birds: New directions for cultural policy”. In: Hampson, F.O., Hart, M., Rudner, M. (Eds.), *Canada Among Nations 1999: A Big League Player?* Oxford University Press, Don Mills, pp. 155–174.
- Acheson, K., Maule, C.J. (2000). “Rethinking Canadian magazine policy”. *Gazette* 62 (3).
- Acheson, K., Maule, C.J. (2001a). “Canadian magazine policy: International conflict and domestic stagnation”. In: Stern, R.M. (Ed.), *Services in the International Economy*. University of Michigan Press, Ann Arbor, pp. 395–414.
- Acheson, K., Maule, C.J. (2001b). “No bite, no bark: The mystery of magazine policy”. *American Review of Canadian Studies* Autumn, 467–481.
- Acheson, K., Maule, C.J. (2004). “Convention on cultural diversity”. *Journal of Cultural Economics* 28 (4), 243–256.
- Anderson, B. (1991). *Imagined Communities*. Verso, London.
- Banerjee, A.V. (1992). “A simple model of herd behavior”. *Quarterly Journal of Economics* 107, 797–817.
- Barber, B.R. (1995). *Jihad vs McWorld*. Times Books, New York.
- Baumol, H., Baumol, W. (2002). “Maledizione! or the perilous prospects of Beethoven’s patrons”. *Journal of Cultural Economics* 26 (3), 167–184.
- Behrman, S.N. (1951). *Duveen*. Random House, New York.
- Benhamou, F. (2004). “Comment”. *Journal of Cultural Economics* 28, 263–266.

- Bergstrom, J. (1998). "Emigres or exiles? The French directors' return from Hollywood". In: Nowell-Smith, G., Ricci, S. (Eds.), *Hollywood and Europe: Economics, Culture, National Identity 1945–95*. BFI Publishing, London, pp. 86–103.
- Bikhchandani, S., Hirshleifer, D., Welch, I. (1998). "Learning from the behavior of others: Conformity, fads, and informational cascades". *Journal of Economic Perspectives* 12 (3), 151–170.
- Burke, P. (1999). *The Italian Renaissance: Culture and Society in Italy*. Princeton University Press, Princeton.
- Cahn, S., Schimmel, D. (1997). "The cultural exception: Does it exist in GATT and GATS frameworks? How does it affect or is it affected by the agreement on TRIPS?" *Cardozo Arts & Entertainment Law Journal* 15, 281–314.
- Cannon-Brookes, P. (1996). "Cultural-economic analyses of art museums: A British curator's viewpoint". In: Ginsburgh, V., Menger, P.-M. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier, Amsterdam, pp. 255–274.
- Carmody, C. (1999). "When 'cultural identity was not an issue': Thinking about Canada-certain measures concerning periodicals". *Law & Policy in International Business* 30, 231–320.
- Caves, R.E. (2000). *Creative Industries*. Harvard University Press, Cambridge.
- Caves, R.E. (2003). "Contracts between art and commerce". *Journal of Economic Perspectives* 17 (2), 73–83.
- Cowen, T. (1998). *In Praise of Commercial Culture*. Harvard University Press, Cambridge.
- Cowen, T. (2002). *Creative Destruction: How Globalization is Changing the World's Cultures*. Princeton University Press, Princeton.
- Dale, M. (1997). *The Movie Game: The Film Business in Britain, Europe and America*. Cassell, London.
- De Vany, A., Walls, W.D. (1996). "Bose–Einstein dynamics and adaptive contracting in the motion picture industry". *The Economic Journal* 106, 1493–1514.
- De Vany, A., Walls, W.D. (2002). "Does Hollywood make too many R-rated movies? Risk, stochastic dominance, and the illusion of expectation". *Journal of Business* 75, 425–451.
- Dixit, A. (1992). "Investment and hysteresis". *Economic Perspectives* 6 (1), 107–132.
- Dixit, A.K., Pindyck, R.S. (1993). *Investment under Uncertainty*. Princeton University Press, Princeton.
- Doyle, G. (2000). "The economics of monomedia and cross-media expansion: A study of the case favouring deregulation of TV and newspaper ownership in the UK". *Journal of Cultural Economics* 24, 1–26.
- Eberts, J., Ilott, T. (1990). *My Indecision is Final: The Spectacular Rise and Fall of Goldcrest Films, the Independent Studio that Challenged Hollywood*. Atlantic Monthly Press, New York.
- Francois, P., van Ypersele, T. (2002). "On the protection of cultural goods". *Journal of International Economics* 56, 359–369.
- Frey, B.S. (1994). "The economics of musical festivals". *Journal of Cultural Economics* 18, 29–39.
- Frey, B.S., Busenhardt, I. (1996). "Special exhibitions and festivals: Culture's booming path to glory". In: Ginsburgh, V.A., Menger, P.-M. (Eds.), *Economics of the Arts: Selected Essays*. Elsevier, Amsterdam.
- Ghoneim, A.F. (2005). "The audio-visual sector in Egypt". In: Guerrieri, P., Iapadre, L., Koopmann, G. (Eds.), *Cultural Diversity and International Economic Integration: The Global Governance of the Audio-Visual Sector*. Edward Elgar, Cheltenham, pp. 192–217.
- Ginsburgh, V. (2003). "Art Markets". In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham, pp. 140–156.
- Goldstone, L. (2003). "Cultural Statistics". In: Towse, R. (Ed.), *A Handbook of Cultural Economics*. Edward Elgar, Cheltenham, pp. 177–182.
- Gray, J. (1998). *False Dawn: The Delusions of Global Capitalism*. Granta Books, London.
- Guerrieri, P., Iapadre, L., Koopmann, G. (Eds.) (2005). *Cultural Diversity and International Economic Integration: The Global Governance of the Audio-Visual Sector*. Edward Elgar, Cheltenham.
- Hoskins, C., McFadyen, S., Finn, A. (1997). *Global Television and Film: An Introduction to the Economics of the Business*. Clarendon Press, Oxford.
- Hoskins, C., Mirus, R. (1988). "Reasons for the U.S. dominance of the international trade in television programmes". *Media, Culture, and Society* 10, 499–515.
- Iapadre, L. (2004). "Comment". *Journal of Cultural Economics* 28, 267–273.
- Jeancolas, J.-P. (1998). "From the Blum–Byrnes agreement to the GATT affair". In: Nowell-Smith, G., Ricci, S. (Eds.), *Hollywood and Europe*. BFI Publishing, London.

- Lawall, S.N. (2003). *The Norton Anthology of World Literature*, second ed. W.W. Norton, New York.
- Litvak, I.A., Maule, C.J. (1974). *Cultural Sovereignty: The Time and Reader's Digest Case in Canada*. Praeger, London.
- Magurn, R.S. (1955). *The Letters of Peter Paul Rubens*. Harvard University Press, Cambridge.
- Mason, C. (2004). *The Art of the Steal: Inside the Sotheby's–Christie's Auction House Scandal*. Putnam, New York.
- Mathews, G. (2000). *Global Culture/Individual Identity: Searching for Home in the Cultural Supermarket*. Routledge, London.
- Messerlin, P., Siwek, S.E., Cocq, E. (2004). *The Audiovisual Services Sector in the GATS Negotiations*. AEI Press/GEM, Washington/Paris.
- Millen, R.F., Wolf, R.E. (1989). *Heroic Deeds and Mystic Figures: A New Reading of Rubens' Life of Maria de' Medici*. Princeton University Press, Princeton.
- Mosetto, G. (1992). "A cultural good called Venice". In: Towse, R., Khakee, A. (Eds.), *Cultural Economics*. Springer, Berlin, pp. 247–256.
- Nozick, R. (1974). *Anarchy, State, and Utopia*. Basic Books, New York.
- O'Hagan, J. (1992). "The Wexford Opera Festival: A case for public funding?" Towse, R., Khakee, A. (Eds.), *Cultural Economics*. Springer, Berlin, pp. 61–66.
- Passman, D.S. (1997). *All You Need to Know About the Music Business*. Simon & Schuster, New York.
- Rajadhyaksha, A., Willemsen, P. (Eds.) (1999). *Encyclopaedia of Indian Cinema*. Oxford University Press, Oxford.
- Ramsdale, P. (2001). *International Flows of Selected Cultural Goods 1980–98*. UNESCO Institute for Statistics, Paris.
- Robertson, R. (1992). *Globalization: Social Theory and Global Culture*. Sage Publications, London.
- Rushton, M. (1999). "Methodological individualism and cultural economics". *Journal of Cultural Economics* 23, 137–147.
- Schiller, H. (1992). *Mass Communications and American Empire*. Westview Press, Boulder.
- Schulz, B. (1994). "German museums and the American influence". In: Robison, O., Freeman, R., Riley II, C.A. (Eds.), *The Arts in the World Economy*. University Press of New England, Hanover, pp. 136–140.
- Schulze, G. (1999). "International trade in art". *Journal of Cultural Economics* 23, 109–136.
- Scott, A. (1997). *The Limits of Globalization: Cases and Arguments*. Routledge, London.
- Scott, A.J. (2000a). "The cultural economy of Paris". *International Journal of Urban and Regional Research* 24 (3), 567–582.
- Scott, A.J. (2000b). "French cinema: Economy, policy and place in the making of a cultural-products industry". *Theory, Culture and Society* 17, 1–38.
- Sedgwick, J., Pokorny, M. (Eds.) (2004). *An Economic History of Film*. Routledge, London.
- Solomon, M. (2003). *Late Beethoven: Music, Thought, Imagination*. University of California Press, Berkeley.
- Sontag, S. (2003). "The world as India: Translation as a passport within the community of literature". *Times Literary Supplement*, June 13, p. 15.
- Taylor, R. (2000). *The Battleship Potemkin: The Film Companion*. I.B. Tauris, London.
- Throsby, D. (2002). "The music industry in the new millennium: Global and local perspectives". Paper prepared for The Global Alliance for Cultural Diversity, UNESCO, Paris.
- Tomlinson, J. (1991). *Globalization and Culture*. University of Chicago Press, Chicago.
- Trites, R.S. (2001). "The Harry Potter novels as a test case for adolescent literature". *Style* 35 (3), 472–485.
- Tunstall, J. (1977). *Media Are American*. Columbia University Press, New York.
- Van der Ploeg, F. (2004). "Comment". *Journal of Cultural Economics* 28, 257–261.
- Welch, E.S. (1997). *Art and Society in Italy 1350–1500*. Oxford University Press, Oxford.

## THE MAKING OF CULTURAL POLICY: A EUROPEAN PERSPECTIVE\*

FREDERICK VAN DER PLOEG

*European University Institute, Florence, Italy,  
University of Amsterdam, CESifo and CEPR*

*'If it is art, it is not for everybody and if it is for everybody, it is not art.'*

Arnold Schönberg, 1945

### Contents

Abstract	1184
Keywords	1184
1. Introduction	1185
2. Indicators of size and participation for cultural sectors in Europe	1187
2.1. Cultural employment	1187
2.2. Public cultural expenditures	1190
2.3. Size of creative sectors	1193
2.4. Cultural participation	1194
2.5. Museums	1195
3. Cultural goods and services	1196
3.1. Special features	1196
3.2. Value of culture	1199
4. Valid and invalid arguments for subsidizing culture	1200
4.1. Convincing arguments	1200
4.2. Unconvincing arguments	1202
4.2.1. Demand-side arguments	1202
4.2.2. Supply-side arguments	1203

\* Former State Secretary for Education, Culture and Sciences of the Netherlands (1998–2002). Some of the arguments in this chapter come from my main White Paper [van der Ploeg (1999)] and from van der Ploeg (2003). See OCW (2003) for the making of cultural policy in the Netherlands. I thank Vladimír Bína for advice on data for the sizes of cultural sectors in Europe and Jan Honout for helpful discussions. I am also grateful to the participants and the discussant Françoise Benhamou of the conference on 'The Economics of Art and Culture', Princeton University, 10–12 September 2004 and to Sir Alan Peacock for his helpful comments. Finally, I owe an enormous debt to David Throsby for his help in redrafting the manuscript and his many detailed suggestions for improvement of the text.

*Handbook of the Economics of Art and Culture, Volume 1*

*Edited by Victor A. Ginsburgh and David Throsby*

*Copyright © 2006 Elsevier B.V. All rights reserved*

*DOI: 10.1016/S1574-0676(06)01034-9*

5. Quality versus popularity in cultural funding in Europe	1204
5.1. Different ways of funding culture	1204
5.2. Programming diversity and funding of TV	1207
6. The arm's length principle: Institutions matter	1208
6.1. Institutional structures	1208
6.2. Delegation and control	1210
6.3. Analytical example of delegation of cultural policy	1212
7. Subsidiarity, local cultural clubs and federalism	1213
8. International cultural policy: Different approaches in Europe	1215
9. Summary and conclusions: Some lessons for cultural policy-making in Europe	1217
References	1218

## Abstract

No good comparable data on sizes of cultural sectors of the countries of Europe exist. Still, local and national governments of Europe spend substantial resources on culture and cultural sectors contribute significantly to employment and national income. After briefly describing special features of cultural goods and clarifying some misconceptions about the value of culture, valid and invalid arguments for subsidizing culture are discussed. Although it is easy to justify government support for preservation of heritage, this is more difficult for the performing arts. Due to changing technologies and advent of E-culture classic public-good arguments for government intervention in broadcasting and other cultural activities become less relevant. Different institutions lead to different cultural landscapes. Theories of delegation suggest delegating the judgement on artistic qualities and execution of cultural policy to an independent Arts Fund, with the Minister of Culture concentrating on formulating a mission for cultural policy and making sure it is implemented properly. Insights from the theories of local public goods and federalism are applied to the making of cultural policy in Europe. Different approaches to international cultural policy in Europe are discussed. The overview concludes with lessons for the making of cultural policy in Europe.

## Keywords

cultural policy, heritage, performing arts, museums, quality, participation, vouchers, tax incentives, quality, politicians, bureaucrats, delegation

*JEL classification:* H2, H4, P51, Z11

## 1. Introduction

In Europe the market and the arts are often unhappy bedfellows. Many music lovers believe that only subsidized symphony orchestras and classical music ensembles perform good music. Commercial orchestras and musical ensembles may perform musicals, operettas and the popular operas of Verdi, Puccini, Rossini and Wagner or play classical music for the millions, but they will not play difficult or more esoteric pieces of classical music for fear of frightening away the public. If they do, they will lose money. Similarly, many argue that only subsidized theater will stage the more difficult and artistically valuable theater repertoire; commercial theater will concentrate on the lighter stuff such as comedy, cabaret or Christmas pantomimes. Commercial theater in the West End of London can survive due to millions of tourists that understand the English language, but this is unlikely elsewhere in Europe. The big London museums such as the Tate, the Tate Modern and the British Museum can also survive with less subsidy than in other European towns. Still, top museums in the European capitals thrive on one old master, for example, the ‘superstars’ Velasquez in Madrid’s Prado, the Mona Lisa in the Louvre or Rembrandt’s Night Watch in the Rijksmuseum of Amsterdam. Of course, a museum that strives to show and maintain a large and varied collection and undertake research needs a larger subsidy. Exposition centers (e.g., the Kunsthal-concept) may be able to thrive commercially or on little government subsidy.

From the Renaissance onwards cultural production in the Low Countries moved to satisfy demands of ordinary people. In contrast, Italy followed the judgement of Vasari and other independent taste critics. In a similar vein John Stuart Mill said “the uncultivated cannot be competent judges of cultivation. Those who most need to be made wiser and better, usually desire it least, and, if they desire it, would be incapable of finding the way to by their own lights”.<sup>1</sup> Today much of the cultural elite of Europe finds commercial culture suspect and argues that subsidies for high culture are essential. Yet even though national, regional and local governments of Europe hand out sizeable cultural subsidies, the belief in the contradiction between market and quality appears to have diminished; some countries now require cultural organizations to bring in a minimum amount of box office receipts and sponsorship income.

Nevertheless commercial culture offers many marvels throughout Europe. In the past this was true as well. William Shakespeare wrote his best work for the people’s theater and managed to pull in the crowds in a fiercely competitive environment; that this is still true is illustrated, for example, by Baz Luhrmann’s film of Shakespeare’s *Romeo and Juliet* with popular stars Leonardo DiCaprio and Claire Danes in the lead roles. Contemporary Europe has excellent museums that flourish without any subsidy at all. There are commercial theater productions whose quality of performances is at least as good from a high-culture point of view as that of subsidized theater groups. There is a growing number of cultural entrepreneurs who reject subsidies, preferring to succeed

<sup>1</sup> Quoted in Blokland (1997, p. 89).

without the bother and humiliation of going through awkward committees to get subsidy. On the other hand, much of subsidized culture does not realize that they will lose significance unless they reach out to new and more diverse audiences and make use of modern marketing techniques. The subsidized arts cater largely to white educated audiences and not to the huge influx of immigrants in many parts of Europe. In addition, technological advances change the character of many cultural goods and question the rationale of government support for the arts. Why (if at all) and how should the government subsidize production and showing of contemporary arts? Why and how should it finance the collection, upkeep and display of cultural treasures? How does the political process influence public support for the arts?

Cultural pessimists criticize the 'market', because it encourages a money-oriented and culturally impoverished way in which people live together. They fear a world where people, artists and cultural organizations are merely treated as buyers or sellers of goods and services. The market demands purchasing power, not background, education or culture. Who pays, joins in. Market forces dumb down expressions of high culture in order to get mass attention. Culture thus becomes part of the entertainment industry. These critics prefer culture to be expressed in ivory towers for refined elite audiences that make an effort to understand what it is all about. Using economics to understand cultural policy is considered philistine. Ministers of Culture are treated with contempt unless they bring more cash. Cultural pessimists share with medics the zealous emphasis on professional autonomy. Professionals, not politicians must decide which artistic expressions are worthy, although professional autonomy must not be abused for rent seeking.

Economists argue that the market mechanism may generate higher welfare than central planning, and that this may apply to the allocation of arts subsidies. Cowen (1998), for example, argues that the market produces a great variety and quality of culture, not just homogeneous pulp. It produces plenty of low culture, but that is what many people want and the market does produce high culture niches for the elite. Globalization and the Internet allow economies of scale and enable the market to produce diversity and variety. Robbins (1963) and Peacock (1969) derive the case for cultural policy from concern with the public interest and the need to correct markets that fail to deliver Pareto-optimal outcomes. Public goods, externalities, natural monopolies, information asymmetries and frictions thus need to be considered. One also needs to consider equity, and non-economic issues such as distinction, connectivity, security and internal motivation. Some shy away from government interference as this may conflict with creativity or, worse, artistic freedom. 'State Art' has connotations with the Third Reich, some dictatorships today and even present-day Italy. Culture may be better served by socially responsible entrepreneurs and by civic responsibility, which has a long tradition in the US and Europe. Grampp (1989) rejects even this. He claims that the public interest is not served by cultural policy at all, since subsidies only serve the self-centered interests of members of a passionate minority. Regardless of whether he is right, we must reckon with market *and* government failures in the making of cultural policy.

Europe has different systems of cultural policy, each with its own merits and distortions. In this chapter we distinguish the state-driven bureaucratic systems in, say, France

or Italy from the arm's-length approach of the UK, Netherlands and Scandinavia. We differentiate the UK with an independent Arts Council and no ministerial responsibility from countries such as the Netherlands with an independent Arts Council and ministerial responsibility.<sup>2</sup> Finally, we point to Germany, which has almost no federal cultural policy and delegates most of the making of cultural policy to the Länder. In assessing these various systems we adopt a political-economy perspective and make use of the theories of delegation and of local public goods, clubs and federalism. Before that can be done, however, we must understand the nature of cultural goods and the arguments for subsidizing cultural activities.

The layout of this chapter is as follows. Section 2 presents indicators of the different sizes of cultural sectors in Europe. Section 3 discusses the nature and value of cultural goods. Section 4 considers how technical innovation and Baumol's cost disease affect the case for cultural subsidies. Section 5 shows how the way of allocating cultural subsidies affects quality, diversity and popularity of culture in Europe. Instruments vary from tax incentives and other privileges to direct grants for cultural activities. Section 6 discusses subsidy allocation by grant-giving arts councils, advisory arts councils and bureaucrats. Lobbying and rent seeking are considered. The crucial questions are which powers should be delegated and what mission, guidelines and criteria should be used. Section 7 discusses the principle of subsidiarity and applies the theory of local public goods, clubs and federalism to cultural policy-making in Europe. This analysis helps to explain cultural competition between regional governments and sheds light on federalist systems such as Germany. Section 8 discusses approaches to international cultural policy in Europe. Section 9 concludes.

## **2. Indicators of size and participation for cultural sectors in Europe**

### *2.1. Cultural employment*

Table 1 presents estimates of employment in cultural enterprises in 2002 including non-cultural occupations, and groups them according to the NACE-classification.<sup>3</sup> Employment in the cultural sector amounted to about 3.9 million working persons in Europe (excluding Poland and Malta); 3.5 million in the EU-15, representing 2.1 percent of the working population. Employment in cultural enterprises was relatively high in Denmark (2.6 percent), Sweden (2.7 percent), Finland (2.9 percent), the Netherlands (2.5 percent), Ireland (2.4 percent) and the UK (2.8 percent). It was relatively low in southern Europe; e.g., Italy (1.4 percent), Spain (1.7 percent) and Portugal (1.2 percent). More

<sup>2</sup> In this latter case the Arts Council recommends, but the Minister of Culture decides and is held responsible by the parliament.

<sup>3</sup> We use the statistical classification of economic activities in the EU (NACE) for cultural activities in Table 1 and the international standard classification of occupations (ISCO) for cultural occupations in Tables 2 and 3.



Table 1  
Shares of employment in cultural enterprises in Europe in 2002 (percent)

	EU-15	EFTA	New members	Total
Publishing	26	31	17	26
Artistic and cultural activities	56	51	65	56
Retail sales of books, newspapers and stationery	7	7	7	7
Architecture	11	11	11	11
Total	100	100	100	100

*Source:* Definition and Production of Harmonized Statistics on Culture in Europe, Chapter 2: Employment in cultural activities (2002), Eurostat, February 2004.

than half of people employed in the cultural sector had an artistic or cultural activity. Table 2 uses the ISCO-classification and looks at cultural occupations only. In 2002 there was about 2.9 million people working in cultural occupations in Europe, 2.5 million in the EU-15, 0.1 million in the EFTA and 0.3 million in the new EU countries. This represented 1.5 percent of the employed working population of Europe, albeit slightly less in new EU countries. The share of cultural occupations in total employment was particularly high in Sweden (2.3 percent), Finland (2.1 percent), Denmark and the UK (1.9 percent), the Netherlands (2.0 percent), and Spain (2.2 percent). The shares in France (1.2 percent), Belgium and Austria (1.3 percent), Luxemburg (1.0 percent), Germany (1.5 percent), Greece (1.4 percent) and Portugal (0.7 percent) were lower than the EU-average. These figures refer to both commercial and subsidized cultural employment.

Table 3 gives the shares of the various types of cultural occupations in Europe for the year 2002. The largest category is writers and creative or performing artists (45 percent). The second largest group in the EU-15 and EFTA is artistic and entertainment associate professionals (25 and 34 percent, respectively), but in new EU countries the second largest group is information professionals including archivists and librarians (21 percent). Shares of various cultural occupations do not vary much across countries of Europe. A large fraction of cultural occupations is in the commercial sector.

Benhamou (2000) empirically explains growth in employment in the audiovisual and performing arts for France and Great Britain during 1981–1992. The two main explanatory factors were a growing number of people with cultural occupations working in cultural industries, and structural changes in the labor market (more flexibility, fragmentation of work opportunities). The Enterprise Allowance Scheme contributed to higher self-employment in Britain while the specific social security benefits for artists led to higher temporary employment in France. As more comparative data become available from Eurostat, more research on shifts in cultural employment in Europe will be possible.

Table 2  
Employment in cultural occupations in Europe in 2002 ('000)

	BE	DK	GE	GR	SP	FR	IR	IT	LU	NL	AU	PO	FI	SW	UK	EU 15	EFTA	New EU	Total
Information	6.0	5.8	39.6	1.3	16.6	43.8	2.5	25.8	0.3	6.7	1.8	2.5	6.5	11.7	36.4	207.3	10.5	70.0	287.9
Creation and performance	26.4	21.4	243.6	19.3	60.8	144.7	14.0	183.5	0.5	94.4	28.6	17.3	29.4	56.7	176.0	1116.6	42.6	151.7	1310.9
Image and sound	5.0	4.9	52.4	6.5	29.7	26.9	2.0	30.7	0.3	13.5	4.2	5.9	2.7	11.2	50.8	246.7	6.1	33.9	286.9
Art and entertainment	3.6	14.8	114.3	16.0	53.1	37.4	7.4	41.3	0.4	42.3	12.7	5.2	8.1	14.5	241.3	612.4	36.5	48.3	697.5
Architects	9.9	6.2	103.9	13.9	26.9	27.0	2.0	66.5	0.4	5.0	2.8	5.0	3.8	4.8	31.0	309.1	10.0	22.4	341.4
Total	50.9	53.2	553.7	56.9	187.2	279.8	27.9	347.9	1.9	161.9	50.2	35.9	50.5	98.8	535.6	2492.3	105.8	326.2	2924.5
Percentage	1.3	1.9	1.5	1.4	2.2	1.2	1.6	1.6	1.0	2.0	1.3	0.7	2.1	2.3	1.9	1.5	1.7	1.0	1.5

Key: Percentage of employed working population.

Table 3  
Shares of various cultural occupations in Europe in 2002 (percent)

	EU-15	EFTA	New EU	Total
Information	8	10	21	10
Creation and performance	45	40	47	45
Image and sound	10	6	10	10
Art and entertainment	25	34	15	24
Architects	12	9	7	11
Total	100	100	100	100

*Key:* Information consists of archivists, librarians and related information professionals (ISCO 243). Creation and performance are writers and creative or performing artists (ISCO 245). Image and sound correspond to photographers and image and sound recording equipment operators (ISCO 3131). Art and entertainment are decorators and commercial designers (ISCO 3471), radio, television and other announcers (ISCO 3472), street, nightclub and related musicians, singers and dancers (ISCO 3473), and clowns, magicians, acrobats and related associate professionals (ISCO 3474). EFTA includes Iceland, Norway and Switzerland. New EU are the new countries joining EU in 2004: Czech Republic, Estonia, Cyprus, Latvia, Lithuania, Hungary, Malta, Poland, Slovenia, Slovakia and Bulgaria.

*Source:* Definition and Production of Harmonized Statistics on Culture in Europe, Chapter 1: Employment in cultural occupations (2002), Eurostat, February 2004.

## 2.2. Public cultural expenditures

Throsby (1994) suggests that public cultural expenditures varied from 0.02 percent of GDP for the US to 0.14 percent for the UK and 0.21 percent for Germany. Little reliable data for Europe exist, however, on a comparable basis. Preliminary indicators on total gross public cultural expenditures have been published for Austria, Finland, France, Germany, Italy, the Netherlands and Spain. Per capita public cultural expenditures are highest in France (180 euro), Austria (179 euro) and are highest in France (180 euro), Austria (179 euro) and the Netherlands (174 euro). For most EU countries it is hard to distinguish between *gross* and *net* cultural expenditures and between *current* and *capital* expenditures. It is difficult to single out receipts earmarked for cultural sectors and to avoid double counting transfers to lower levels of government. The reliability of figures on cultural expenditures at municipal and local levels is further hampered by classification problems.

Table 4 presents national, regional and local public cultural spending for three European countries, namely Finland, the Netherlands and Spain. The data are not comparable and there are considerable gaps: the Finnish data exclude subsidies for public broadcasting, whereas Dutch data include these subsidies; the data exclude cultural subsidies from other ministries than the ministry of culture and media, contributions from government regulated sponsors such as national lotteries, and tax subsidies for maintenance of monumental buildings; the data also exclude risk capital for film ventures, purchase of works by living visual artists, etc. Data collection also suffers from organizational

Table 4  
National, regional and local public cultural spending (million euros)

	Finland 2000			Netherlands 1999			
	Central	Local	Total	State	Provinces	Towns	Total
Heritage	73.0	41.9	114.9	167.4	50.8	221.0	439.2
Archives	14.5	–	14.5	2.9	0.5	52.2	55.6
Libraries	182.6	128.6	311.2	23.5	28.1	343.1	394.7
Books & press	20.4	–	20.4	18.8	–	–	18.8
Visual arts	9.4	–	9.4	58.0	19.1	42.2	119.3
Architecture	1.7	–	1.7	7.6	–	–	7.6
Performing arts	82.1	73.8	155.9	167.0	12.7	64.9	244.6
Audio/visual/multi-media	13.5	–	13.5	904.7	70.3	5.9	980.9
Interdisciplinary	90.0	97.6	187.5	34.7	10.4	443.3	488.5
Total	482.5	341.4	823.9	1384.5	191.9	1172.6	2749.0

	Spain		
	Central 1999	Regional 1999	Local 1996
Heritage		215.7	65.0
Archives & libraries		67.2	–
Reading & publications		8.3	–
Music, theater, cinema		127.5	–
Promotion-dissemination		73.8	512.3
Cooperation-promotion abroad		49.2	–
Administration & services		19.9	–
Total		561.6	577.3

*Key:* Finnish data only refer to current expenditures, but Dutch and Spanish data do include capital cultural expenditures.

*Source:* Task Force on Cultural Expenditure and Finance, Eurostat, October 2002.

problems, lack of a systematic framework and changes in accounting conventions. Despite these problems, the data suggest that it is wrong to focus attention on cultural spending by national governments alone. In Finland, the Netherlands and Spain national-level expenditures made up only 58.6 percent, 50.4 percent (only 32.3 percent if the public broadcasting bill of 732.5 euro is excluded) and less than 28 percent, respectively, of total public cultural spending in the years shown. Cultural spending by the regions/provinces and the municipalities is thus very important. In Germany with almost no federal cultural budget this is even more the case.

Bille, Hjorth-Andersen and Gregersen (2003) describe the efforts of the Nordic Contact Group for Cultural Policy Research to arrive at a joint Nordic common standard for data classification. They stress the difficulties in defining what exactly is meant by cultural policy. Is it only subsidized culture? Does it include teaching of literature or drawing in primary schools, preservation of heritage or valued added in the pop

music industry? Apart from libraries and newspapers the Group did not arrive at a common standard for data classification; public spending on the creative arts, theater, music and dance, film and media, libraries, museum, archives, etc. were included, but public spending on cultural education, radio and TV and capital items were not. With these definitions, culture's share of gross domestic product during the last quarter of a century was shown to have remained practically the same in Denmark, Finland, Norway and Sweden, namely 0.5–0.6 percent. Economic growth has thus led to an expansion of cultural budgets in real terms.

Little has been done to explain public cultural expenditures from a macroeconomic or political economy perspective as has been done for general government spending, for example, by [Alesina and Perotti \(1995\)](#). However some studies do explain spending on particular forms of culture. For example, [Krebs and Pommerehne \(1995\)](#) empirically examined political and other determinants of public support for the performing arts. [Schulze and Rose \(1998\)](#) analyzed funding of the 'Kulturorchester' in the various regions of Germany. Orchestras are heavily funded in Germany: 94.6 percent of public subsidy comes from local communities or municipalities and more than eighty percent of orchestra funding (compared to 4 percent for the US and 20 percent for Japan) comes from direct public subsidies. Germany has a very high density of orchestras (1 for every 0.6 million citizens) compared with Great Britain (one for every 4 million citizens) and the US (one for every 1.6 million citizens). Local politicians and local circumstances play a key role and decentralization of culture is even anchored in the German constitution. Schulze and Rose found that regional public funding of orchestras in Germany increased with the size of the population in the Länder. Larger towns also had bigger orchestras, but surprisingly the proportion of higher educated people had no effect on German funding of orchestras. The level of orchestra support depended positively on the size of the overall and the cultural budget and negatively on the level of public debt. Schulze and Rose also found that conservative and liberal politicians tend to support classical orchestras more than Social Democratic and Green politicians do.

A further example is that of [Getzner \(2002\)](#), who analyzed determinants of total public cultural spending in Austria during 1967–1998. He found that cultural spending, gross domestic product and the relative price indices of gross domestic product and government spending are co-integrated. The ratio of cultural spending to gross domestic product may vary in the short run, but rises steadily with gross domestic product. Cultural goods seem to have income elasticities greater than one, which lends support for Wagner's law of 'the growth of state activities' as described in [Peacock and Scott \(2000\)](#) and [Frey \(2000\)](#). Public cultural spending also rises with the price index of government spending (which is evidence for Baumol's cost disease). The relative importance of cultural spending thus grows over time as the national income grows. In contrast to [Schulze and Rose \(1998\)](#), cultural spending in Austria thus seems relatively unaffected by the ideology of the ruling political parties or by 'wars of attrition' between political parties described by [Alesina and Drazen \(1991\)](#). As new data become available, comparative econometric studies into the determinants of cultural spending in European countries will become more feasible.

Table 5  
Cultural subsidies per visit and box office receipts in the Netherlands

Performing arts and museums (number of institutions)	Subsidy per visit (average 1994–1997)	Proportion box office receipts to subsidy in 1997
Dance (12)	48 euro	18:82
Symphony orchestra's (10)	41 euro	24:76
Baroque orchestra's (4)	3 euro	82:18
Musical ensembles (17)	16 euro	36:64
Opera (3)	120 euro	22:78
Operette (1)	40 euro	22:78
Theater (27)	49 euro	14:86
Youth theater (13)	25 euro	18:82
Mime (6)	41 euro	14:86
Puppet theater (4)	12 euro	37:63
Scientific musea (4)	56 euro	3:97
Arts musea (4)	7 euro	30:70
Culture historical musea (15)	13 euro	19:81

Source: van der Ploeg (1999).

Public spending on culture contrasts with income from box office and sponsors. For example, Table 5 relates art subsidies in the Netherlands to box office receipts during the mid 1990s. Over this time the state subsidized every opera seat (not counting contributions to costs of concert hall and orchestra) on average by 120 euro. Dance needed less, 48 euro. Theater seats received an average subsidy of 49 euro; seats in concert halls received 41 euro. Each museum visit was subsidized to the tune of 11 euro, a lot less. More accessible forms of culture needed less subsidy than more elitist cultural expressions. Orchestras playing popular music of the 18th century, Baroque orchestras and operettas thus needed less support than opera or symphony orchestras. Paintings of Van Gogh are loved throughout the world, so the Van Gogh Museum needed less subsidy per visit than other museums.

### 2.3. *Size of creative sectors*

Collection of employment and income data on the creative industries at a European, national and regional level does not exist on a systematic basis, in contrast to many other sectors of economic importance; it is thus difficult to conduct comparative research on the size of the cultural sectors or of the creative industries of Europe. Collection of such data deserves high priority. In the absence of comparative data, we simply refer to two examples of independent measurement of the size of the creative industries for particular countries. First, for the UK, the Department of Culture, Media and Sport [DCMS (2001)] suggests that the share of creative industries in the UK gross domestic product is about 5 percent, while employing 1.3 million people. Second, García, Fernández and

Zoffo (2003) present statistics of the contribution of creative cultural and leisure industries to employment and national income for the center, sectors and regions of Spain. They find that they contribute about 4.5 percent of the gross national product and give work to 7.8 percent of Spanish employees of which only a tiny fraction relates to publicly financed culture. About 70 percent of value added is taken up by the performing, musical and audiovisual arts and publishing and printing. Most of it is concentrated in Madrid and Catalonia.

#### 2.4. Cultural participation

The Euro-barometer surveys<sup>4</sup> give a rough comparison of cultural participation in Europe. The Nordic countries and the Netherlands seem to be ahead in use of PCs and the Internet, while southern and East-European countries lag behind. The people of Nordic and Baltic countries, the Czech Republic and the Netherlands read more books and newspapers than the EU-average. Reading in Mediterranean and Eastern-European countries is below the EU-average. A similar picture emerges for visits to cultural institutions and artistic activities. It will be interesting to study the influence of variables such as education level, age and gender on participation once Eurostat makes these data available. In the mean time SCP (2001) offers a partial comparison of participation in various cultural activities based on national surveys in nine European countries – see Table 6 and Bína (2003). In all countries the less educated show less interest in traditional forms of culture *and* popular culture (e.g., pop concerts); in other words, cultural

Table 6  
Cultural participation in eight European countries

	Classical concert			Pop concert			Museum			Theater		
	All	Young	LE	All	Young	LE	All	Young	LE	All	Young	LE
Netherlands	16	8	9	25	38	12	31	26	17	27	25	15
Belgium	31	46	14	27	59	19	48	50	33	49	50	32
France	9	6	5	16	–	–	32	39	24	39	35	31
GB	13	7	9	18	46	15	32	32	25	39	35	31
Denmark	16	–	–	–	–	–	55	–	–	26	–	–
Finland	11	10	3	12	37	8	43	49	34	38	36	30
Italy	10	10	5	19	39	12	29	35	18	18	20	8
Spain	7	8	–	10	17	7	28	41	–	14	18	6

Key: Percentages of those 15–75 years of age that visited a classical concert, pop concert, museum or theater during the past twelve months with differentiation for young people in the age group 15–24 and the less educated in the bottom tertile (LE) in the 1990's.

Source: Bína (2003).

<sup>4</sup> *Europeans' Participation in Cultural Activities in 2001* (EU15) and *New Europeans and Culture in 2003* (EU members in 2004 plus Bulgaria, Romania and Turkey).

Table 7  
Museum statistics for Europe 2003

	Number	# visits	% free	Income	Entry fees	Staff	Participation
FI	201	91	37	19.6	8.7	1590	37
FR		111					23
GE	2729	121					33
IT	384	52	47	81.0			28
LA		62				1735	59
LU	26	43	24	6.0		223	32
NL	541	130	30	414.4	56.9	8935	32
NO	146	183	46	236.9	18.9		45
PR	218	59				2648	16
RO	519	7					
SL	102	72	23	17.1	2.1	2292	
SP	878	106	56			10,951	22
SW	184	185	43	327.6	30.7		52
UK	1102	127		491.0		16,777	42

*Key:* Number (of museums opened at least 200 days a year), # visits (visits per 1000 inhabitants), % free (percentage of free admissions of total visits), income and entry fees (millions euro), and participation percentage of population of 14 years and older have visited a museum in 1994. # visits for Germany is excluding free entries and for the UK is based on visits to 1182 out of 1850 responding museums.

Data available for Finland (FI), France (FR), Germany (GE), Italy (IT), Latvia (LA), Luxemburg (LU), Netherlands (NL), Norway (NO), Portugal (PR), Romania (RO), Slovak Republic (SL), Spain (SP), Sweden (SW) and UK.

*Source:* EGMUS, Museum Statistics 2003, Eurostat; participation data from Eurobarometer 2001.

interest does not seem to depend so much on whether it is high or low culture, but on whether the participant is better educated or not. This suggests that cultural education at young ages may overcome some of the inequality gap in cultural interest and cultural participation. The decline of interest for traditional culture particularly affects young people in the Netherlands, even though the rise in educational levels is similar to elsewhere in Europe. However, comparative data on cultural participation are needed before such judgements can be made more firmly.

## 2.5. Museums

The European Group on Museum Statistics (EGMUS) collects available data on museums in Europe. Since national museum surveys are conducted very differently in different countries, present data are not comparable. However, EGMUS has now designed a common core of questions to be inserted in or extracted from national surveys, in order to guarantee comparability at the European level. Some partial comparisons can be deduced from the data presented in Table 7. For example, Norway and Sweden have a large number of visits per 1000 inhabitants, followed by the Netherlands, the UK



Germany, France and Spain. Citizens of Italy, Portugal, Luxemburg, Latvia and Romania visit museums rather less than elsewhere. Museums throughout Europe often have free admissions and entry fees make up only a relatively small fraction of total museum income. Most countries have many public and private museums covering art, archaeology, history, ethnology, science and technology. The Nordic countries, Latvia and the UK have a relatively high percentage of the population that has visited a museum while museum participation in France, Spain, Italy and Portugal is below the European average.

### 3. Cultural goods and services

#### 3.1. *Special features*

Cultural experiences can be aesthetic, touching, memorable and mind baffling. They may even make people think differently about life. In contrast, instant entertainment may be fun but the touching and baffling effects fade away quickly. However, it is important to understand that cultural experiences are also normal economic goods with substitutes (sport, family dinners) and complements (newspapers, magazines, transport, etc.). In fact, most cultural goods and services have, in varying degrees, both private good (rival, excludable) and public good (non-rival, non-excludable) properties. The latter may derive from generalized or local community benefits supposed to arise from the arts or simply from externalities. If non-market benefits exist, free markets will under-provide cultural goods and services due partly to free riders not wishing to pay, and thus subsidy may be warranted.

Other chapters in this volume explain this in more detail, but it may help to give a few examples. Pay-TV and, to a lesser extent, visits to the performing arts or museums are private goods. One can internalize all benefits, so prices reflect true costs, the market functions and no subsidy is required. If performing arts and museums generate positive educational externalities,<sup>5</sup> they may have some public good aspects and merit some subsidy. However, if congestion costs are large due to noisy, uninterested people destroying the enjoyment of others attending an open-air concert, the public good aspect is diminished and a (higher) fee must be charged. Maddison and Foster (2003) estimate that the congestion cost imposed by the marginal visitor to the British Museum is eight pounds.<sup>6</sup> The pure public good case for performing arts and museums is thus not obvious. On the other hand, the marginal cost of the enjoyment of an extra person is zero for broadcasts of radio or TV and monumental buildings. Though jamming can exclude people from broadcasts, it does not make economic sense. Similarly, it is

<sup>5</sup> For example, De Swaan (2001) argues that language is a public good. Drama, literature and film may thus justify subsidy despite being excludable, rival goods.

<sup>6</sup> Based on valuation data collected before realization of the Museum's Great Court by Norman Foster.

difficult or costly to charge passers-by for the benefits of restored heritage or splendid architectural design. Furthermore, projects like the Guggenheim museum in Bilbao designed by Frank Gehry or the Centre Beaubourg designed by Renzo Piano and Richard Rogers may pull in tourism, scarce knowledge workers and new business to Bilbao and the depressed Les Halles area of Paris. Another example of such a city externality might be the building of the largest trade fair in Britain in Birmingham partially in response to the City of Birmingham Symphony Orchestra and to the relocation of part of the Royal Ballet there. Thus open-channel radio and television and maintaining heritage may have public good aspects and may deserve subsidy.

Rapid technological innovation changes the nature of cultural goods. In the past one could record fine classical music on a cassette from radio or a record, but quality was not perfect. Today consumers get their favorite music from Napster-like sites. The quality is very good. CDs are rapidly changing from rival (private) to non-rival (public) goods. Record companies find it tougher to make profits and may invest less in CDs, while extra income from concerts, books, interviews, merchandise and other related activities is becoming more important for performing companies. These technological advances, together with the Internet, globalization of the arts market and Baumol's cost disease all contribute to a shift from unique, autographic art to reproductive, allographic art. The trend towards specialization and division of labor implies that artists make designs and others produce, reproduce and distribute the art object to the public at large. Examples are books, CDs, records, lithographs, posters, photographs, DVDs, CD-ROMs and pay-per-view TV. This shift has lowered prices and increased accessibility of classical and contemporary art for large sections of society. Similarly, there has been a shift from physical cultural experiences to the Internet, where websites of libraries, museums, archives and performing arts companies enable access at any time and any place in the world at almost no cost. Yet just as the extra demand for DVDs has not undermined demand for cinema, free virtual cultural expressions on the Internet do not seem to have eroded demand for seeing physical displays of culture. Finally, there has been a shift in colleges of art from autonomous, subsidized art to applied, commercial arts such as fashion or design. These graduates prefer empowerment (loans for tools, etc.) to income support.

Some culture is produced under increasing returns to scale and winner-takes-all markets. Harris and Vickers (1985) show that there is over-investment in new technologies if different firms race to be the first to get the patent. But if patent markets work badly, there is under-production of cultural goods; in the Internet age it is harder to protect property rights on creative expressions and this may discourage artists to produce. Rengers and Plug (2001) build on Throsby (1996) and estimate a joint model of the choice of visual artists to opt for subsidies or market funding of their work. They show that subsidizing the visual arts through grants and commissions enhances the winner-takes-all tendency for the market at large.

High culture differs from instant entertainment. First, culture is an acquired taste. Preferences for cultural consumption are not given, but shaped by education and experience. Rather than saturation described by Gossen's second law of diminishing marginal

utility, cultural goods show *over time* increasing marginal utility. The hundredth literary book one reads gives more satisfaction than any of the previous books, since the frame of reference will be bigger; the latest concert of Bach music impresses more than previous ones, since one gradually discovers the unifying themes in Bach's oeuvre. Furthermore, culture is a memorable experience; seeing *Hamlet* at a young age can lead to lifelong memories.

Second, Bourdieu (1979) argues that high culture has snob appeal. When the masses start appreciating some forms of high culture, the elite loses interest and moves on to other forms. It is akin to the theory of clubs where the utility of particular cultural goods falls if more people enjoy them. Baumol and Bowen (1966) noted that audiences for the performing arts in the US and UK consist mainly of middle-aged people, professionals, managers and white-collar workers with high education and high incomes. Throsby and Withers (1979) found similar profiles for Australian and US audiences. SCP (2000) finds for the Netherlands a diminishing interest for classical art of people between 16 and 40 years and growing interest among older people. Prieto-Rodríguez and Fernández-Blanco (2000) use the Spanish *Structure, Conscience and Class Biography 1991 Survey* to estimate a bivariate probit model to characterize the average profile of audiences for classical and popular music. Both groups seem to have an 'innate' taste for music, suggesting that music lovers are 'omnivores' who enjoy classical and popular music. Education, cultural backgrounds of parents and occupational status have strong positive effects on classical music listening for young adults (30–45 years), but not on popular music listening. Sintas and García Álvarez (2002) analyze social consumption of culture by Spanish people using both Bourdieu's 'distinction hypothesis' and Lancaster's (1966) theory of characteristics of products. They conclude that culture in Spain has symbolic value associated with social class and allows people to express and distinguish themselves. Nevertheless, the rising education level since the Civil War has not increased consumption of high culture proportionally.

Yet even if only a minority values high culture, the government may wish to encourage the majority to have an interest in it. Such paternalism overrules consumer sovereignty. However, Bille (1997) shows that the Danish public is willing to pay at least as much as the Royal Theater in Copenhagen receives in public subsidies, even though the visitors comprise only 7 percent of the population and are richer and better educated. This suggests that many people are happy to support the arts even if they do not visit themselves. If so, there is no need for paternalism, since people attach option or other values to performing arts; enjoy reading critical reviews or watching recordings of performances; value the derived benefits for television, the film industry, cultural education, cultural heritage and traditions; and appreciate the international prestige.<sup>7</sup>

A final special feature of cultural goods and services is that their enjoyment takes time. This means that the full cost includes forgone after-tax wage income as well as the cost of ticket and travel. Hence the elasticity of demand is less sensitive to the change in

<sup>7</sup> See also Frey and Pommerehne (1989).

the ticket price, especially for higher earners. This offers scope for price discrimination where venues try to charge high earners with little spare time more.

### 3.2. *Value of culture*

Throsby (2001) explains the aesthetic, decorative, spiritual, social identity, historical, symbolic and authenticity value of art. Art also has use, exchange, store, status, option or bequest value. Some argue that the value of arts is (like human life) infinite. Noonan (2003) offers an instructive meta-analysis of many empirical contingent valuation studies of willingness to pay for cultural goods varying from local TV, opera and UNESCO World Heritage Sites. On the latter Maddison and Mourato (2001) study willingness to pay for Stonehenge and Carson, Mitchell and Conaway (2002) do the same for the Fés Medina.

Arts Councils often use ‘quality’ as the prime criterion. But who decides what ‘quality’ is? Is it artists themselves, expert members of an arts council, art critics or the public? Is culture the Western canon of established high culture or the culture offered by newcomers and other civilizations? Is ‘quality’ high culture or dumbed-down culture? Does quality only emerge in the confrontation with a public? The optimal level of quality requires the marginal cost to equal marginal benefit of quality. Maximum quality occurs where government ensures zero marginal cost of extra quality, but that is not necessarily socially optimal.

Stocks of cultural goods need an intertemporal approach to value. A crude estimate by directors of Dutch museums of the market value of their collections was – depending on the current prices of Van Gogh paintings – 20 billion euro. Only about 5 percent of museum assets are on display, the rest is in storage; nevertheless, museums strive for bigger collections and buildings. Bookkeeping of museums ignores the opportunity cost of their collections in their accounts. Museums act as if their collection, their most important production factor, is almost free. Hotelling’s arbitrage principle for exhaustible resources suggests that one is indifferent between selling the assets and investing the proceeds on the one hand and keeping and displaying, lending or storing the collection on the other. Thus the following calculation could be performed: expected net gains in the value of the collection *plus* net gains from exhibiting the collection on a permanent or temporary basis *plus* returns from loaning the collection to other museums, companies or the public *plus* gifts from sponsors, donors and friends to help with purchase of new items and exposition *minus* cost of storage *minus* rate of depreciation and wear and tear of the collection *EQUALS* the opportunity rate of return (the market rate of interest) *minus* subsidies for the collection from local or national governments.

If there is an interested and fee-paying public, it may pay to exhibit items. Otherwise, it may be better to store them. Since curators have a strong incentive to use funds on storage and research, Frey (2000, Chapter 3) argues for increasing incentives for collection mobility (nice for regional museums with lesser collections), regulated sales of objects and exhibitions. The opportunity costs, roughly 1 billion euro at a 5 percent interest rate, are much larger than the subsidies museums receive. Yet they do not fea-

ture in the cost–benefit analysis of museums, so there is little incentive to make use of valuable collections. Making these costs transparent would improve the management of museums and may resolve the potential conflict of interest between curators who want to study and protect, *and* exposition managers and the public who want the collections to be on display.

#### 4. Valid and invalid arguments for subsidizing culture

##### 4.1. Convincing arguments

First, there is a strong case to invest in cultural awareness and cultural education of children, especially for developing a taste for cultural experience goods that leave a lasting impression rather than instant entertainment. It is important not to force high culture upon school children, but to let them have a say in what they want to enjoy and discuss. One possibility is to give vouchers, so children can go four to six times a year to a museum, film, theater play, dance performance, concert or opera. National and local governments in Europe give cultural organizations special subsidies for developing special activities for young people. For many children cultural education is their only chance to develop an awareness of and taste for high culture. The main rationale for subsidizing cultural education is that, as noted above, culture is an acquired taste and an investment in the future social stock of cultural capital.

Second, as subsidized high culture is primarily enjoyed by higher socio-economic groups, there is a case for bringing high culture to the rest of the population. By bringing high culture to public parks, pop temples, libraries, community halls, shopping precincts, etc., one can reach out to new and more diverse audiences. This would also provide a genuine experience as discussed in [Pine and Gilmore \(1999\)](#). The action plans of central and local governments of the UK and the Netherlands allow for interesting experiences in the twilight zone between high and low culture. Examples include serious theater and ballet at pop festivals, school children rapping with symphony orchestras, pop artists from immigrant countries backed up by symphony orchestras, the staging of Wagner's *Ring* in central parks, poetry on billboards, etc. These action plans force the arts to make their output more of a non-rival, non-excludable good and less of a luxury good, so there may be a case for government subsidy.

Third, there often is a case for stimulating demand rather than supply of cultural goods. For example, the Netherlands has shifted from income support for artists to interest subsidies for buyers of contemporary visual art. Artistic Programming Funds may encourage provincial venues to program more adventurous avant-garde culture. It takes time to cultivate an audience. Also, it is worthwhile to have a longer run for successful productions. Giving a 'bums on seats' premium to the performing arts provides greater incentives for drawing bigger audiences and generating income from the market.

Fourth, since non-rival and/or non-excludable cultural goods are public goods, the market will fail to deliver them at a socially optimal level, and government should step

in with subsidies, tax incentives, regulation or public sector provision. These public cultural goods should be provided according to the rule of [Atkinson and Stern \(1974\)](#), who propose that the sum of marginal benefits of culture to all people in society *equals* the relative price of cultural goods *times* the marginal cost of public funds *minus* the effects of cultural goods on the tax base. The well-known Samuelson rule states that the sum of the marginal benefits of cultural goods expressed in resource units must equal the relative price of cultural goods. The Atkinson and Stern rule modifies this in two ways. First, the cost is raised, as they are financed by distortionary taxes. This extra cost is high if the wage elasticity of labor supply is high and pre-existing tax distortions (e.g., the tax rate on labor) are high. Second, making cultural goods more widely available may make people enjoy them more and work less hard and pay fewer taxes. This erosion of the tax base raises the cost of public goods. In fact it may be more likely that more cultural goods will attract more high-skilled workers, businesses and tourists and thus raise tax revenues; obviously, this would lower the marginal cost of public goods. In sum, the demand for public cultural goods relative to private goods will be large if the costs of cultural goods is low (e.g., due to supply subsidies or tax facilities), if the marginal cost of public funds is low (i.e., if labor supply is inelastic and the tax rate on labor is low), and if cultural goods have a positive impact on economic activity.

There is a danger that expert advice on allocation of subsidies will be biased in favor of established cultural organizations, hence it is important to ask the Arts Council or the Arts Fund to make room for new cultural initiatives and support applications that appeal to new and more diverse audiences. An extra requirement may be that the performing arts should obtain at least a certain percentage of their income from the box office. For example, in the Netherlands cultural organizations must obtain at least 15 percent from outside income. [Table 5](#) shows that this is easily satisfied for the well-loved classics played by the baroque orchestras and to a lesser extent by the symphony orchestras, musical ensembles, opera and puppet theaters, but dance, theater and mime may have difficulty in satisfying this requirement. The Dutch public pays only a small part of the true cost of the performing arts; people may find an opera ticket expensive, but rarely realize that it is only 22 percent of the full cost. Similarly, people may not realize that they only pay 14 percent of the full cost of a theater ticket. Such high subsidies may be hard to justify, since performing arts are neither non-rival nor non-excludable and thus do not qualify as public goods unless one relies on the social cultural capital argument. In particular, the social value of theater may derive from keeping the language alive.

Fifth, there is a strong case for government subsidy for cultural goods with existence, option or intergenerational value, especially if they contribute to the social stock of cultural capital and are shown to the public. This applies to subsidizing restoration and maintenance of monumental churches, farms, castles, bridges, locks and landscape values. It also applies to other heritage such as archives of history, photographs, films and performing arts, libraries, museums and archaeological treasures, particularly if they are on display to the public. Each of these adds to the social stock of cultural capital and, if they deteriorate, harm is done to the welfare of future generations. The option value argument says that, even if one never visits certain parts of the country, one does not want

the heritage in those parts to disappear. Some argue that the opera provides option value and adds to the social stock of capital as well. Obviously this is also true for funding research and development in the arts, especially as fundamental innovations in the arts are very difficult to protect through patents and/or author rights.<sup>8</sup> There is also a strong case for government support for architecture, spatial planning and arts objects in public spaces, because these clearly add to the social stock of cultural capital and often to civic pride as well.

Sixth, failures on the supply side may merit government intervention. Cultural goods industries may have difficulties in being set up, so there may be a case for temporary subsidy. This is the infant-industry argument. Since many poor countries often have a rich and unique cultural tradition in both visual and performing arts, there is a case for temporary cultural aid in economic development programs. This may boost tourism, yield foreign currency and act as a lever for sustainable growth [UNESCO (1998)]. Tibetan horn players, Balinese dancers and Cape Verdian singers have unique selling points, which can be popularized through the global media, modern reproduction methods such as CDs and DVDs and live performances. These artists can help brand their own country, like the Beatles and Abba did, and thus boost their economies.

Finally, the market tends to avoid risk unless high profits are expected. Avant-garde cultural activities face bigger credit-market constraints than more profitable, run-of-the-mill cultural activities (musicals, cabaret, bestsellers, etc.). The government should thus stimulate the provision of risk-bearing capital for risky cultural activities that are regarded as valuable. Government-funded participation companies may help getting risk-bearing capital from the market, especially if combined with fiscal incentives. Governments can also stimulate cultural investment by granting interest rate incentives for cultural projects. Friends of a particular museum or theater may then be willing to lend money in the form of soft loans.

## 4.2. *Unconvincing arguments*

### 4.2.1. *Demand-side arguments*

A popular argument is that demand for cultural goods depends on their supply and that the government should thus subsidize supply, otherwise people will never know the value of cultural goods that may never materialize. For example, without subsidies for the opera, people may never know that Gluck's work existed and thus demand will not occur. Still, this does not justify subsidizing supply of high culture unconditionally. It does suggest a boost to the demand for high culture through cultural education, vouchers for parts of the population that normally do not go to expressions of high culture, action programs reaching out to new and more diverse audiences, and programming subsidies.

Another fallacious argument is that cultural goods should be affordable to everyone and not only to rich art lovers; the cultural lobby typically argues that one should lower

<sup>8</sup> See the Schumpeterian arguments in Wijnberg (1994, 1995).

the price for everybody by subsidizing these cultural goods. However higher-income consumers profit much more from these subsidies than lower-income consumers. The dead-weight burden of lump-sum subsidies for the supply of cultural goods may be very high. It is much more efficient to boost demand among lower-income groups or children through education, vouchers, action programs or other means. This is a more effective way to ensure that people can choose themselves which culture they want to experience. If one is concerned about a tendency for middle-of-the-road taste at the expense of avant-garde and high culture, teachers and programmers of venues might try to gently cultivate a taste for experience goods. Note that supply subsidies allocated by committees of experts typically stimulate high culture for the elite; although this seems in line with the moral ideas of Thomas Aquinas that the prominence of someone's position in the community should imply a larger share of the common goods, this medieval view is hardly endorsed today.

Yet another fallacious demand-side argument is that price increases lead to even bigger falls in demand and thus to fall in revenues, hence supply subsidies must be given. This is unconvincing, since demand for cultural products is often inelastic so raising prices does raise revenues. If this means that concert halls, theaters, etc. are not fully utilized, there is a better case for boosting demand rather than supply for concerts, theater, etc.

Finally, a popular argument for granting government subsidy is that cultural goods are merit goods; if society demands an insufficient amount of these goods even though it can afford them, the government should subsidize them. Although this paternalistic argument may justify subsidizing cultural education, it is not a convincing one in support of across-the-board cultural subsidies.

#### 4.2.2. *Supply-side arguments*

A popular argument is that cultural goods are produced under decreasing average costs, so subsidy is needed to ensure full utilization of concert halls, theaters, museums, etc. and ensure profitable operations. This argument is unconvincing. If venues do not pull in audiences, they can experiment with cheap last-minute tickets and spend more time and effort on marketing. This phenomenon provides a case for stimulating demand through education, vouchers, action plans and programming subsidies, not for extra supply subsidies. Another unconvincing argument is that culture contributes to employment, especially in an economy with a tight labor market. But the government is not obliged to provide orchestras, theater companies, etc. in sufficient numbers simply to employ all art graduates. Nevertheless the employment argument may be valid for certain professions (e.g., restoration) if one fears that otherwise valuable skills will be lost for future generations. However the fact that many artists are poor is not a valid argument, except insofar as one should assist all poor people regardless of their occupation.



The performing arts and other cultural goods are labor intensive and thus suffer from Baumol's cost disease.<sup>9</sup> The argument that this justifies government subsidy to the performing arts is unconvincing, since increases in productivity elsewhere in the economy gives rise to huge increases in purchasing power; if people value this labor-intensive culture, they will use their new riches to pay for it. With Cobb–Douglas preferences, the ratio of output of cultural goods to other goods dwindles away and prices of cultural goods rise at the rate of technical progress in the rest of the economy, but absolute levels of employment and output in the cultural sector remain constant. In other words jobs in the cultural sector are not destroyed since technological progress elsewhere boosts purchasing power just enough to maintain spending on culture despite rising prices of culture. In fact, there are two good reasons why the share of culture in national income may grow and the arts will flourish, even without more government support. First, food, drink, shelter and health are necessary goods while culture is a luxury good. With non-homothetic (e.g., Stone–Geary) preferences the budget share of the arts will rise over time as people grow richer. Second, an elasticity of substitution between culture and other consumption goods of less than unity will generate a budget share of culture that rises over time as well. With an inelastic demand for culture, sufficient revenue will be generated to compensate for rising costs.

In any case, Baumol's cost disease leads to offsetting trends. The rise in the relative price of labor-intensive cultural expressions causes a shift towards less labor-intensive culture. For example, a vocal artist may not be accompanied by a symphony orchestra but by a chamber orchestra or even by a computer. Technology may induce new economies of scale and substitution in consumption, as evidenced by the popularity of the CD, video, DVD or poster as a substitute for the concert, film, theater play or painting, respectively. The possibilities for an artist of operating on a world market have grown enormously with the advent of globalization and the Internet.<sup>10</sup> Note however, that Baumol's cost disease causes a shift from art for the elite to art for the masses, and may induce dumbing-down at the expense of diversity and experiment; if so, there may be a case for subsidizing innovation, experiment and diversity in the arts.

## 5. Quality versus popularity in cultural funding in Europe

### 5.1. *Different ways of funding culture*

There are three ways of supporting the arts. First, one can grant privileged positions possibly in exchange for satisfying certain requirements from the government. For example, public broadcasting organizations almost everywhere in Europe get first right

<sup>9</sup> See further in Chapter 15 by Brooks in this volume.

<sup>10</sup> Note that all of these developments can lead to the emergence of superstars; see the classic Rosen (1981) and Chapter 25 by Adler in this volume.

in the allocation of air space for TV and radio on top of a hefty subsidy; in return they program certain (minimal) amounts of information, education, arts and culture. Commercial broadcasters in Europe (apart from the UK) have more freedom. Another example is the fixed book price used in much of continental Europe (Austria, Denmark, France, Germany, Greece, Italy, the Netherlands, Portugal and Spain), where booksellers are granted monopoly power for each book that is published.<sup>11</sup> This leads to higher prices, lower sales and lower welfare, as shown in [Figure 1\(a\)](#). Critics also argue that the fixed book price is bad for low-income consumers and hurts the democracy of culture. On the other hand others stress that books are heterogeneous goods, produced and sold under monopolistic competition; they argue that the agreement allows for cross-subsidies from bestsellers towards less popular books and results in a more diverse supply of book titles and bookshops. This cross-subsidy argument is unlikely to be valid if bestsellers are highly price elastic and thus permit little monopoly power, while less popular books are price inelastic and allow a lot of monopoly power. In these circumstances the agreement induces substantial welfare costs. This may happen if bestsellers are easily digestible, require little time to read and have higher price elasticities while, say, poetry readings demand a lot of time and thus have lower price elasticities. In any case, a lot of energy and time is spent on defending the fixed book price agreement. This leads to wasteful rent seeking as discussed in [Tullock \(1980\)](#).

Many privileges and monopoly positions granted by European governments are eroded by technological change. For example, the fixed book price agreement may be undermined if people order books on the Internet through amazon.com and other virtual book suppliers. [Hjorth-Andersen \(2000\)](#) documents how the advent of new printing technology and new media have cut costs and led to an avalanche of new book titles in the Danish book market. Although it is much easier to get a book published for an author, turn-round of books has increased enormously and it is much more difficult to attract large groups of readers. In the broadcasting field, digital cameras and recording and editing equipment make low budget radio and TV possible, while the monopoly of public and commercial broadcasters is challenged by the arrival of the Internet, digital frequencies and narrowcasters. The Danish Dogma group as well as the recently killed film producer Theo van Gogh show that it is possible to make low-budget films with digital, hand-held cameras.

Second, the tax system and price subsidies can be used to grant demand price or supply price subsidies.<sup>12</sup> For example, the UK and Ireland do not allow fixed book price agreements but do have a zero VAT on books. Many countries in Europe exempt tickets for museums and the performing arts from VAT as well. The Netherlands has interest-free loans for buying visual arts. Almost all countries of Europe allow tax deductions for restoring monuments. [Figure 1\(b\)](#) shows that each of these price subsidies lowers prices at the gate and boosts output of culture. Since culture takes time to enjoy, the effective

<sup>11</sup> See [van der Ploeg \(2004\)](#) and [Chapter 21](#) by Canoy, van Ours and van der Ploeg in this volume.

<sup>12</sup> We abstract here from tax deductions to philanthropists, which are more prevalent in the US than in Europe.

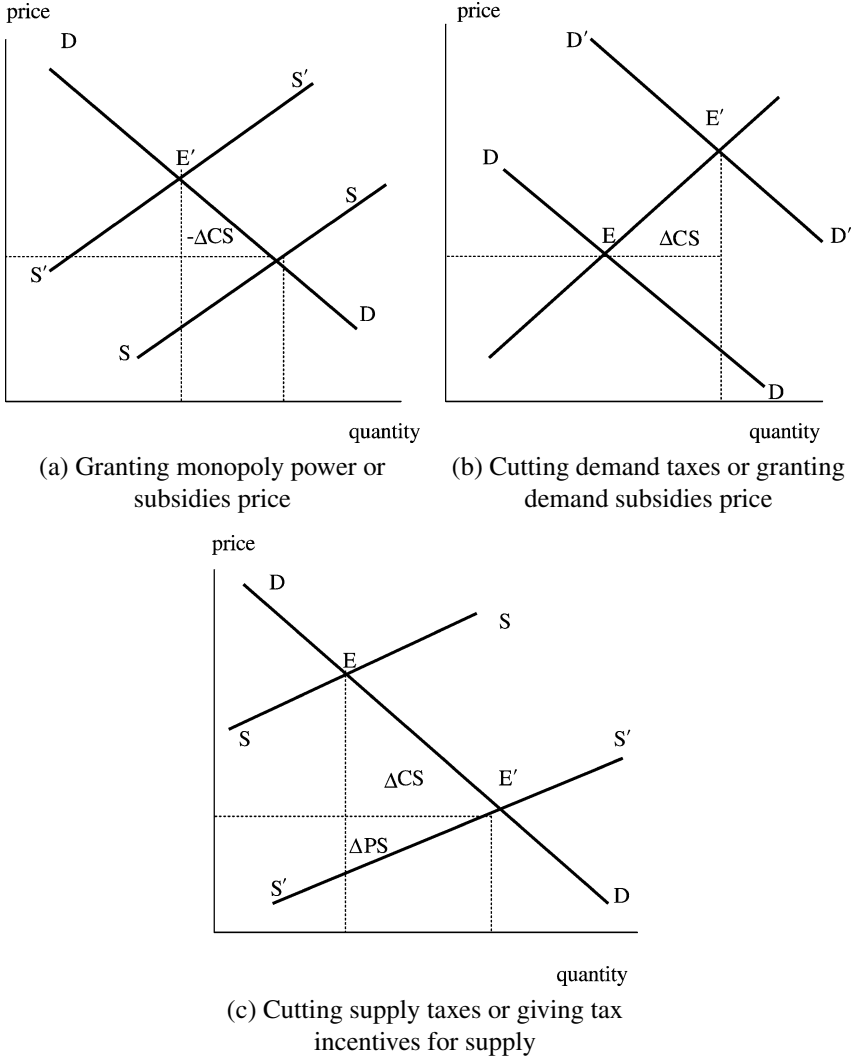


Figure 1. Incidence of different types of government intervention in the cultural sector. Key:  $\Delta CS$  and  $\Delta PS$  refer to the change in the consumer surplus and producer surplus, respectively.

demand elasticity is likely to be small. In that case, punters benefit a lot by having to pay substantially less while suppliers of cultural goods do not benefit much, especially if the price elasticity of supply is large. The potential welfare gains from the demand price subsidy, measured by consumer and producer surpluses, is then not very large and may well be negative if one takes account of the welfare costs of raising tax revenues to finance the subsidy. Figure 1(c) indicates that boosting supply through *supply price*

subsidies lowers prices and raises volume. For example, Ireland is well known for its tax incentives for producing films. This allows the production of more Irish films, and the lower prices for the public boost demand as well.

In principle, there is no difference in using demand or supply price subsidies in competitive environments. If demand is very inelastic and supply elastic, most of the benefit of the supply price subsidy is shifted to consumers while demand is not boosted very much. In that case, there is not much point in supply price subsidies. Although most cultural tax incentives apply across the board, it makes sense to avoid dead-weight losses by trying to target these instruments at the most needy projects; for example, interest-free loans might only apply to living visual artists; film incentives might require films to have substantial European contents; and deductions for monuments might only be obtainable if the property is indeed a valuable, historic monument. Also, if one wants to help low incomes with cheap opera tickets, one must avoid giving unnecessary subsidies to those pundits who can afford the true cost of opera. The welfare losses from these unnecessary subsidies are higher if demand and supply are more inelastic. One should avoid substitution, so cultural producers should not receive subsidies for activities they would have done anyway without subsidy.

Third, public support for the arts in Europe comes mainly in the form of *supply volume* subsidies allocated on the basis of advice of cultural experts as, for example, in the UK, the Netherlands, Belgium and Scandinavia. In France and Italy civil servants seem to play a bigger role. The positive effect on volume is partially offset by the negative effect of lower prices on supply, particularly if the demand elasticity is small and the supply elasticity is large. Supply volume subsidies can be better targeted and thus cause less dead-weight loss. However, such subsidies are allocated by civil servants, committees of experts or politicians, which causes wasteful lobbying and rent seeking. Committees of experts may also induce a bias towards 'arts for artists' and the elite, but extremes of the cultural spectrum may benefit. Some therefore have a preference for generic demand-price subsidies that lean with the market. The government then changes from participant to one who ensures a level playing field among cultural producers. This may induce median voter outcomes that reinforce middle-of-the-road culture. Indeed, Urratiaguer (2002), building on Throsby (1990) and Abbé-Decarroux (1994), finds that quality as judged by experts or drama critics or indicated by the artistic reputation of manager-directors are poor explanatory variables of the demand for subsidized theater in France. The main factor explaining demand seems to be the reputation of the theater company. This empirical work suggests that middle-of-the-road programming and bigger audiences may follow after a shift from supply-volume grants to demand-price subsidies.

## 5.2. *Programming diversity and funding of TV*<sup>13</sup>

The question of quality vs. popularity in European cultural policy is especially relevant in the field of television. To analyze the effects of financing modes on programming di-

<sup>13</sup> For a discussion of the rationale for the license fee, see O'Hagan and Jennings (2003).

versity and welfare, a framework of monopolistic competition is needed. The pioneering contribution of Spence and Owen (1977) shows that any private system of finance biases against programs that have a low own-price elasticity of demand ('minority taste programs') and that are expensive to produce. The price fails to fully reflect the average intensity of preferences for certain programs and, given fixed costs, such programs are not produced. This bias occurs with pay-TV, but is worse if programs are financed via advertising revenues. The latter produces fewer programs and leads to a less diverse menu of programs than pay-TV. As the cross-elasticity of substitution among programs increases, advertising is more likely to approach a second-best structure than pay-TV. Many countries of Europe have cable-TV. It is important to permit advertising on pay-TV, make sure the cable operator allows entry of other suppliers of programs, and ensure open entry with elastic supply of channels. Doyle (1998) extends the analysis and considers various regulatory instruments (such as direct instructions to channels on programming minimal amount of culture, arts and education and a tax on profits). Chae and Flores (1998) argue that certain characteristics of TV programs such as the extensiveness of the market, makes them more likely to be broadcast than shown on pay-TV. Bourreau (2003) analyses the trade-off between mimicking and counter-programming. He shows that profile differentiation is higher under pay-TV than under advertising support where competition on price is infeasible. He also shows that program quality is higher under advertiser support than under pay-TV. These studies ask the right questions and they also apply to other areas of culture. Of course, in contrast to the US, Europe finances a lot of public broadcasting. These studies justify governments doing this if they make possible the production and broadcasting of more expensive TV-programs that cater for special niche audiences. That is something the market can only deliver in the form of narrowcasting. To ensure the diversity of programs of public broadcasts, Germany relies on representative viewers' committees, the UK on detailed program requirements, and the Netherlands on diverse broadcasting corporations corresponding to liberal, catholic, protestant, social-democratic, libertarian, or evangelical pillars. The UK limits the entry of commercial broadcasters and forbids the BBC from advertising. In return for monopolistic advertising profits Channel 4 and ITV face extensive programming requirements.

## **6. The arm's length principle: Institutions matter**

### *6.1. Institutional structures*

Europe has three basic systems of allocating subsidies to cultural goods. The French and Italian systems are top-down and state-driven, so bureaucrats and politicians decide. This is not very transparent with plenty of scope for lobbying and friends' politics, with prestige projects standing a better chance and with unpleasant connotations of state art. At the other end is the British system; the government allocates a budget to the Arts Councils and leaves them the responsibility for allocating subsidies. The UK Arts

Councils operate as an independent Art Fund, so it is hard for the British government to set priorities and formulate criteria for allocating subsidies. The British system has clear advantages: less room for political lobbying and rent seeking and no danger of the state trying to impose its taste on the cultural sector. The disadvantage is that the government has little room to influence the direction of cultural policy.<sup>14</sup>

The Netherlands uses an intermediate system, which has an independent Arts Council that gives experts' advice on artistic merit and on how budgets should be allocated. The main difference from the British case is that it is the Dutch Minister of Culture and not the Arts Council that takes the final responsibility for the allocation of cultural subsidies. The government operates at arm's length and does not judge the artistic quality of different cultural activities and organizations, but there may be political and managerial reasons for deviating from the advice of the Arts Council. The government attempts to set the rules of the game well in advance, so announces in a White Paper the framework (priorities, criteria and budgets) that the Arts Council needs to adopt in their advice on the allocation of subsidies. The main advantage of this system is that the government can shape cultural policy without meddling in artistic judgements about individual cultural activities and organizations. However, with ministerial responsibility, there is maximum scope for political lobbying and rent seeking. Winners claim their subsidy and get on with making theater, opera, dance or whatever, while losers find plenty of opportunity to make their case in the media. Parliamentarians who want to distinguish themselves by associating themselves with the arts further reinforce this. The result is that it is easier to lobby for extra budget for the arts than for science, where the research councils allocate the subsidies and there is no ministerial responsibility.

The Arts Council system relies on committees, which are subject to well-known problems of committee decision-making [Tullock (1971)]. There is a danger that arts council meetings are long and dull, suffer from comment pollution and members merely proffer personal opinions without getting to grips with the issues concerned. It is thus important to give committees of cultural experts clear criteria and priorities. Debate concentrates on the victims, so cultural policy tends to be incremental and catering for the status quo. Klamer (1996) argues that politicians will lose the rhetoric arguments, since the cultural sector is much better able to make its case. It is thus crucial that the government clearly lays down the rules of the game as well as criteria and priorities in advance of the applications for subsidy and refrains from making artistic value judgements about individual cultural expressions. Otherwise, it may succumb to pressure to overthrow negative artistic judgements of the Arts Council, setting a legal precedent to other cultural organizations to have their subsidies reinstated in court.

Each system for the allocation of cultural subsidies has different drawbacks. Stimulating demand across the board carries little danger of rent seeking and lobbying and avoids protection of the status quo, but suffers from dead-weight losses and bias towards middle of the road culture. The French or Italian system suffers from top-down

<sup>14</sup> Delegation of cultural tasks to an Arts Fund or the Arts Council is known as the arm's length principle.

behavior, lobbying, rent seeking, political and bureaucratic favoritism, and a bias towards prestige projects. The British system does not suffer too much from lobbying and rent seeking, since it avoids the taste of politicians and bureaucrats influencing individual decisions. Both the British and the Dutch system are transparent. The Dutch system allows more room for the government to shape cultural policy without trying to influence artistic judgements, but suffers from political lobbying and rent seeking. The Dutch government has made a determined effort to give more room for diversity and innovation. It also has given a clear mission to the Arts Council. In any system it is important to keep the cultural sector on the ball by ensuring a competitive field with free entry. This requires equal access to cultural subsidies for everybody, but may also be helped by vouchers, performance contracts and benchmarks. One can stimulate cultural entrepreneurship by defining ends and performance indicators while allowing for the costs of targeting. The government should use the right sticks and carrots and avoid destroying intrinsic artistic motivation. Attention should be paid to how to manage cultural institutions and guidelines for cultural governance. Unfortunately, board members of cultural organizations are often older white men from the banking sector and big business. To be in touch with new and diverse audiences, one must look for more representative boards of supervisees.

## 6.2. *Delegation and control*

Recent theories of delegation and control can be used to explain whether non-elected bureaucrats or art experts or elected politicians should take responsibility for shaping cultural policy and allocating funds to cultural organizations. Hart, Schleifer and Vishny (1997) analyze whether public services such as prisons should be delegated to private business or under control of politicians. Dewatripont and Tirole (1999) discuss the use of advocates for policy makers, while Dewatripont, Jewitt and Tirole (1999) consider career concerns rather than explicit contracts as a motivation of government agencies. Maskin and Tirole (2004) study career concerns and non-accountable agents with intrinsic motivations. Alesina and Tabellini (2004) build on these contributions to explain whether tasks should be given to bureaucrats or politicians. From a normative perspective they show that politicians are better suited to particular tasks than bureaucrats if:

- differences in performances are due to effort rather than to individual talent or technical ability;
- preferences of the electorate and parliamentarians are unstable and uncertain, so flexibility is valuable and the task or mission of bureaucrats cannot be specified ex ante while politicians are accountable and can ex post be voted out of office;
- time inconsistency is unlikely to apply;
- politicians cannot strategically distort policy choices in favor of short-term objectives at the expense of long-term welfare;
- the stakes of organized interest groups are small and the legal system is poorly designed, so corruption is widespread; and

- side payments to compensate the losers are desirable and relevant or bundling of policies is crucial to obtain majorities in parliament.

Epstein and O'Halloran (1999) argue that the degree of delegation is chosen so as to maximize the benefit to politicians. Alesina and Tabellini (2004) argue that in these circumstances politicians want to keep those tasks that generate substantial rents, campaign contributions and/or bribes and all kinds of redistributive tasks. However, Fiorina (1977) points out that politicians delegate tasks that have a high risk of policy failure so that bureaucrats can be blamed. Monetary policy requires sophisticated skills, has relatively few distributional effects and suffers from time inconsistency problems. Also, the electorate's preferences for low inflation are stable and not controversial and recessions can be blamed on central bankers. As in Rogoff (1985), there is a clear-cut case for delegating monetary policy to an independent central bank. Similar arguments apply to regulation of utilities. In contrast, foreign policy should not be delegated to bureaucrats as preferences change a lot and it is difficult to ex ante specify the goals. Redistributive policy should not be delegated either, since politicians may want to capture some of the associated rents.

What are the implications of these studies for the delegation of cultural policy? Which tasks in the domain of cultural policy should be given to the Minister of Culture, parliamentarians, bureaucrats and art experts? Many of the insights of the theory of delegation apply. Being popular and winning elections in the short run motivates politicians; bureaucrats at the Ministry of Culture are motivated by career concerns and have a longer-term perspective; art experts are motivated by their standing and reputation with their peers in the cultural sector, and they also want to be seen to be independent of political pressure. Cultural subsidies generate substantial rents. Many politicians like to be seen to be a patron of the arts and around election time are happy to call on 'friends' in the arts to lend their theaters and to help during the campaign. But they are unsuited to decide on the exact allocation of funds; bureaucrats who make use of the advice of expert judgement on artistic quality do this better. Yet many politicians do not resist the temptation to interfere in the allocation of funds to please electoral lobbies and engage in redistribution from, say, the rich metropolis to culture-starved regions. This is particularly prevalent in a system with ministerial responsibility. There is often pressure from the parliament to focus on short-term benefits of cultural subsidies with no concern for the long run; building a new theater or opera house generates short-term prestige, but without making funds available for ambitious programming it is unclear whether it will contribute to a thriving cultural climate. The making of cultural policy suffers from very serious time inconsistency problems. This should not be mistaken for varying or unstable cultural preferences. The first five of the normative reasons described above thus strongly argue in favor of the arm's length principle in the making of cultural policy, and although it is true that the Minister of Culture can package cultural policies and make side payments to obtain a majority (see the sixth reason), the financial and cultural experts of the Arts Council can do that as well.

Nevertheless, delegation of the execution of cultural policy to an independent Arts Fund does not deny a role for the Minister of Culture. It is crucial that he or she specifies



a clear yet broadly defined mission for the Arts Fund, stating its priorities and criteria and also the available budgets; the mission could state, for example, that more funds should go to cultural education, vouchers, cultural diversity, the regions or international cultural policy, but not to which particular cultural organizations unless no judgement of artistic quality is involved. The Minister of Culture should not be allowed to change the rules of the game halfway through the complicated process of granting subsidies or to decide on subsidies for individual cultural organizations, but should ensure that the Arts Fund acts according to the objectives stated in the mission and only interfere if the Arts Fund deviates from the mission. Carpenter (2001) argues that the rise of the regulatory state offers bureaucrats the chance to decide as well as to implement legislation. This also applies to Arts Councils and Arts Funds. They can count on generous space in the press and are able to obtain support for their case. In the absence of a strong and visionary Minister of Culture, there is a danger that the Minister of Culture simply rubberstamps the proposals of the Arts Council or Arts Fund.

Granting cultural subsidies by the above means is not the only option. Another system is to allocate cultural subsidies by referendum as in Switzerland. For example, Schulze and Ursprung (2000) analyze a public referendum for the Zurich Opera House. However governments of the European Union have not used referenda in cultural policy.

### 6.3. Analytical example of delegation of cultural policy

The above argument can be analyzed formally as follows. Suppose policy output  $Y$  is given by the sum of the policy maker's unobservable talent or ability  $A \sim N(A_M, \sigma^2)$ , the policy maker's effort  $X$ , and noise  $\varepsilon \sim N(0, \tau^2)$ , that is  $Y = A + X + \varepsilon$ . Bureaucrats and experts want others to have a good perception of their ability, so their reward is  $E[E(A/Y)] = A_M + \beta E[A + \varepsilon + X - E(X) - A_M]$ , where  $\beta \equiv \sigma^2 / (\sigma^2 + \tau^2) < 1$  is the signal-to-noise ratio. If costs are convex in effort, i.e.,  $C(X)$ ,  $C' > 0$ ,  $C'' > 0$ , the Arts Council chooses effort level  $X = C'^{-1}(\beta)$  and puts in less effort if there is a lot of noise relative to the variance of talent. With imperfect monitoring, bureaucrats and art experts put in too little effort which causes a loss of welfare. A Minister of Culture is concerned with re-election, so his or her reward is  $\text{Prob}[Y > A_M + E(X) + \varepsilon]$ . This implies rational voters, who assume that the alternative to the incumbent is a politician with average talent and who in equilibrium puts in the same effort as the incumbent. It follows that the Minister of Culture sets  $C'^{-1}(X) = 1/\sqrt{[2\pi(\sigma^2 + \tau^2)]}$ . Imperfect monitoring ( $\tau > 0$ ) reduces effort of both the Arts Council and the Minister of Culture. However, more uncertainty about talent boosts effort of the Arts Council and cuts effort of the Minister of Culture. The Arts Council experts are motivated by career concerns and fully internalize benefits of higher expected ability. The Minister of Culture wishes to secure re-election. It follows that bureaucrats and art experts are better suited for tasks requiring special abilities or technical competence that not everyone (and a Minister of Culture definitely not) is likely to have. They are not more gifted than the Minister of Culture, but that they have stronger incentives to be gifted.

Assume now perfect monitoring and that cultural policy is concerned with two tasks: facilitating a diverse spectrum of cultural activities of high artistic quality ( $Q$ ) and promoting participation of large and diverse audiences ( $P$ ). The success of achieving these tasks is determined by effort  $X_i$ ,  $i = Q, P$ , and ability  $A$  of the person who makes cultural policy, so  $Q = A + X_Q$  and  $P = A + X_P$ . The costs of the efforts of the policy maker are additive and convex,  $C(X_Q + X_P)$ ,  $C' > 0$ ,  $C'' > 0$ . Ex ante the electorate is uncertain about whether it prefers more artistic quality or more participation. It has utility  $U[\delta Q + (1 - \delta)P]$ , where  $\delta = 1$  with probability  $p$  and  $\delta = 0$  with probability  $1 - p$ . The Minister of Culture formulates in a White Paper the mission or task of the Arts Council, say  $Y \equiv \lambda Q + (1 - \lambda)P$  with  $0 < \lambda < 1$ , and defends this in parliament before asking the Arts Council to use it. The mission of the Arts Council is not contingent on the realization of the preferences of the electorate  $\delta$  (i.e., it is an incomplete contract) and thus does not follow the whims of the voters. The experts on the Arts Council are concerned with the perception of their ability, hence their reward is  $E[E(A/Y)] = E[A + \lambda X_Q + (1 - \lambda)X_P - \lambda E(X_Q) - (1 - \lambda)E(X_P)]$ . If  $p > 1/2$ , it is optimal to ask the Arts Council to focus completely at artistic quality  $Q$  (i.e.,  $\lambda = 1$ ). The optimal strategy, as far as the Arts Council is concerned, is to set  $X_P = 0$  and  $X_Q = C'^{-1}(1)$ . By giving the Arts Council a fixed mission, society is stuck with the risk that they cannot respond if ex post cultural participation is important as well. The Minister of Culture can respond to changing preferences (i.e., the realization of  $\delta$ ) and devotes effort only to the task ex post preferred by the electorate. This advantage is not so strong if voters are not too risk averse and relative certain about their ex post preferences. In that case it is best to let the Minister of Culture set the mission of the Arts Council.

## 7. Subsidiarity, local cultural clubs and federalism

To determine the best level (local, regional, national or international) for the making of cultural policy, the subsidiarity principle employed by the European Union is useful. This implies that it is best to conduct cultural policy at as low a level of government as possible, because politicians at the local level are better informed about preferences of their electorate and are more readily rewarded and punished than politicians at a national or international level. However, culture thrives best in a competitive climate so it may be better to allocate part of the subsidies at a national or international level; this would ensure that the quality of, say, a local symphony orchestra is judged against symphony orchestras in other regions and would require cultural policy at a higher level of government. One reason for cultural policy at an international or national level is that some culture generates by its very nature positive cross-border externalities. In addition there may be substantial economies of scale; for example, small countries make do with one big opera house.

The theory of local public goods surveyed in [Rubinfeld \(1987\)](#) and [Scotchmer \(2002\)](#) points out the importance of migration between local jurisdictions. This has obvious

applications to the making of a cultural policy in Europe, in particular in the Länder of Germany. Tiebout (1956) argues that the provision of local public goods can be viewed as an efficient competitive market for private goods where people reveal their preferences for these public goods by ‘voting with their feet’. Efficiency requires perfect information, costless migration, and no cross-border externalities; it also requires each jurisdiction to be large enough so there are enough people to produce the local public good at minimum average cost. Free entry of people is a very strong assumption. Nevertheless, one does see more skilled and educated people migrating to bigger cities (London, Barcelona, Paris, Berlin and Amsterdam) with a bigger supply of local cultural goods, while provincial towns are starved of their potential audiences and find it more difficult to support high culture. The optimal size of a jurisdiction is small if preferences are homogenous for each region, cross-region spill-over effects are small and economies of scale are unimportant.

The club model of local public good provision trades off economies of scale versus externalities arising from congestion. Groups of people with similar incomes and similar preferences arrange themselves in a club. The Tiebout model allows citizens to have different incomes and uses a local property tax to finance local public goods. Cultural public goods have an impact on land and property values. Building a theater, opera house or museum makes neighborhoods more attractive places to live and attracts citizens who pay higher local taxes and push up land and house prices. An interesting direction of future research in cultural economics is to analyze the effects of local cultural infrastructure and activities on land and house prices and to estimate the demand for these goods in terms of local socio-economic characteristics and local and inter-governmental support. The use of micro data on cultural demand seems particularly promising. Such demand studies should focus on different attributes (e.g., international reputation versus popular appeal) of the local cultural goods. A crucial complication is how to aggregate preferences, since people differ very much in their taste for culture. Bergstrom and Goodman (1973) show that under certain strong conditions one can postulate a simple majority-rule political process by estimating demand for public goods based on a median-voter model.<sup>15</sup>

An important issue for Europe is whether local governments should balance their budget, what local public goods should be provided locally, and whether and how revenues ought to be shared between levels of government. The literature on fiscal federalism and multiple layers of government attempts to give answers to these questions. It is doubtful whether there are good economic reasons for conducting cultural policy at the level of the European Commission; a large part of the Commission’s cultural budget is devoted to prestige projects such as the Cultural Capital of Europe with little international spill-over effects, and much of the cultural exchange that is supported is hampered by costly bureaucratic procedures. Since not all countries of Europe adhere to the arm’s

<sup>15</sup> Rubinfeld (1987) discusses the empirical aspects of estimating such demand functions for local public goods and the Tiebout model in some detail.

length principle, it is difficult to avoid political intervention. Topics that are of a European nature (e.g., competition policy, trade policy) are typically not the responsibility of Ministers of Culture. The European Councils of the Ministers of Culture have very tiny agendas and serve mainly the purpose of symbol politics.

## **8. International cultural policy: Different approaches in Europe**

Most countries of Europe make an effort to develop international cultural policy – i.e. cultural policy involving or directed towards other countries, including developing countries. Typically, such policy is the responsibility of both the Ministry of Foreign Affairs and the Ministry of Culture.<sup>16</sup> In some cases the international positioning of culture is done with the departments responsible for media and sport (UK), education (Finland), sciences (Austria), sport and tourism (Ireland), or communication (France). In some countries cooperation between the two key ministries has resulted in a central institute for international cultural exchange (Denmark, the Netherlands) and in others these ministries work together on a continuing basis (France, Finland). In others the Ministry of Foreign Affairs takes the lead (Austria, UK and Sweden) and may make use of cultural institutes abroad (such as the British Council and the French, Swedish or Finnish Institutes). Sometimes the Ministry of Culture is primarily responsible (Hungary) and elsewhere there are discussions leading in that direction (Ireland, Sweden). In some cases international cultural policy is shaped in collaboration with the regions (UK). Most countries engage in bilateral and multilateral cultural treaties (in particular France with its many cultural specialists and *Alliances Françaises*) and join forces regionally (e.g., the Nordic Council, Finland with the Baltic countries). They also develop international cultural policy together with the Council of Europe, the EU and UNESCO.

Different definitions of international cultural policy formulation are used. Most countries use a very broad concept of culture (Denmark, Finland, France, Austria, UK and Sweden). Some of these countries tie international cultural policy closely with development policy (Sweden, Denmark and Finland) or with efforts to stabilize regions such as the Balkans (France, Austria and UK). International cultural policy often pays considerable attention to education (Finland, France and UK) and boosting creative industries (Denmark, Finland, Sweden, UK and France). Other countries employ a narrower concept focusing at the arts, heritage and libraries (Ireland, Hungary).

In all countries one of the goals of international cultural policy is to help home cultural organizations to travel abroad. With a certain degree of national pride countries want to promote their own culture abroad and gain mutual respect for each others' culture. These forms of cultural diplomacy may help to further other objectives of the Minister of Foreign Affairs, e.g., to help to win foreign contracts for home enterprises.

<sup>16</sup> See, for example, Annalin (2003) and the website [www.culturalpolicies.net](http://www.culturalpolicies.net).

The UK seems to have the biggest emphasis on economic interest. If this is the case, international cultural policy formulation is best seen as a side-kick of foreign affairs and has little to do with inducing a thriving international cultural exchange. From the point of view of the cultural sector it may be better to leave the initiative with the Minister of Culture. Then priority may be given to the highest artistic quality of culture offered abroad rather than popular expressions of culture geared towards other motives than the arts. It also seems more likely that priority will be given to fund and attract foreign expressions of culture that are not offered at home, e.g., Third World music, Berber culture, Bombay films, etc. (Denmark, France, Sweden and the Netherlands). This added diversity has intrinsic value for the cultural climate of a country. Another reason is that it encourages more competition among home producers of culture, which also leads to a more thriving cultural climate. From this perspective some countries host the best foreign young visual artists, film directors, architects, musicians, etc. if they are judged to be of higher merit than home-grown talent. Not all countries agree (France).

It is important to measure the effects of international cultural policy. Although this has not been done on a systematic basis, casual evidence can be assembled. For example, promotion of Swedish and French music abroad seems to have been successful; in France music sales abroad rose from 1.5 million items in 1992 to 39 million items in 2000, even though the French government probably judges this spectacular growth insufficient to counter the dominance of the Anglo-Saxon music industry. France also strives to promote its films, TV and radio programs abroad, again with some success. Denmark has shown considerable success abroad with the films of its Dogma group, while Finland obtains similar results in music, design and dance. It is not only Britain that is trying to promote a 'cool' image; Austria also wishes to change towards a more modern image, building on the popular music of Kruder and Dorfmeister, for example. Both countries make use of branding, corporate identities and joint marketing approaches to promote their culture abroad.

There is a heated debate about whether one should adopt protectionist cultural policy or not. Some countries find it better to empower artists, so they can compete on international markets (UK, the Netherlands). Other countries (led by France) disagree strongly and rationalize their protectionist tendencies with the goal of promoting cultural diversity. This goal is not controversial if it involves using special instruments to sell French films, books or music outside France, offering ethnic minorities the chance to express their culture, or bringing in more non-European culture within national borders. However, it is controversial if it boils down to keeping as much culture as possible from the US out of Europe. Although there may be a case for promoting own-language products and blocking foreign-language products, this is generally a dangerous route to take and is often in direct conflict with the objectives of liberalizing international trade.<sup>17</sup>

<sup>17</sup> For further discussion of these issues, see [Chapter 33](#) by Acheson and Maule in this volume.

## 9. Summary and conclusions: Some lessons for cultural policy-making in Europe

The governments of Europe believe that culture cannot be left to the whims of the market. They argue that culture is an experience good, takes time to enjoy and appreciate, and has strong social externalities. It is therefore unpalatable if only the cultural elite enjoy culture, because culture is like language, in essence a social phenomenon. If culture is non-excludable and/or non-rival, it is a public good. Culture has social value, but it also has existence value, innovation value, option value and intergenerational value. None of these factors are properly internalized by the market.

The challenge for most European policy makers is to boost high culture, including the synergies with low culture, and strive for a democracy of culture by making sure that more people have the competence to appreciate and understand cultural expressions. Democracy of culture is not concerned with dumbing-down high culture, but with making high culture available and accessible to broader, new and more diverse audiences. This enhances the public good character of culture. Furthermore, cultural expressions only become meaningful if they are confronted with a critical audience. Culture has to compete with many other leisure activities, which explains why cultural participation over the last quarter of a century has been stable while the level of education has risen substantially.<sup>18</sup> Culture in Europe has become a more normal, integral part of an omnivore diet of excursions, holidays, visits to leisure parks and even sport. High culture in Europe faces the danger over the next few decades of becoming marginalized as the leisure industry becomes even more professional, the young invest less and less in cultural competence, the circle of genuine art lovers becomes smaller, and culture is to be found more often only in places of entertainment. Hence, governments of Europe work on a cultural policy that stimulates cultural education for school children, develops a taste for cultural goods that make a lasting impression, brings high culture to where ordinary people are (parks, squares, pop temples, etc.), boosts the programming and demand for high culture, finances cultural expressions that are non-excludable and/or non-rival or has social, existence, option or bequest value, and provides room for experiment and research and development.

Governments should avoid dead-weight losses and substitution, on the one hand, and lobbying and rent seeking on the other hand. If subsidies to art producers are allocated by committees of experts (as on arts councils), there is a real danger of art for the sake of art and an erosion of the public support for the arts. The French model is less attractive than the Dutch or British model, because it gives too much influence to bureaucrats and politicians and prestige objects. The Dutch model allows a greater possibility for shaping cultural policy than the British model, but is more susceptible to lobbying. One should keep the committees of experts small, appoint the experts for periods that are not too long, and allow interested laymen on the committees as well. Since the task of deciding on the artistic merit and financial needs of different cultural organizations

<sup>18</sup> See van den Broek and de Haan (2000).

is a difficult task, and since lobbies in the cultural sector are strong and politicians tend to focus on the short term and act in a time-inconsistent fashion, recent theories of delegation suggest it is best to delegate the execution of cultural policy to an independent Arts Fund while the Minister of Culture is restricted to setting out a mission with clear priorities, criteria and budgets and making sure the Arts Fund adheres to the mission.

Demand-oriented subsidies such as vouchers, interest free loans, matching grants, public-finance partnerships, etc., stimulate the cultural sector to produce art for which there is a demand in society. The danger with this is that safe, boring, better-known art wins over innovative art. Hence, special facilities are needed to stimulate experimental art and research and development.

The performing arts are labor-intensive and, like education, health care and the police, suffer from Baumol's cost disease. Productivity growth in the arts lags behind other sectors, resulting in relative price increases and greater pressure for bigger subsidies. Still, the newly-found riches of technical progress elsewhere lead to extra demand for culture, especially if culture is a luxury good. There may be a shift from unique to reproductive art expressions, from small-scale to large-scale productions and from labor-extensive to labor-intensive productions. In addition, since the performing arts are good at lobbying, this may take up a growing proportion of the total culture budget at the expense of other cultural needs. This is somewhat of a paradox, since the performing arts are much less a public good than reproductive art forms, the more so if subsidies for the performing arts profit a highly educated, high-income minority living in the big cities. Hence, governments should be careful to provide sufficient funds for cultural causes that benefit current and future generations such as monuments, archaeology and museums.

Governments in Europe should avoid addiction to cultural subsidies and sustaining the status quo by allowing equal opportunities for newcomers and having a level playing field. In the Renaissance there was fierce competition between producers of art and also some degree of specialization. Today a healthy dose of domestic and foreign competition and more market-oriented support from governments can be important drivers bringing about a revival of the creative and innovative arts sectors of Europe.

## References

- Abbé-Decarroux, F. (1994). "The perception of quality in the demand for services. Empirical application to the performing arts". *Journal of Economic Behavior and Organization* 23 (1), 99–107.
- Alesina, A., Drazen, A. (1991). "Why are stabilizations delayed?" *American Economic Review* 81, 1170–1188.
- Alesina, A., Perotti, R. (1995). "The political economy of budget deficits". *IMF Staff Papers* 42 (1), 1–31.
- Alesina, A., Tabellini, G. (2004). "Bureaucrats of politicians?" CEPR Discussion Paper No. 4252, London.
- Annalin (2003). *Scan Internationaal Cultuurbeleid*. Annalin, Zandhuizen.
- Atkinson, A.B., Stern, N.H. (1974). "Pigou, taxation, and public goods". *Review of Economic Studies* 41, 119–128.
- Baumol, W.J., Bowen, W.G. (1966). *Performing Arts – The Economic Dilemma*. Twentieth Century Fund, New York.

- Benhamou, F. (2000). "The opposition between two models of labour market adjustment: The case of audiovisual and performing arts activities in France and Great Britain over a ten year period". *Journal of Cultural Economics* 24, 301–319.
- Bergstrom, T.C., Goodman, R.P. (1973). "Private demands for public goods". *American Economic Review* 63, 280–296.
- Bille, T.H. (1997). "The willingness-to-pay for the Royal Theatre in Copenhagen as a public good". *Journal of Cultural Economics* 21, 1–28.
- Bille, T.H., Hjorth-Andersen, Chr., Gregersen, M. (2003). "A tale of tables. On cultural public expenditures". In: Dueland, P. (Ed.), *The Nordic Cultural Model*. Nordic Cultural Institute, Copenhagen. Chapter 10.
- Bína, V. (2003). "Cultural participation in the Netherlands". Mimeo, Ministry of Education, Culture and Sciences, The Hague.
- Blokland, H. (1997). *Publiek Gezocht*. Boom, Amsterdam/Meppel.
- Bourdieu, P. (1979). *La distinction – critique sociale due jugement*. Editions de Minuit, Paris.
- Bourreau, M. (2003). "Mimicking vs. counter-programming strategies for television programs". *Information Economics and Policy* 15, 35–54.
- Carpenter, D. (2001). *Reputations, Negotiations, and Policy Innovation in Executive Agencies 1862–1928*. Princeton University Press, Princeton.
- Carson, R.T., Mitchell, R.C., Conaway, M.C. (2002). "Economic benefits to foreigners visiting Morocco accruing from the rehabilitation of the Fés Medina". In: Navrud, S., Ready, R.C. (Eds.), *Valuing Cultural Heritage: Applying Environmental Valuation Techniques to Historic Buildings, Monuments and Artifacts*. Edward Elgar, Cheltenham.
- Chae, S., Flores, D. (1998). "Broadcasting versus narrowcasting". *Information Economics and Policy* 10, 41–57.
- Cowen, T. (1998). *In Praise of Commercial Culture*. Harvard University Press, Cambridge, MA.
- DCMS (2001). *Creative Industries Mapping Document*, Creative Industries Task Force. UK Department for Culture, Media and Sport, London.
- De Swaan, A. (2001). *Words of the World. The Global Language System*. Basil Blackwell, Oxford.
- Dewatripont, M., Jewitt, I., Tirole, J. (1999). "The economics of career concerns: I. Comparing information structures & II. Applications to missions and accountability of government agencies". *Review of Economic Studies* 66, 183–217.
- Dewatripont, M., Tirole, J. (1999). "Advocates". *Journal of Political Economy* 107, 2–39.
- Doyle, C. (1998). "Programming in a competitive broadcasting market: Entry, welfare and regulation". *Information Economics and Policy* 10, 23–39.
- Epstein, D., O'Halloran, S. (1999). *Delegating Powers*. Cambridge University Press, Cambridge.
- Fiorina, M. (1977). *Congress: The Keystone of American Democracy*. Yale University Press, New Haven, CT.
- Frey, B.S. (2000). *Arts & Economics – Analysis & Cultural Policy*. Springer, Berlin.
- Frey, B.S., Pommerehne, W.W. (1989). *Muses and Markets: Explorations in the Economics of the Arts*. Blackwell, Oxford.
- García, M.I., Fernández, Y., Zoffo, J.L. (2003). "The economic dimension of the culture and leisure industry in Spain: National, sectoral and regional analysis". *Journal of Cultural Economics* 27, 9–30.
- Getzner, M. (2002). "Determinants of public cultural expenditures: An exploratory time series analysis for Austria". *Journal of Cultural Economics* 26, 287–306.
- Grampp, W.D. (1989). *Pricing the Priceless – Art, Artists and Economics*. Basic Books, New York.
- Harris, C., Vickers, J. (1985). "Patent races and the persistence of monopoly". *Journal of Industrial Economics* 52, 193–209.
- Hart, O., Schleifer, A., Vishny, R. (1997). "The proper scope of government: Theory and application to prisons". *Quarterly Journal of Economics* 111, 1127–1161.
- Hjorth-Andersen, Chr. (2000). "A model of the Danish book market". *Journal of Cultural Economics* 24, 27–43.
- Klamer, A. (Ed.) (1996). *The Value of Culture – On the Relationship between Economics and Arts*. Amsterdam University Press, Amsterdam.



- Krebs, S., Pommerehne, W. (1995). "Politico-economic interactions of German public performing arts institutions". *Journal of Cultural Economics* 19, 17–32.
- Lancaster, K. (1966). "A new approach to consumer theory". *Journal of Political Economy* 74 (2), 69–107.
- Maddison, D., Foster, T. (2003). "Valuing congestion costs in the British Museum". *Oxford Economic Papers* 55, 173–190.
- Maddison, D., Mourato, S. (2001). "Valuing different road options for Stonehenge". *Conservation and Management of Archaeological Sites* 4, 203–212.
- Maskin, E., Tirole, J. (2004). "The politician and the judge: Accountability in government". *American Economic Review* 94, 1034–1054.
- Noonan, D.S. (2003). "Contingent valuation and cultural resources: A meta-analytic review of the literature". *Journal of Cultural Economics* 27, 159–176.
- OCW (2003). *Cultural Policy in the Netherlands*. Ministry of Education, Culture and Sciences, Zoetermeer.
- O'Hagan, J., Jennings, M. (2003). "Public broadcasting in Europe: Rationale, license fee and other issues". *Journal of Cultural Economics* 27, 31–56.
- Peacock, A.T. (1969). "Welfare economics and public subsidies to the arts". *Manchester School of Economic and Social Studies* 37, 323–335.
- Peacock, A.T., Scott, A. (2000). "The curious attraction of Wagner's law". *Public Choice* 102 (1), 1–17.
- Pine, B.J., Gilmore, J.H. (1999). *The Experience Economy*. Harvard Business School Press, Boston.
- Prieto-Rodríguez, J., Fernández-Blanco, V. (2000). "Are popular and classical music listeners the same people?" *Journal of Cultural Economics* 24, 147–164.
- Rengers, M., Plug, E. (2001). "Private or public? How Dutch visual artists choose between working for the market and the government". *Journal of Cultural Economics* 25, 1–20.
- Robbins, L.C. (1963). "Arts and the state". In: Robbins, L.C. (Ed.), *Politics and Economics: Essays in Political Economy*. Macmillan, London, pp. 53–72.
- Rogoff, K. (1985). "The optimal degree of commitment to an intermediate monetary target". *Quarterly Journal of Economics* 100, 1169–1190.
- Rosen, S. (1981). "The economics of superstars". *American Economic Review* 71, 848–858.
- Rubinfeld, D.L. (1987). "The economics of the local public sector". In: Auerbach, A.J., Feldstein, M. (Eds.), *Handbook of Public Economics*, vol. 2. North-Holland, Amsterdam, pp. 571–646. Chapter 11.
- Schulze, G.G., Rose, A. (1998). "Public orchestra funding in Germany – an empirical investigation". *Journal of Cultural Economics* 22, 227–247.
- Schulze, G.G., Ursprung, H.W. (2000). "La donne e mobile – or is she? Voter preferences and public support for the performing arts". *Public Choice* 102 (1), 131–149.
- Scotchmer, S. (2002). "Local public goods and clubs". In: Auerbach, A.J., Feldstein, M. (Eds.), *Handbook of Public Economics*, vol. 4. North-Holland, Amsterdam, pp. 1997–2042. Chapter 29.
- SCP (2000). *Het Bereik van de Kunsten – Het Culturele Draagvlak* 4. SCP, The Hague.
- SCP (2001). *The Netherlands in a European Perspective*. Social & Cultural Report 2000. SCP, The Hague.
- Sintas, J.L., García Álvarez, E. (2002). "The consumption of cultural products: An analysis of the Spanish social space". *Journal of Cultural Economics* 26, 115–138.
- Spence, M., Owen, B. (1977). "Television programming, monopolistic competition, and welfare". *Quarterly Journal of Economics* 91, 103–126.
- Throsby, D. (1990). "Perception of quality in demand for the theatre". *Journal of Cultural Economics* 14, 65–82.
- Throsby, D. (1994). "The production and consumption of the arts: A view of cultural economics". *Journal of Economic Literature* 33, 1–29.
- Throsby, D. (1996). "Economic circumstances of the performing artist: Baumol and Bowen thirty years on". *Journal of Cultural Economics* 20, 225–240.
- Throsby, D. (2001). *Economics and Culture*. Cambridge University Press, Cambridge.
- Throsby, D., Withers, G.A. (1979). *The Economics of the Performing Arts*. Arnold, New York, Melbourne.
- Tiebout, C. (1956). "A pure theory of local expenditures". *Journal of Political Economy* 64, 416–424.
- Tullock, G. (1971). "Public decisions as public goods". *Journal of Political Economy* 79, 913–918.

- Tullock, G. (1980). "Efficient rent seeking". In: Buchanan, J.M., Tollison, R.D., Tullock, G. (Eds.), *Towards a Theory of the Rent-Seeking Society*. Texas A&M Press, College Station, pp. 3–15.
- UNESCO (1998). *World Culture Report*. UNESCO, Paris.
- Urratiaguer, D. (2002). "Quality judgements and demand for French public theatre". *Journal of Cultural Economics* 22, 185–202.
- van den Broek, A., de Haan, J. (2000). *Cultuur tussen Competentie en Competitie*. Boekmanstudies/SCP, Amsterdam/Den Haag.
- van der Ploeg, F. (1999). *Cultuur als Confrontatie – Uitgangspunten voor het Cultuurbeleid 2001–2004*. Ministerie van Onderwijs, Cultuur en Wetenschappen, Zoetermeer.
- van der Ploeg, F. (2003). "In art we trust?" *De Economist* 150, 333–362.
- van der Ploeg, F. (2004). "Beyond the dogma of the fixed book price agreement". *Journal of Cultural Economics* 28, 1–20.
- Wijnberg, N.M. (1994). "Art and technology: A comparative study of policy legitimation". *Journal of Cultural Economics* 18, 3–13.
- Wijnberg, N.M. (1995). "Selection processes and appropriability in art, science and technology". *Journal of Cultural Economics* 19, 221–235.

## CULTURAL POLICY: AN AMERICAN VIEW

DICK NETZER

*New York University, USA*

### Contents

Abstract	1224
Keywords	1224
1. Introduction	1225
2. A brief history	1225
3. The role of the arts councils	1229
3.1. What the arts councils do	1229
3.2. The National Endowment for the Arts	1231
3.2.1. Development of programs	1231
3.2.2. The “culture wars”	1232
3.3. State arts councils	1234
4. Direct support for the arts and culture	1235
4.1. Levels of support	1235
4.2. The case for direct subsidy by subnational governments	1238
5. Indirect support for the arts and culture	1240
5.1. Levels of support	1240
5.2. Criticisms	1241
6. Economic issues in support for the arts	1242
6.1. Market failure and willingness to pay	1242
6.2. Direct vs. indirect support	1242
6.3. Crowding-in or crowding-out?	1243
7. Heritage policy	1244
7.1. Public sector role in heritage protection	1244
7.2. Heritage tax incentives and regulation	1245
7.3. Heritage valuation	1247
8. Further research	1248
9. A concluding observation	1248
References	1249

**Abstract**

American cultural policy is made and executed by multiple governmental entities, with a large part of direct public expenditure coming from sub-national governments rather than the central government. Much the largest share of government support is provided by tax concessions for gifts to not-for-profit enterprises that are the predominant producers of the subsidized arts, and by tax concessions to private owners who agree to preserve their heritage buildings and sites. The multiplicity of government decision makers and the reliance on “arms-length” relations between government and not-for-profit private providers of cultural services results in cultural policy that appears to be and is incoherent, but is characteristic of American public policy in many spheres.

**Keywords**

cultural policy, arts councils, National Endowment for the Arts, culture wars, state and local governments, heritage tax incentives

*JEL classification:* Z10

## 1. Introduction

A standard definition of policy is “a definite course or method of action selected to guide and determine present and future decisions” (Merriam-Webster Dictionary Online, 2001). In that sense, there is not and never has been a clearly articulated cultural policy in the US. Instead, there is a very large number of distinct policy measures and proposals, often contradictory and poorly articulated. To some extent this is a consequence of the organizational pluralism that characterizes American government: the large and varied roles played by sub-national governments, and the pluralism within central and sub-national units of government. It is extremely rare that a single organizational entity within the federal government, and to a slightly lesser extent the state governments, has exclusive responsibility for a major area of public policy, from national defense to culture. A list of sources of direct federal government support of the arts and humanities in 2001 includes 17 different organizational entities as direct financial supporters of the cultural sector [Seaman (2002, p. 26)].

American cultural policy, then, comprises the collection of public-sector actions that affect the production and consumption of cultural goods and services. It also relates to how those actions have changed over time. But in fact, despite noisy controversy during the 1980s centering on the National Endowment for the Arts, cultural policy in the United States has been rather stable over the past century or more. There has been modest direct government financial support<sup>1</sup> of the arts and culture, and that support has increased at a moderate rate. Most of the increase has been in the form of steady increases in the budgets of cultural organizations that are government enterprises: federal, state and local government museums, non-commercial television and radio stations owned and operated by state and local government and by state universities, and so on. Public expenditures for heritage protection have also risen.

## 2. A brief history

The first museum in the United States was created by the New-York Historical Society, founded in 1804 in Federal Hall (where the first president of the United States was inaugurated in 1789). The Society’s collections included a large number of paintings by leading American painters, and 485 watercolors of birds by John Jay Audubon, a major figure in American painting in the first part of the 19th century. However, American cultural policy at the national level had an important but isolated founding moment in 1836. James Smithson, a successful British scientist, had drawn up his will in 1826, naming his nephew as sole heir. The will stipulated that, should the nephew die without heirs (as he did in 1835), the estate should go “to the United States of America, to

<sup>1</sup> “Direct government support” means grants of government funds, rather than tax preferences for private gifts to non-profit cultural organizations or exemption of such organizations from general taxes.

found at Washington, under the name of the Smithsonian institution, an establishment for the increase and diffusion of knowledge among men” [*Smithsonian Institution Libraries* (2002, p. 1)]. The motives behind this remain mysterious. Smithson died in 1829 leaving a legacy worth more than \$500,000. In 1836, Congress accepted the legacy, not without angry dispute, and eight years later established the Smithsonian Institution. The beginning collection was a scientific not artistic one, consisting of the thousands of plant and animal specimens brought back by the United States Exploring Expedition of 1838–1842 to the Pacific and the Antarctic.<sup>2</sup>

There were no further actions by the Federal government for decades to follow, apart from the commissioning of large paintings and sculpture (mostly bad but indubitably heroic) for the Capitol building that was under construction during the middle of the 19th century. For the country as a whole, the history really begins later in the century, outside the national capital, with the founding of the first large museums<sup>3</sup> and performing arts companies in major cities. Most of them were private not-for-profit entities, but there were also some state and local government museums (often historical museums) as well as Federal museums in Washington, the latter meant to entertain visitors to the capital<sup>4</sup> as well as to serve the educational missions customary to museums. Support for all these organizations was greatly reinforced after 1913 when personal income and death taxation began, with essentially unlimited deductions for “charitable gift deductions”, defined to include (among other things) cultural activities conducted by not-for-profit entities.<sup>5</sup> Government-owned cultural institutions have continued to receive private gifts of money and works of art. For example, the National Gallery of Art in Washington was created following a gift from Andrew Mellon, a billionaire who shortly before the gift was made was the Secretary of the Treasury; it is a Federally-supported museum and receives large private gifts of cash and works of art. Indeed, the differences between government-owned and private not-for-profit museums in regard to their financing and governance are sometimes obscure, and often matter little in practice.<sup>6</sup>

In the last three decades of the 19th century, art museums opened in a number of major cities, including New York, Boston, Brooklyn, Chicago, Cincinnati, Detroit,

<sup>2</sup> This expedition’s story is recounted in *Gurney* (2003) and *Philbrick* (2003).

<sup>3</sup> See *Meyer* (1979, pp. 17–36).

<sup>4</sup> The city was directly governed by the national Congress until well after World War II, with the House of Representative’s “District Committee” acting as the city council, approving the budgets of the city officials (who were Federal employees), determining the local tax rates, and deciding upon all policies applicable to the city.

<sup>5</sup> These tax preferences apply to a wide range of charitable gifts. Deductions are unlimited with respect to transfers at death (that is, bequests may reduce the taxable estate to zero). The value of gifts during one’s lifetime that can be deducted in any one year is limited to a stated percentage of before-tax income, but the non-deductible portions may be deducted in future years. The portion of the value of the gift reflecting capital appreciation over time is not taxed at the time of the gift at all.

<sup>6</sup> In sample surveys conducted by the American Association of Museums in 1999 and 2003, more than one-fourth of the respondents reported their governing authority as state or local governments, including state and local government colleges and universities. Nearly all others were operated by non-profit organizations.

Philadelphia, Pittsburgh, St. Louis, San Francisco, and Washington.<sup>7</sup> The organizers were well-off businessmen and the museums were private non-for-profit enterprises. But in most cases, the city government provided land and/or money to build and support the new museums. For example, the New York City government in the 1870s provided \$500,000 of the construction costs of the newly-established Metropolitan Museum of Art.<sup>8</sup> Interestingly, most charged non-trivial admissions fees, but had days with no charges at all. [Smolensky \(1986\)](#) demonstrates conclusively that the motivating rationale for establishing these museums was to exploit the educational value – or externality – of museums for ordinary residents of these cities. The organizers were persuaded that museum attendance was a worthy educational exposure for working people. Both state and local governments at this time and for a few decades afterwards also created historical and natural history museums, but usually with relatively little local government financing. But otherwise neither state nor local governments were significant as cultural patrons, save for what was done at the state government universities.

Direct government production of cultural goods and services has remained relatively small; indeed, it is relatively less important today than a century ago. There were two short-lived exceptions to this. In the Great Depression of the 1930s, a large number of artists were temporarily employed as artists, an almost accidental by-product of Federal work relief programs. A similar though much smaller such program existed for a few years in the late 1970s, as a Federally-financed counter-cyclical measure. In both cases, cultural activities were favored objects of publicly-financed counter-cyclical programs simply because at almost any time many cultural workers are not fully employed and have the low incomes that qualify them for these programs. Moreover, most cultural workers at any given time have well-developed projects that they can pursue, if financed. That means that the time between the passage of the law and the start of program operations tends to be short, relative to the lead time needed for most other projects.

By the early 1960s, establishment people persuaded politicians that direct monetary support from government, in the form of grants to not-for-profit cultural organizations, was essential to the local and national well-being. [Baumol and Bowen \(1966\)](#) appear to have made the single most persuasive intellectual contribution to this sea change in American cultural policy. They provided a market-failure case for subsidy. Their analysis of slow productivity increases in the performing arts, far less than increases in labor costs in the economy as a whole, forecast increasing financial distress for the sector, a forecast that was readily comprehensible by political leaders and the public at large. The amounts of money provided increased rapidly for a few years after 1965 when the National Endowment for the Arts and the National Endowment for the Humanities were created. Government funding of the Endowments peaked in 1980 (in current dollars), was cut sharply during the 1980s and has grown unevenly at modest rates since then.

<sup>7</sup> Some museums that opened during this period were financed and operated by individuals and families, such as the distinguished Isabella Stuart Gardner Museum in Boston. They eventually were converted into private, not-for-profit museums.

<sup>8</sup> [Jackson \(1995, p. 755\)](#).

Soon after World War II, the New York City government began to provide substantial support to a few major cultural institutions that were in serious financial difficulties. Initially, the rationale was that these institutions were located on City-owned land and/or in City-owned buildings. The first institutions so subsidized had immediate financial problems. The program evolved into one in which 30-plus institutions now receive annual support in the city budget to pay for some or all of the costs of building maintenance and operation, including security costs. The annual budget for this has been approximately \$120 million for some time.

The New York City local government arts agency example has been emulated in other American cities. The number of such agencies has been estimated to be more than 2000, an estimate that includes many private, not-for-profit (rather than governmental) agencies which receive little cash support from local government. The governmental agencies in the fifty largest cities spent \$350 million in the financial year 2001, of which \$306 million came from local government funds; New York City alone accounted for roughly 40 percent of the latter total.<sup>9</sup> Most of those agencies receive part of their public funding from the proceeds of special local taxes, such as hotel room occupancy taxes [Americans for the Arts (2001)]. In a National Endowment for the Arts publication, the local arts agencies are described as “primarily concerned with creating opportunities for artistry to occur” [NEA (1988, p. 394)] by sponsoring festivals and special exhibitions and providing housing for creative events. Some do provide operating support for long-existing cultural institutions, but at levels far below those in New York City. Also, there are cases in which local governments have issued bonds to finance, in part, new or expanded buildings such as concert theaters and museums.

In the 1960s, the New York State government created an arts council, roughly modeled on the UK Arts Council, to provide continuing support of operating budgets of large and small cultural institutions in the state; that concept became a reality in the 1970s when the state arts council budget was sharply increased. Meanwhile in the early 1960s American advocates for the arts also used the Arts Council of Great Britain as a model in pressing for national government provision of direct cash support of cultural institutions and cultural activities. They argued that the national government, which for years had been providing base-level financing of museums and other cultural activities in Washington, should now extend its financial support to most established and to-be-created performing arts companies and art museums. They proposed a new National

<sup>9</sup> It appears, from data in *Americans for the Arts* (2001), that total expenditure of public funds by local arts agencies in smaller cities in the financial year 2000 was modest. However, Cordes (2004, p. 205) suggests that arts spending by local government may have been equal to two to three times the total state government spending through state arts councils in 1999. That would mean local government support of between \$800 million and \$1.2 billion, implying that local governments *outside* the 50 largest cities were spending, per capita, two to four times as much as those in the big cities. This is highly implausible. It implies that the mean arts support (in total amounts of money) in smaller cities is higher than the mean for the cities ranking 2nd to 49th (that is, excluding New York). It may be that the source cited by Cordes includes non-governmental funds raised and spent by local arts councils, and also includes local arts councils that are entirely non-governmental.



Table 1  
Federal appropriations to the National Endowment for the Arts. Fiscal years 1965–2003<sup>1</sup>

Fiscal year	Appropriation in millions of current dollars	Appropriation in millions of 2000 dollars <sup>2</sup>
1966	2.9	12.5
1970	9.1	33.1
1975	80.1	210.8
1980	154.0	285.0
1985	163.7	234.8
1990	171.3	210.0
1995	162.3	176.2
2000	97.6	97.6
2003	115.7	109.2

Sources: National Endowment for the Arts, *2003 Annual Report*, p. 89; price deflator downloaded from the Bureau of Economic Analysis website, [www.bea.doc.gov/bea/dn/nipaweb/SelectTables](http://www.bea.doc.gov/bea/dn/nipaweb/SelectTables).

<sup>1</sup>Federal fiscal years ended December 31 until 1976; since fiscal years have ended on October 31.

<sup>2</sup>Converted to constant 2000 dollars by use of the price deflator for gross domestic product.

Endowment for the Arts (NEA), and a parallel National Endowment for the Humanities (NEH) to support other cultural activities. The NEA would also provide support for other fields such as literature and the visual arts in which individual artists were the dominant form of activity. In the event, the total amount of Federal money provided for the NEA was far too small to permit so ambitious a role, even in the late 1970s when the funding of the Endowment in real terms was at its peak (see Table 1).

Yet another federal government intervention into cultural matters was the creation and funding of the Corporation for Public Broadcasting in 1968 to help support the rapidly increasing number of non-commercial radio and television stations known as “public broadcasting”. A large percentage of the on-the-air time of these stations is devoted to cultural programming. Most public radio stations are operated by universities, while the predominant operators of the public television stations are non-profit organizations and state governments. The entire system had revenues of \$2.3 billion in 2002, 43 percent from government (mostly state and local) and 57 percent from individuals, corporations and foundations.<sup>10</sup>

### 3. The role of the arts councils

#### 3.1. What the arts councils do

Apart from the New York City program, the programs of the NEA, the state arts councils and the local government arts councils are similar in design. None has sufficient funds

<sup>10</sup> [www.cpb.org/pubcast#whopays](http://www.cpb.org/pubcast#whopays), November 2004.

to support a major share of the operating costs of cultural institutions, nor to make large capital grants for buildings or acquisition of works of art. Instead, they support selected projects that applicants propose, such as special exhibitions at art museums, the commissioning of a new work by an orchestra or opera or theater company, improvement of climate control systems, art festivals, projects designed to expand the audiences for whatever it is that the applicant does, or part of the costs of expanded touring programs. Those projects must be consistent with the applicants' overall programs, and the grants ordinarily provide only a portion of the full costs of the proposed projects.

In the early days of the NEA, one of its distinctions was making grants to individual artists in numerous fields such as literature, the visual arts, music composition, and choreography. As discussed further below, the grants were often controversial, because some resulted in the production or presentation of work that was offensive to audiences and/or politicians or work that was ridiculed by critics as well as audiences. As a result, grants to visual artists were ended in the 1980s and grants to individual artists in other programs were reduced or eliminated. However, some of the state arts councils continue to make grants to individual artists, often by subsidizing organizations that actually make the choice of grantee.<sup>11</sup>

In the operation of most arts councils (New York City is again the major exception, where most of the money goes to cultural organizations on a list that is specified in advance and seldom changes), the council and its staff have broad discretion in making grants. However, in addition to voting money to the state arts council to use at its discretion (subject to some general restrictions), state legislatures often specify grants in detail. These so-called "line-items" in state arts agency budget allocations usually reflect the preferences of the more important members of the state legislature, who are convinced that the state arts council will not look with favor on projects thought to be highly popular in a particular legislative district; these often involve grants to build a named building or concert hall. "Line-items" tend to account for larger percentages of the total amount allocated to the state arts councils in recession years when the regular allocation to the arts council has been reduced. For example, line-items comprised 5 or 6 percent of the total in the middle-1990s (in 14–16 states), but 13–17 percent in 1999–2001, when they appeared in 20 or more states [[National Assembly of State Arts Agencies \(2001b\)](#)].

It is to be expected that most of the interventions by government cultural agencies will be on the supply side. Inevitably, most requests for funding are made by suppliers or would-be suppliers rather than consumers. But in the presence of very strong demand, private entities will respond and Baumol's disease will be irrelevant, as the evidence from the popular arts shows. This is so obvious that long-term development of audience demand is invariably a policy objective. Nonetheless, Endowment grants have mostly

<sup>11</sup> Some state constitutions prohibit gratuitous gifts of money to individuals and for-profit firms, which requires the use of non-profit intermediaries if grants to individual artists are to be made. A policy decision not to make grants to individual artists avoids that complication and also the political difficulties experienced by the NEA.

gone to expanding supply. For example, both the Endowment and major foundations have made grants to support new compositions of serious music, in face of the reality that few compositions ever have even a single performance.<sup>12</sup> Nevertheless, some of the state arts councils have made, and continue to make, grants that foster demand, such as developing new ways of marketing performances and distributing tickets.<sup>13</sup>

### 3.2. *The National Endowment for the Arts*

#### 3.2.1. *Development of programs*

After its establishment in 1965, the new Endowment quickly became a marginal supporter of the well-established cultural organizations, and a more significant supporter for arts disciplines that had barely existed when the Endowment was created, notably dance. In 1960, there were very few permanent dance companies in the US, and only a few dance performances each year outside the few cities that had dance companies in residence, notably New York and San Francisco. The Endowment very early in its life created a “Dance Touring Program” that relied on consumer choice rather than on centrally-made decisions as to which dance companies to support. That program provided managers of venues such as theaters at colleges and universities with inducements to schedule more dance, by offering them a long list of companies from which to choose and by paying one-third of a stipulated fee for a dance company’s visit. Not surprisingly, since all the ticket revenue belonged to the theater operator, dance companies that filled the seats were invited back [Netzer (1987)]. The Endowment also greatly strengthened the regional non-profit theater companies and the regional opera companies, as the data on the growth of theater and opera companies, revenue and attendance in Table 2 show.

In addition to the relatively small amount of money spent by the Endowment each year, it has exerted some leverage. A grant from the Endowment could be used to certify – for other donors – that the organization had demonstrated its worthiness. At the very least, recognition by the Endowment lowered search costs for other would-be donors. Moreover, the Endowment in making grants does so for specific projects, not for the cultural organization as such. It also requires that its grants be matched by funds from other sources, which could increase private giving. Such “crowding in” is unlikely to have been very successful, because any funds that the recipient organization has can be applied to the match.<sup>14</sup>

<sup>12</sup> O’Hare (1980) and Felton (1978, 1980) examined possible ways of encouraging composers to devote more time to composing rather than in more remunerative work like teaching.

<sup>13</sup> One celebrated project was the New York State Arts Council’s TKTS ticket booths in Times Square which distributed unsold tickets to Broadway shows at half-price or less on the day of the performance.

<sup>14</sup> See further in Section 6.3 below.

Table 2  
Live performing arts, 1980–2000. (Revenue in millions of dollars. Attendance in millions. Data are for seasons ending in the year shown)

	1980	1985	1990	1995	2000
For-profit theater (Broadway and touring)					
Attendance	20.9	15.5	19.1	24.6	23.1
Gross ticket sales	327	435	649	1107	1175
Non-profit professional theaters: <sup>1</sup>					
Attendance	14.2	14.2	15.2	18.6	22.0
Total income	113.6	234.7	307.6	444.4	791.0
Earned income	67.3	146.1	188.4	281.2	466.0
Contributed income	46.3	88.6	119.2	163.1	325.0
Non-profit professional opera companies: <sup>2</sup>					
Attendance	5.5	6.7	7.5	6.5	6.7
Expenses	122.4	216.4	321.2	435.0	636.7
Symphony orchestras: <sup>3</sup>					
Attendance	NA	24.0	24.7	30.9	31.7
Total revenue					
From concerts	NA	168.6	253.3	368.6	481.5
Other earned income	NA	83.8	72.1	91.4	115.9
Grants and gifts	NA	167.3	226.4	312.2	464.7
Government grants	NA	42.2	55.6	55.5	59.9
Private grants and gifts net of fundraising costs	NA	125.1	170.7	256.7	404.8
Endowment income	NA	NA	52.1	76.2	136.6

Source: Statistical Abstract of the United States, 2002, Table 1214.

<sup>1</sup>Companies reporting to Theatre Communications Group, New York, NY. The number reporting almost doubled over this period.

<sup>2</sup>Companies reporting to Opera America, which increased from 79 to 98 over this period.

<sup>3</sup>Prior to 1995, covers 254 US orchestras; beginning 1995 covers 1200 orchestras. The additional orchestras are much smaller than those previously covered.

### 3.2.2. The “culture wars”

Another role of the Endowment was an unhappy one: it has been something of a lightning rod. Its existence demonstrated a Federal government commitment to the arts and culture, but also provided a target for criticism of expenditures that were more conspicuous than other Federal cultural outlays like support of the Federal museums. Such criticism led to what journalists have called “the culture wars” [Marquis (1995)] which began in 1981. They arose as follows. By the late 1970s, the NEA was making (necessarily very small) grants to virtually every non-profit organization in a given discipline each year. It had programs in disciplines that had been flourishing commercially with-

out subsidy such as crafts, folk arts, musical theater and architecture. In making specific grant choices, the NEA gave complete discretion to its discipline panels which were composed of artists in the field; this process was especially hazardous in making grants to individual artists. Until the program of grants to individual visual artists was ended, the Endowment was under continual political attack. Its survival was in question at numerous points in the 1980s and 1990s, often because of grants to allegedly “obscene” recipients. The Endowment’s defenders cited a variety of reasons why these attacks were wholly irrational and hurtful to the public good. But in 1990 the US Congress inserted in the law governing the Arts Endowment the requirement that the Chairperson of the Endowment ensure that the judges of grant applications take into consideration “general standards of decency and respect for the diverse beliefs and values of the American people”, a requirement upheld by the US Supreme Court.<sup>15</sup> This was greeted with outrage by the “artworld”. Rushton (2000) has laid out the various rationales for public funding of the arts and has concluded that the decency-and-respect requirement is consistent with each such rationale.

The major consequence of the long controversy was the end of most grants to the visual arts. Also, there have been important changes in the ways that the Endowment makes grant decisions. It is no longer organized along disciplinary lines – dance, theater, music, museums, etc. – but by major purpose.<sup>16</sup> A second major internal change was reaffirmation in words and in practice that final decisions are not to be made by panels and/or program directors: they are made by the person chairing the Endowment, as is done in the US Federal government generally.

There have been noisy controversies and political threats to the continuation of direct government support on occasion since 1980 at the state government level, as well as at the Federal level. The NEA’s funding in recent years has been equal to or less than its funding in 1980 in current dollars (see Table 1), but because of increased funding of other Federal agencies – including the Federal museums and the Corporation for Public Broadcasting – total Federal spending for the arts and culture has increased somewhat. However, its composition has changed considerably, in the direction of limiting the discretion of the agencies and eliminating subsidy of some cultural activities. One example is the end of Federal support for virtually anything that can be described as contemporary visual arts, which flourish despite this. Artists and their advocates often decry what they see as increasing reluctance to support anything unconventional on the part of the government agencies that provide direct cash support to cultural organizations. Almost any change in the staffing or internal organization of those agencies is viewed as another shift toward the conventional.<sup>17</sup>

<sup>15</sup> The Court noted that the Endowment’s limited budget requires that it reject the majority of applications and that some will be rejected on the grounds of content.

<sup>16</sup> The Humanities Endowment has been so organized from its inception.

<sup>17</sup> See *New York Times*, October 5, 2002.

### 3.3. State arts councils

Following the reduction in NEA funding during the 1990s, a considerable percentage of the funds nominally appropriated to the Endowment is now passed on directly to the state arts councils, which since the 1980s have existed in every state. The state arts councils became the new hope for sustained government support of cultural organizations, a hope based in part on the example of the New York State council, which set out to provide significant budgetary support to almost 100 “major” organizations, as well as marginal grants to small ones. Although the New York council’s funding has been far below its 1974 peak in real terms since then, state government support (from all states combined) has been much larger than National Endowment funding for some years. The total for all states reached nearly \$300 million by 1990, when the National Endowment’s budget for programmatic grants was less than \$200 million.

Nonetheless, the state arts councils have not been reliable base-level supporters of cultural organizations. This is because there is considerable volatility in their budgets. In the 13 financial years<sup>18</sup> between 1991 and 2003, a period including two recessions and the extraordinary mid-1990s boom, aggregate legislative appropriations to state arts agencies declined in six years, including the recovery year 1993 and the high boom year 1996. There were years of decline in every state, and in an average year in this period declines occurred in 19 of 50 states. Often the year-to-year declines were more than 25 percent. In a few cases, the budget was reduced to zero, and the agency rebuilt in later years on a much smaller scale. The average annual growth over the 13 years was 1.6 percent.<sup>19</sup> There had been similar although less pronounced instability in the 1980s [Netzer (1992a)].

There have been efforts to explain the large variation among states in state government support of the state arts councils measured in both total funds and dollars per capita terms. These efforts are summarized in Heilbrun and Gray (2001, pp. 283–285). The most exhaustive analysis was done by Schuster (1989). For the 14 states for which he could find data suggestive of the demand for funding – the size of audiences, the number of not-for-profit arts organizations and the number of resident artists in the decennial Census year – those variables explained 76 percent of the interstate variation. Dropping the audience size variable, the 50-state equation explained only 10 percent of the variation, and no independent variable was significant. Netzer (1992a) subsequently found that states in which state tax revenue is relatively high tend to spend more for the arts, but the relation was not a strong one. The variation seems idiosyncratic.

<sup>18</sup> Nearly all states begin their financial years on July 1.

<sup>19</sup> Calculated from data in the following publications of the National Assembly of State Arts Agencies, the trade organization for the state arts agencies: Legislative Appropriations Annual Survey, 2000; Legislative Appropriations Annual Survey, 2001; Legislative Appropriations Annual Survey, 2002; Legislative Appropriations Annual Survey, 2003; State Arts Agency Funding Sourcebook (1966–2001 data).

## 4. Direct support for the arts and culture

### 4.1. Levels of support

Total direct government support of the arts and culture in the year 2000 was \$3 billion as shown in Table 3. The “arts council” organizations accounted for a little over one-third of this sum, reflecting the substantial fluctuations in their funding at federal and state government levels over time. For the federal Endowments, this variability reflects the political controversy surrounding the Arts Endowment beginning in the 1980s, as discussed above. For the state arts councils, it reflects budgetary pressures confronting state governments during recessions. In contrast, direct government support for the museums that governments own and operate has increased steadily, as has support of public broadcasting; together these avenues account for more than 60 percent of total funding by the federal government and by the states. More than 80 percent of Federal support goes to those two activities and the heritage programs of the National Park Service. An obvious explanation for these figures is the reality that government departments have well established bureaucracies, located close to the places where budget decisions are made. These organizations are likely to have budget increases over time that parallel the increases enjoyed by other departments and agencies of that government, unlike arts councils that give public money to “outsiders”.

A particular avenue of direct government support for the arts and culture is support for non-profit organizations. Tables 4 and 5, based on different sources of data, show the relative importance of government grants to non-profit organizations that produce cultural services. Table 4 is based on the Economic Census of 2002.<sup>20</sup> Table 5, showing data for the year 1999, is based on inspection of returns filed with the Internal Revenue Service annually as a condition of Federal tax exemption. There are large differences in the data in the two tables, especially for museums.<sup>21</sup> But both show that government grants are small percentages of total revenue for all types of non-profit cultural organizations.

There is one strategic intervention in cultural policy with a very high benefit–cost ratio. It is the Federal government’s indemnity insurance for exhibitions of artworks and artifacts borrowed from other countries by US museums, or lent by US museums for exhibitions in other countries. The insurance may also include works lent by US museums for exhibitions within the US if the exhibitions also include works borrowed from other countries. In other words, the program covers virtually any exhibition that includes valuable works. The total amount of this insurance that may be outstanding at any one time

<sup>20</sup> The Economic Census is conducted every five years; in categories in which non-profit organizations are important, data for firms not subject to Federal income tax are shown separately. Although filing returns for this Census is not compulsory, there are indications that compliance rates are high, for both tax-exempt and taxable firms.

<sup>21</sup> In the periodic financial surveys conducted by the American Association of Museums, the response rate has been quite low. It appears that many large museums do not respond, even to surveys that seem to be in their interests.

Table 3  
Direct government support of the arts and culture, 2000 (in millions of dollars)

	Federal	State	Local	Total
Total, all levels of government	1158	1080	820	3058
Federal, total, all listed programs	1158			
National Endowment for the Arts	85			
National Endowment for the Humanities*	103			
Institute of Museum and Library Services	24			
National Park Service, heritage sites and buildings	200			
Smithsonian Institution*	374			
National Gallery of Art	72			
Public broadcasting*	300			
State governments, total, listed programs		1080		
State arts councils (State funds only)		400		
State government museums, estimated*		180		
Public broadcasting*		500		
Local governments, total			820	

*Sources:* Most of the data were found on the web sites of government agencies, in agency annual reports, and in the annual statistical reports of organizations serving state government arts councils and local arts councils. The estimates for state and local government museums (by the author) were based on the sample surveys of the American Associations of Museums, using the differences in budget size that those surveys report between non-profit museums and state and local government museums and the universe data for the non-profit museums in Cordes (2004). AAM (1999); AAM (2003); National Assembly of State Arts Agencies (2001a, 2002, 2003); Davidson (2002).

\*Some of these funds are used to support activities that cannot, strictly speaking, be described as "cultural", such as science museums and news and public affairs programming on public broadcasting stations.

Table 4  
Sources of income of non-profit firms, 2002 live performing arts and museums (dollar amounts in millions)

	Live performing arts <sup>1</sup>		Museums	
	Million \$	Percent	Million \$	Percent
Earned income <sup>2</sup>	5199	60.2	3083	39.4
Government grants	435	5.0	1105	14.1
Private gifts and grants	2437	28.2	2609	33.3
Investment income and other	577	6.6	1032	13.2
Total income	8641	100.0	7829	100.0

*Source:* United States Census Bureau, 2002 Economic Census, Performing Arts, Spectator Sports and Related Industries, EC02-711-01 (August 2004); and Museums, Historical Sites, and Similar Industries, 2002, EC02-711-02 (July 2004).

<sup>1</sup>Includes both performing arts enterprises and what are called "promoters", mostly non-profit owners of theaters and other performance venues.

<sup>2</sup>Includes membership fees, admission charges, sales of merchandise, and rental fees and royalties on use of the artistic property owned by the non-profit organization.



Table 5  
Government assistance to non-profit cultural organizations, 1999 (dollar amounts in millions)

	Current revenue	Government aid	Percent
All non-profit cultural organizations	13,173	1041	7.9
Performing arts	4087	289	7.1
Museums	2564	246	9.6
Visual arts	29	5	7.7
Media and communications	6404	494	7.7
Music schools	832	171	4.2
Other	22	2	11.9

Data from annual reports filed with the US Treasury Department as a condition of exemption from Federal income tax, and summarized in Table 8.1 of Cordes (2004).

is \$8 billion. The maximum insurance coverage for a single exhibition is \$800 million, and the insurance deductibles are very low, so that as much as 99.9 percent of a large loss is recoverable. The purpose of the program is to permit the borrowing of works for which no commercial insurance cover can be had and to permit very large savings in insurance premiums in cases where commercial insurance coverage would have been feasible. From 1975, when the program began, until October 2003 these savings are estimated to have been \$157 million for 713 exhibitions. In 2003, 62 exhibitions opened with this indemnification.<sup>22</sup> It is a low-cost intervention, given the physical care that is taken in mounting any major exhibition. There is no dispute about this program.

The United States is unusual in the very small extent to which the state pays for the professional training of artists. Logically this could be expected to lead to a failure to exploit available talents. Although some state government universities have such programs, a considerable number of the most illustrious programs are in private universities and conservatories with very high fees that surely deter some people from pursuing the path into an artistic career. In addition, some of the private conservatories and dance schools, including the most illustrious, regularly live close to the edge financially, and occasionally there are failures. Government assistance to the conservatories to assure their continued existence would be another not very high-cost intervention.

The state universities and colleges play an important part in developing audiences for the performing arts through their role as presenters of touring companies; this role is essential to the financial survival of large numbers of performing arts groups, especially dance companies, for some of which touring may account for 95 percent of their earned income. It is inconceivable that the development of a national audience for dance, which was limited to a very few cities before 1960,<sup>23</sup> could have occurred without the existence

<sup>22</sup> [www.nea.gov/grants/apply/Indemnity/Intro.html](http://www.nea.gov/grants/apply/Indemnity/Intro.html).

<sup>23</sup> Kendall (1983) recounts the progress of dance in the US in the course of celebrating the contributions of the Ford Foundation to that progress.

of numerous institutions capable of both physical presentation and financial guarantees to fledgling dance companies.

#### 4.2. *The case for direct subsidy by subnational governments*

In American politics, the case for direct subsidy for the arts is made almost entirely on narrow utilitarian grounds, with little attention to the positive externalities that cultural economists write about. The argument typically is that subsidies for arts and culture by state and local governments have very large and favorable impacts on local economic growth. Under some conditions this can be true, notably when the cultural attractions are numerous and famous and people actually travel to that city mainly or partly because of them. But there are few instances of a Venice or Paris among the hundreds of American places to which the argument is said to apply. In many American cases, the economic impact argument is no more than dissembling in a noble cause [Seaman (1987)]: the cultural attractions are substitutes for other local consumer goods rather than exported services. Even for New York, where a plausible economic-impact case has been made for cultural subsidies [Port Authority of New York and New Jersey (1993, 1994)], extravagant claims sometimes are made regarding the alleged local economic impact of government cultural subsidies. A letter to the editor from the head of the main arts advocacy group in New York City that was published in *The New York Times* [Bourscheidt (2002)] asserts that “In return for \$100 million in city support, non-profits arts groups generate \$5 billion in economic activity”. Studies of the local economic impact of the arts in the US usually make generous claims, but few claim a 50:1 multiplier.

Nevertheless, local cultural goods can be a significant contribution to the local quality of life, a proposition that makes the dominance of state and local financing of cash subsidies to culture, in contrast to the minimal role of the Federal government, congruent with American values rather than an indication of American hostility to the arts and culture. A considerable fraction of local government subsidy is used to support the most important local cultural institutions, including venues for first-rank visiting artists as well as local artists and supporting players.

Advocacy groups sometimes note an important and plausible goal of increased subsidy, namely to enable individual artists to devote less time to non-artistic labor and more time to artistic labor. It is well known that most artists hold multiple jobs. For example, in a NEA survey of choreographers, Netzer and Parker (1993, p. 47) found that the mean number of hours per week that was devoted to the first of two or more jobs other than choreography was 23.6 hours, and there was a negative relation between working in non-choreography jobs and the number of new works produced.<sup>24</sup> Throsby (1992, 1994) has formally modeled artists’ three-way labor market choices,

<sup>24</sup> Such supply-side intervention seems appropriate for dance. Each season mounted by any respectable American dance company is expected to include some new works. Many dance companies include European cities in their tours (which are essential to their financial stability), and managers of these companies appear to be convinced that the tour must include new works if it is to be invited to perform in subsequent

among artistic labor, non-artistic labor and leisure. Any policy measure that increases the hourly earnings from artistic labor will stimulate a shift in the supply curve away from non-artistic labor in favor of artistic labor.

Another cause popular with advocates is less plausible. It is asserted that because uncompensated and voluntary cultural activities, notably in music and theater, appear to be large and increasing, direct cash government subsidies to professional cultural activities should increase. The reasoning appears to be this: if the sector flourishes with so little subsidy, imagine what the sector would be with more subsidy [McCarthy et al. (2001)]. However, it seems equally plausible that increased subsidies might result in a reduction in uncompensated voluntary labor, as has been the case with many activities in the US in the past thirty years. A notable example is the sharp decline in uncompensated work by parents and others in the state schools, accompanied by a large increase in paid staff such as school crossing guards.<sup>25</sup>

A new version of Baumol's "cost disease" is a recent addition to the advocates' case for more subsidy. McCarthy et al. (2001) forecast declines in the output of professional performing arts groups in middle-size American cities because of increasing costs of production and sluggish growth in direct government subsidy. The analogy is with the recent history of professional league sports in the US and some other countries. In these sports the top-ranked divisions and teams are extraordinarily successful enterprises, in large part because of television revenue, while lower-ranked divisions and teams languish financially and often collapse; those that survive do so by means of subsidies from the more successful enterprises and sale of athletes' contracts to the successful teams. In reality, in the performing arts in smaller cities the financial problem is much less severe because the companies are less ambitious in their programming and often operate with a mixture of paid professional staff and unpaid amateur artists, and volunteer non-artistic staff. That mix of paid and unpaid, professional and amateur labor was the operating mode for most performing arts companies outside the largest US cities as recently as 1960.

However questionable some of the arguments made by advocates, it is entirely legitimate for those with strong preferences for cultural goods and services to attempt to persuade others that the public goods generated by subsidies to cultural activities and organizations are real and large. As Bruno Frey has long argued, when the political decisions are made at the local rather than central government level, there can be more confidence in the process of public choice, especially when the decision process is by direct individual choice. By voting positively on the issue in an election, voters are indicating that they believe the various external benefits such as option value, bequest value

years. In contrast, American orchestras and opera companies rarely perform new works. O'Hare (1980) calculated that most new compositions are never performed; Heilbrun (2001) documents the declining diversity of American opera companies over the period 1982/83–1997/98.

<sup>25</sup> The question of substitution of paid for voluntary labor is examined in numerous papers dealing with "crowding out" in the provision of health care and social services.

and prestige value are real enough to warrant payments of taxes. Frey's (1999) argument suggests that one of the strengths of American cultural policy is that most direct subsidy to the cultural sector is provided by state and local governments rather than by the national government. The national government's main roles are to handsomely support the operations of national cultural institutions in the nation's capital, to provide physical protection for a large number of heritage sites and buildings, and even more importantly, to provide indirect support through tax preferences to private givers, thus leaving decision-making about each gift highly dispersed among thousands of private parties.

## 5. Indirect support for the arts and culture

### 5.1. Levels of support

Indirect support of the arts and culture through the tax system is of considerably larger magnitude than direct subsidy. "Private contributions to arts, culture and humanities"<sup>26</sup> have been in the \$10–11 billion range in recent years, as shown in Table 6. A conservative estimate is that the cost of these gifts to the public sector, in terms of the reduced revenue from income and wealth-transfer taxes, is \$4 billion annually. Income tax credits for preservation of privately-owned historic structures amounted to \$9.2 billion over the years 1978–1997 [Listokin, Listokin and Lahr (1998, pp. 433–439)]. Also, there is the exemption from local property tax of land and buildings owned by not-for-profit cultural organizations, an exemption dating back to the early 19th century [Diamond (2002)]. This exemption is not important for most types of cultural entities, but only for those that own valuable land and buildings. Logically it should encourage cultural organizations to over-invest in land and buildings, but museums are the only type of not-for-profit cultural enterprise for which land and buildings are crucial and expensive factors in the organization's production function.

Not-for-profit entities are exempt from corporate income tax in the US by definition, i.e. any excess of receipts over expenditure in a given year cannot be appropriated by the "owners" of the enterprise. There is a Federal tax on "unrelated business income" of not-for-profit organizations, but this is defined in a manner that results in infrequent and very small payments by only a small minority of not-for-profit entities. In countries with value added taxes, the treatment of not-profit-organizations' receipts under VAT is of considerable importance. The US counterpart – state and local government taxes on retail sales of goods and services – are imposed at rates that are considerably lower than most VAT rates, and the receipts of not-for-profit cultural organizations from admissions charges are never subject to these sales taxes; their receipts from sales of food and drink and other merchandise also largely escape sale taxes. Thus, the tax exemption that matters is the ability of donors to deduct the value of gifts from income and wealth-transfer taxes.

<sup>26</sup> That is a considerably broader category than "non-profit performing arts and museums".

Table 6  
Private philanthropy for the arts and culture, 1980–2000 (in millions of dollars)

Source and allocations	1980	1985	1990	1995	2000
Total funds, for all uses	48.6	66.9	101.4	124.0	210.9
Individuals and charitable bequests	43.6	57.4	88.6	106.1	176.0
Foundations <sup>1</sup>	2.8	4.9	7.2	10.6	24.6
Corporations	2.2	4.6	5.5	7.3	10.3
Gifts to arts, culture and humanities					
Total <sup>1</sup>	3.2	5.1	7.9	10.0	11.5
From foundations	0.4	0.7	0.6	0.8	1.8
From corporations	0.1	0.2	0.2	0.7	1.0
From individuals	2.8	4.4	7.3	9.2	9.7

Source: Statistical Abstract of the United States, various years; in some places, estimates have been made by the author.

<sup>1</sup>Roughly 5% of all charitable contributions are not allocated are by recipients in the data source.

## 5.2. Criticisms

The income and wealth-transfer tax preferences for charitable gifts date from 1913, when Federal government taxation of income and wealth transfers began. They have been criticized, mainly on the grounds that deductibility for tax purposes transfers decision-making power regarding the allocation of resources in matters of public concern from government to givers, in proportion to the size of their gifts. Gifts to the arts and higher education are notoriously dominated by gifts from rich donors, unlike gifts to churches.<sup>27</sup> However, the prevalent American view of subsidy in this form appears to be that it is the ultimate decentralization of choice regarding who and what is subsidized. Americans generally see this not as the rich deciding on who and what receives subsidy at government expense, but as the only practical way to shift the decision-making from bureaucracies and politicians to a vastly larger number of individuals. Moreover, because gifts to culture are so heavily concentrated among the rich, there is some measure of progressivity in incidence, probably more so than is the case with respect to direct cash subsidies from government [Netzer (1992b)].<sup>28</sup>

<sup>27</sup> See Feld, O'Hare and Schuster (1983) and Chapter 36 by Schuster in this volume.

<sup>28</sup> In the volume that includes Netzer (1992b) two of the discussants, Estelle James (pp. 244–255) and Henry Aaron (pp. 237–243), argue that several papers including the Netzer paper may overstate the redistributive effects of income and wealth tax deductibility by not explicitly taking into account the likely low price elasticity of demand for some of the services subsidized by charitable contributions (notably for cultural goods and services) on the part of lower-income households. The subsidies probably do not increase their consumption of cultural goods by very much. The proportion of audiences that are low income households is, therefore, a poor measure of the distribution of the benefits from tax-deductible gifts. However, Aaron notes that in the performing arts, the wage subsidies to artists financed by tax-favored charitable contributions probably are real and significant.

## 6. Economic issues in support for the arts

### 6.1. Market failure and willingness to pay

The great majority of Americans tolerate, even take pride in, occasionally consume, and often join as amateurs in the production of, cultural goods and services. But they evidently do not consider increases in the supply of professionally produced cultural goods to be sufficiently important to warrant them paying more than infinitesimal amounts of taxes to directly support culture. This attitude is comfortable in part because cultural output has been increasing for some time at a rate equal to or greater than the increase in total consumer expenditure, in the absence of large subsidies from tax-derived public revenue as was seen in Table 2.<sup>29</sup> Attendance data for performing arts organizations and museums are an indication of this increase in output. Moreover, surveys of “public participation in the arts” done by the US Census Bureau for the National Endowment for the Arts every five years beginning in 1982 suggest that between one-third and one-half of adult Americans who are university graduates attend at least one performance a year of both plays and classical music and well over half visit an art museum at least once. Even for opera and ballet, that proportion is 10–15 percent [US Census Bureau (2001, Table 440)].<sup>30</sup>

The essential argument for subsidy to the cultural sector in one form or another is that there are a number of classes of external benefits that will not be realized in purely market transactions [Frey and Pommerehne (1989, pp. 16–30)]. The magnitude of the external benefits and how those are valued by the individual or the collective decision-making entity – a country, province, or city – should determine the extent of subsidy. Frey suggests that the impact of government subsidy on creativity and the quality of artistic output is affected by “what kind of state” provides the subsidy, arguing that a decentralized and democratic state is likely to better than a centralized and authoritarian state [Frey (1999, pp. 72–75)]. Frey and Pommerehne (1995) make a case for direct democracy – voting in referenda – using the Swiss experience. The alternative, given that decision by referenda is unusual except in Switzerland and some places in the United States, is the contingent valuation method, using a variety of indirect approaches.

### 6.2. Direct vs. indirect support

In the *Journal of Cultural Economics*, November 2003, eight papers evaluate contingent valuation in the arts itself. Noonan presents a meta-analysis of more than 100 such

<sup>29</sup> Relative prices of tickets for the listed performing arts appear to have risen considerably over the period, which suggests that the price elasticity of demand is low, while the income elasticity is moderately high. However, Felton (1992) found, in an analysis of 50 performing arts companies, a considerable degree of variation in price elasticities among the companies and some suggestion that higher prices may be thought by many consumers that they are indicators of higher quality, as Throsby (1983) proposed.

<sup>30</sup> There has been some skepticism about the response to the surveys, to the effect that they seem benign exaggerations. Earlier surveys showed even higher percentages and seemed inconsistent with estimates of total paid attendance at professional performances [cf. Netzer (1986, 1992b)].

studies. In concluding papers, Epstein (2003) and Throsby (2003) tell readers that despite the drawbacks of the approach, its use is a “regrettable necessity”. Tax reductions to encourage individuals to make gifts of money and art works arguably reduce the error inevitable in contingent valuation methods. In arithmetic terms this must be true, because the value of the gifts themselves represent direct valuation of externalities by donors; it is only the portion of the gift that represents tax forgiven by the state that requires contingent valuation. The lower the marginal tax rate confronted by individual givers, the smaller the amount of externalities whose economic value is unknown.

There is an extensive literature about American “tax expenditures”, that is, using tax concessions rather than direct spending of public money to subsidize activities considered desirable in some sense, beginning in the 1970s [Surrey (1973)]. Much of it is not very relevant to the economics of art and culture. Nevertheless for at least 50 years public finance economists have argued that the broader the coverage of income and/or wealth by a tax, the less the excess burden or deadweight loss – i.e. the economic distortion caused by payment of the tax itself – because the excess burden is a function (most likely exponential) of the marginal tax rate confronting taxpayers [Auerbach (1985)]. Exemptions, deductions and other exclusions from the tax base must cause marginal tax rates to be higher. That would suggest that generous deductions allowed in calculating tax liability will increase the marginal tax rate and thus the excess burden.<sup>31</sup> However, if the alternative is outright subsidy, the marginal tax rate must be even higher because the size of the budget is increased by the direct subsidy relative to gifts by taxpayers, unless this exercise begins by assuming a marginal tax rate of 100 percent. If the highest marginal income tax rate is 40 percent, as it is in the United States, then the 60 percent of the gift that is paid entirely from the taxpayer’s own resources reduces the deadweight loss from taxation, compared to a direct expenditure of government funds of the same amount in the form of subsidy.

### 6.3. *Crowding-in or crowding-out?*

During the early life of the NEA, there was considerable concern among the managements and governing boards of large and prestigious cultural organizations, especially museums, that direct Federal government support might reduce private philanthropic donations to their organizations. There is a considerable economic literature on the possible “crowding out” of private money and voluntary labor by incremental government spending for parallel purposes.<sup>32</sup> Some of this focuses specifically on cultural activities and organizations. Earlier work addressed the issue of the possible negative effect on donor motivations; almost all, like Abrams and Schmitz (1978), found that there was

<sup>31</sup> A commonly used rule of thumb is that the excess burden of the Federal income tax at the rates prevailing since 1986 is 1.6 times the tax collected.

<sup>32</sup> There is an even larger literature on the possible “crowding out” of private investment and consumption spending by government macroeconomic measures.

some crowding out but that it is not complete, implying that donors view government grants as imperfect substitutes for private giving.

Borgonovi and O'Hare (2004) approached the issue somewhat differently, employing both econometric and interview data to examine crowding (crowding-in as well as crowding-out) at the level both of individual institutions and of the cultural sector as a whole. Their interviewees (managers of cultural institutions and foundation donors) almost uniformly disputed the notion that there is any crowding in, although there was some in the early history of the two national Endowments. Their econometric model used data for the period 1955–1999. The empirical results suggest that Endowment grants do not have significant effects on grants received by individual organizations institutions or by the sector as a whole.

Some work has addressed the effect of grants on fund-raising activities by charities; the usual finding is that there have been significant reductions in spending for fund-raising caused by government grants. Andreoni and Payne (2003) analyzed panel data for 233 non-profit arts organizations and 534 social services organizations over a 10-year period, with fund-raising expenditures averaging more than 10 percent of their total revenue from contributions, an average that is stable over the period. They found that an increase in government grants of \$1000 on average reduces fund-raising expenditure by \$264, a decline of 52 percent. Conceivably this could be efficient if the deadweight loss of fund-raising is greater than the deadweight loss of taxation. If not, then the efficient outcome might be a requirement that for the grant to be continued or increased there must be evidence of increased fund-raising spending. American foundations frequently condition grants on just this.

## 7. Heritage policy

### 7.1. *Public sector role in heritage protection*

Heritage policy in the United States, in sharp contrast to public policy with regard to the arts, is supported by considerable direct public financing, much of it by the Federal government, as well as by generous tax incentives. In part this may be because the supply of physical heritage sites and artifacts is relatively small, so that protecting a large fraction of them is not very costly. Many are not located in cities, so the opportunity costs in the form of possible alternative high-value uses are also low for the most part.

The physical heritage outside museums in the US consists largely of historic buildings and sites. Unlike most countries in Western Europe, there is not a vast supply of paintings and sculpture in public places or in small churches, of a quality and significance that makes them part of the commonly accepted national heritage, however costly it may be to protect them. Indeed, in the US much of the sculpture in public places is arguably not highly regarded either by ordinary people or by experts, and often can be suitably restored at relatively modest cost when the need arises. Many historic structures are built of wood rather than masonry, and thus suffer the disadvantage of deteriorating



more quickly. But the offsetting advantage is that the deterioration is usually easier to repair, skilled carpenters being much more numerous than skilled stone-cutters in the US where more than 75 percent of the housing stock consists of wooden buildings. A large majority of protected historic sites managed by government agencies are predominantly open space with a few wooden structures.

Therefore the issues surrounding heritage policy in the US do not include the concern that there may be too much officially designated heritage and too rapid a growth in the number of such designations, as Françoise Benhamou (1996, 1998) has suggested is the case in France and the United Kingdom. No doubt the same is true in much of the rest of Europe; even with conservative hypothetical illustrations of the value of heritage structures and artifacts in these countries, the annual costs of offsetting physical capital consumption could be a large fraction of GDP, for example, in Italy [Netzer (1998)].

As noted above, the public sector's role in heritage protection in the US is substantially larger than in the arts per se. This role is administered in a number of ways. For example, the National Park Service (a bureau of the Department of the Interior) includes in its natural resource protection mission the protection of historic places, sites, trails and monuments, most of them battlefields (or related otherwise to wars fought in North America), important places in the exploration and settlement of the American West in the 19th century, and sites significant in the 17th and 18th century that illustrate aspects of the workings of the economy. The Park Service owns and manages more than 250 such places, at an average annual cost of at least \$1 million, devoting more than 10 percent of its annual budget of \$2.0 billion to these efforts.<sup>33</sup> There also are state and local government efforts to preserve historic sites but it has been the practice to persuade the Federal government to assume ownership and responsibility for historic preservation involving sites with limited private uses wherever possible.

## 7.2. Heritage tax incentives and regulation

As in the arts, there is substantial indirect assistance to heritage protection in the form of tax incentives. There is a National Register of Historic Places, and owners of buildings located in places that are in the Register are eligible for tax credits for preservation work. The Register began in 1968 and contains more than 65,000 listings of places, each of which may have many buildings whose owners can take advantage of the tax savings. The main Federal tax incentive program, which began in 1978, had provided tax credits that reached \$19.2 billion by 1997. It is estimated that there are three million privately-owned structures included in Federal, state and local government registers [Listokin, Listokin and Lahr (1998, pp. 433–439)].

Another aspect of protection of heritage is the regulation of what private property owners may do to their properties, if those properties have been designated historical “landmarks” or “monuments” or are located in historical districts so designated by local

<sup>33</sup> [www.data2.ipc.nps.gov/parksearch/atoz.cfm](http://www.data2.ipc.nps.gov/parksearch/atoz.cfm).

governments (the generic term is “local historic preservation district”). In 1955, there were 20 such districts; the earliest were the Georgetown section of Washington, the Vieux Carré in New Orleans, “Historic Charleston” (South Carolina) and Beacon Hill in Boston. Twenty-five years later, there were more than 800, and more than 2000 by 1996 [Listokin, Listokin and Lahr (1998, pp. 438–439)]. In these districts, changes to the exterior of buildings require permission by an official agency, in order to preserve the historic exterior appearance of the buildings and to prevent demolition and replacement by new structures whose appearance is inconsistent with the architecture that characterizes the district.<sup>34</sup> Owners of buildings in the districts can claim “economic hardship” but this is often hard to demonstrate; in New York small buildings like churches located in high-demand areas often can sell the development rights (to build much higher buildings on a nearby site outside the historic district) at high prices.<sup>35</sup>

In addition, historic district designation is often thought to be accompanied by large increases in property values, much larger than in the remaining areas of the city. However, the literature on this shows mixed results, which is to be expected in the light of the substantial variation in real estate markets across the country. Early studies compared price level changes in designated historical districts with price level changes in other parts of the same cities, without rigorous controls for other differences between historic and other districts [Scribner (1976); Gale (1991)]. Subsequent studies used the hedonic pricing method to isolate the implicit price of the attribute of being in a historic district. Asabere and Huffman (1994) found substantial price increases in Philadelphia associated with both historic facade easements (which essentially give away the owner’s discretion in making facade changes) and being in a historic district. Coulsen and Leichenko (2001) found net positive effects on property values of houses in Abilene, Texas in a Census tract with historic houses, relative to tracts with no historic designation.

The most persuasive study is that by Noonan (2004), who examined 73,000 sales of residential property in Chicago in 1990–1999.<sup>36</sup> The initial finding was that landmark designation had a strong effect on sales prices, adding between 6 and 15 percent. However, Noonan then used the repeat-sales approach, an approach familiar to cultural economists who have used it in analyzing price trends for paintings and other works of art over long periods. The approach controls for unobserved property characteristics, as it does for the uniqueness of a work of art. Using this method, Noonan found that landmark designation in Chicago did not have positive price effects, and may even have had negative effects on price appreciation over time.<sup>37</sup>

<sup>34</sup> Although there are districts characterized by a cluster of cultural activities, historic districts are mostly a celebration of architecture, rather than the “cultural districts” discussed in Chapter 31 by Santagata in this volume.

<sup>35</sup> If “economic hardship” can be demonstrated, the action can be construed as a “taking” of private property for public use. US Constitution requires that owners be compensated for their losses when this occurs.

<sup>36</sup> Cited with the author’s permission.

<sup>37</sup> Public property tax records in New York suggest that residential property prices in historic districts have risen far more rapidly than in the rest of the city, since the districts were designated. This is no doubt partly

Conceivably, if there are substantial increases in market values occasioned by historic designation, local property taxes, which are based on market value in American law, will rise. But such increases in taxable property values are not the general case, since in most cities local government valuers are hesitant to do this. Moreover, the Federal income tax law has been interpreted by courts to permit very large deductions from taxable income and thus income tax payments when property owners grant “easements” to a not-for-profit organization under which no change in the external appearance of the building can be made without the consent of the holder of the easement – which consent is rare. The US Tax Court has found in several cases that giving away that particular property right reduces the market value of the property by 11 percent, which has become the standard tax rule. However, when the historic designation of a district is associated with very large increases in the market value of property in the district, strengthening the rules against changes in appearance will increase, not decrease, market value.

### 7.3. Heritage valuation

A study by Frey and Oberholzer-Gee (1998) is one of the relatively few efforts to address the question of the evaluation of the benefits of heritage protection. Where a more-than-nominal admission price to the building or site is charged, the immediate benefits to those who pay the charge has been revealed. But there is much protected heritage for which charging is entirely infeasible or feasible for limited aspects of the building or site only. Most protected heritage also provides option and bequest value. Possible approaches to measurement of these externalities include contingent valuation, hedonic pricing in the markets for real property, cost–benefit analysis and the travel-cost method. However, the approach preferred by Frey and Oberholzer-Gee is direct democracy, i.e. voting in referenda on specific projects and their budgets. That is common in Switzerland, and occasionally happens in the US for arts projects but not for heritage purposes.

Navrud and Ready (2002) examine the application of various techniques used for environmental policy evaluation to the evaluation of cultural heritage, including the travel-cost approach, which was developed in the 1950s by Marion Clawson and Jack Knetsch as a means of valuing the protection of places of natural beauty in the US, notably national parks in the West and South [Clawson (1959); republished in 1992]. One pragmatic advantage of the approach is that there are good data on travel costs provided by Federal and state government transportation departments, including estimates of the value of time devoted to travel. The approach seems appropriate because much of the protected American heritage is in places that are visited by people who live at some distance from these sites. If these visits had not occurred, physical decay and transformation to other uses would have caused many heritage sites to have been lost

explained by the tax law provision that severely restricts taxable value increases over time on 1–3 family houses, the prevalent type of property in numerous New York City “landmark districts”. Thus, their market values are not reflected in property tax payments. (This is based on a survey by the author.)

during the many years before governments took an energetic role in their protection. It is perhaps surprising that there has been little use of the travel-cost approach in the valuation of heritage in remote sites. One such use is described in [Poor and Smith \(2004\)](#). The heritage site in question is a restored 17th century settlement that was the initial English settlement in Maryland, relatively remote from the larger cities of the region. It is on no main travel route and accessible only by automobile. The value estimates in the alternative forms of their equations vary considerably, but they do support some policy suggestions. Price elasticities were found to be quite low and income elasticities negative, suggesting caution in up-scale advertising.

## **8. Further research**

There has been very little written about how the peculiarly American aspects of cultural policy – notably the heavy reliance of the arts on a revenue stream consisting largely of earned income and private gifts stimulated by tax incentives rather than direct government financial support – affect the composition (rather than the level) of cultural output. For example, “block-buster” museum shows have become the dominant element in arranging exhibitions at major American art museums. Arguably, the block-buster substantially increases revenue from admission charges and generates more and larger private gifts to museums. If so, the result must be to increase the amount of time and space devoted to such exhibitions and to reduce the time and space devoted to exhibitions from the museum’s own collections, compared to the experience in countries with quite different policy regimes.

Similarly, the dependence of American dance companies on revenues from touring in Europe may be a consequence of the negligible level of direct government financial support available at home. Also, there may be programmatic effects of the heavy reliance on sub-national government entities for financial support. The answer to this latter question requires better information on just what local governments actually do in their subsidies to cultural output.

It has been argued that the for-profit sectors in the arts depend heavily on the creativity of the non-profit sector. This is obvious in theater and cinema but less so for other disciplines. The rents generated by this use of material and ideas usually are captured by individual creative people, not by the non-profit companies. Little systematic research on this topic has been done. Also, little is known about the consequences for cultural output (as well as for the size of potential audiences) of the lack of exposure to art and music of children in the great majority of the public elementary and secondary schools.

## **9. A concluding observation**

Like the other chapters in this volume, this one seeks to explain the cultural policy choices in the US as if they had been products of economic analysis, specifically in

terms of market failure. Zimmer and Toepler (1999) challenge that approach. They argue that neo-classical market failure arguments fail to take into account historical and institutional factors that exist in every country, and that lead to quite distinctive cultural policies: “while market failure provides a strong argument why government should intervene, it fails to explain international variations in the extent of public support” (p. 46). Or international variations in the institutional arrangements through which cultural policies are implemented. Their argument is a persuasive one.

## References

- Abrams, B.A., Schmitz, M.D. (1978). “The ‘crowding-out effect’ of Governmental transfers on private charitable contributions”. *Public Choice* 33 (1), 29–39.
- American Association of Museums (1999). *Museum Financial Information, 1999*. American Association of Museums, Washington, DC.
- American Association of Museums (2003). *Museum Financial Information, 2003*. American Association of Museums, Washington, DC.
- Americans for the Arts (2001). *Local Art Agencies Facts, Fiscal Year 2000*. Americans for the Arts, Washington, DC.
- Andeoni, J., Payne, A.A. (2003). “Do government grants to private charities crowd out giving or fundraising?” *American Economic Review* 93 (3), 792–812.
- Asabere, P.K., Huffman, F.E. (1994). “Historic designation and residential market values”. *The Appraisal Journal* 62 (4), 396–401.
- Auerbach, A. (1985). “The theory of excess burden and optimal taxation”. In: Auerbach, A., Feldstein, M. (Eds.), *Handbook of Public Economics*. North-Holland, New York.
- Baumol, W.J., Bowen, W.G. (1966). “Performing Arts: The Economic Dilemma”. Twentieth Century Fund, New York.
- Benhamou, F. (1996). “Is increased spending for the preservation of historic monuments inevitable? The French case”. *Journal of Cultural Economics* 20 (2), 115–131.
- Benhamou, F. (1998). “The evolution of heritage policies”. In: Peacock, A. (Ed.), *Does the Past Have a Future? The Political Economy of Heritage*. The Institute of Economic Affairs, London, pp. 75–95.
- Borgonovi, F., O’Hare, M. (2004). “The impact of the National Endowment for the Arts in the United States: Institutional and sectoral effects on private funding”. *Journal of Cultural Economics* 28 (1), 21–36.
- Bourscheidt, R. (2002). “Letter to the Editor”. *The New York Times*, June 6, 2002. Web address: <http://www.nytimes.com/2002/06/06/L06ARTS/html>.
- Clawson, M. (1959). “Methods of measuring the demand for and value of outdoor recreation”. In: Oates, W.E. (Ed.), *The Economics of the Environment*. Edward Elgar, Brookfield, VT.
- Cordes, J.J. (2004). “The partially subsidized muse: Estimating the value of public support received by non-profit arts organizations”. In: Schwartz, A.E. (Ed.), *City Taxes, City Spending: Essays in Honor of Dick Netzer*. Edward Elgar, Northampton, MA, pp. 198–240.
- Coulsen, N.E., Leichenko, R.M. (2001). “The internal and external impact of historical designation on property values”. *Journal of Real Estate Finance and Economics* 23 (1), 113–124.
- Davidson, B. (2002). *United States Urban Arts Federation – Fiscal Year 2001*. Americans for the Arts, Washington.
- Diamond, S. (2002). “Efficiency and benevolence: Philanthropic tax exemptions in 19th century America”. In: Brody, E. (Ed.), *Property-Tax Exemption for Charities*. Urban Institute Press, Washington, pp. 115–144.
- Epstein, R.A. (2003). “The regrettable necessity of contingent valuation”. *Journal of Cultural Economics* 27, 259–274.
- Feld, A.L., O’Hare, M., Schuster, J.M.D. (1983). *Patrons Despite Themselves: Taxpayers and Arts Policy*. New York University Press, New York.

- Felton, M.V. (1978). "The economics of the creative arts: The case of the composer". *Journal of Cultural Economics* 2 (1), 41–62.
- Felton, M.V. (1980). "Policy implications of a compose labor supply". In: Hendon, W.S., Shanahan, J.L., MacDonald, A.J. (Eds.), *Economic Policy for the Arts*. Abt Books, Cambridge, MA, pp. 186–197.
- Felton, M.V. (1992). "On the assumed inelasticity of demand for the performing arts". *Journal of Cultural Economics* 16, 1–12.
- Frey, B.S. (1999). "State support and creativity in the arts: Some new considerations". *Journal of Cultural Economics* 23, 71–85.
- Frey, B.S., Oberholzer-Gee, F. (1998). "Public choice, cost–benefit analysis, and the evaluation of cultural heritage". In: Peacock, A. (Ed.), *Does the Past Have a Future? The Political Economy of Heritage*. Institute of Economic Affairs, London, pp. 27–53.
- Frey, B.S., Pommerehne, W. (1989). *Muses & Markets: Explorations in the Economics of the Arts*. Basil Blackwell, Oxford.
- Frey, B.S., Pommerehne, W. (1995). "Public expenditures on the arts and direct democracy: The use of referenda in Switzerland". *Cultural Policy* 2 (1), 55–65.
- Gale, D.E. (1991). "The impacts of historic district designation". *Journal of the American Planning Association* 57 (3), 325–341.
- Gurney, A. (2003). *The Race to the White Continent: Voyages to the Antarctic*. W.W. Norton, New York.
- Heilbrun, J. (2001). "Empirical evidence of a decline in repertory diversity among American opera companies". *Journal of Cultural Economics* 25, 63–72.
- Heilbrun, J., Gray, C.M. (2001). *The Economics of Art and Culture: An American Perspective*, second ed. Cambridge University Press, New York.
- Jackson, K.T. (Ed.) (1995). *The Encyclopedia of New York City*. Yale University Press, New Haven, CT.
- Kendall, E. (1983). *Dancing: A Ford Foundation Report*. The Ford Foundation, New York.
- Listokin, D., Listokin, B., Lahr, M. (1998). "The contribution of historic preservation to housing and economic development". *Housing Policy Debate* 9 (3), 431–448.
- Marquis, A.G. (1995). *Art Lessons: Learning from the Rise and Fall of Public Arts Funding*. Basic Books, New York.
- McCarthy, K.F., Brooks, A., Lowell, J., Zakaras, L. (2001). "The Performing Arts in a New Era". RAND, California.
- Meyer, K.E. (1979). "The Art Museum: Power, Money, Ethics". William Morrow, New York.
- National Assembly of State Arts Agencies (2001a). *Legislative Appropriations Annual Survey, Fiscal Year 2001*. Washington, DC.
- National Assembly of State Arts Agencies (2001b). *State Arts Agency Public Funding Sourcebook, 1966–2001*. Washington, DC.
- National Assembly of State Arts Agencies (2002). *Legislative Appropriations Annual Survey, Fiscal Year 2002*. Washington, DC.
- National Assembly of State Arts Agencies (2003). *Legislative Appropriations Annual Survey, Fiscal Year 2003*. Washington, DC.
- National Endowment for the Arts (1988). *The Arts in America: A Report to the President and to the Congress*. NEA, Washington, DC.
- Navrud, S., Ready, R.C. (Eds.) (2002). *Valuing Cultural Heritage: Applying Environmental Valuation Techniques to Historic Buildings, Monuments and Artifacts*. Edward Elgar, Cheltenham.
- Netzer, D. (1986). "Dance in New York: Market and subsidy changes". *American Economic Review* 76 (2), 15–19.
- Netzer, D. (1987). "Changing economic fortunes of dance in the US". In: Grant, N.K., Hendon, W.S., Owen, V. (Eds.), *Economic Efficiency in the Performing Arts*. Association for Cultural Economics, Akron, OH.
- Netzer, D. (1992a). "Cultural policy in an era of budgetary stringency and fiscal decentralization: The US experience". In: Towse, R., Khakee, A. (Eds.), *Cultural Economics*. Springer-Verlag, Berlin, pp. 237–245.
- Netzer, D. (1992b). "Arts and culture". In: Clotfelter, C.T. (Ed.), *Who Benefits from the Nonprofit Sector?* University of Chicago Press, Chicago.

- Netzer, D. (1998). "International aspects of heritage policies". In: Peacock, A. (Ed.), *Does the Past Have a Future? The Political Economy of Heritage*. Institute of Economic Affairs, London, pp. 135–154.
- Netzer, D., Parker, E. (1993). "Dancemakers". Research Division Report #28. National Endowment for the Arts, Washington.
- Noonan, D. (2004). "Neighborhood effects and residential property prices: The case of historical landmarks in Chicago". Paper presented at the 2004 research conference of the Association for Public Policy and Management (APPAM). Cited with the author's permission.
- O'Hare, M. (1980). "The Malthusian nightmare of the composer and his audience". In: Hendon, W.S., Shanahan, J.L., MacDonald, A.J. (Eds.), *Economic Policy for the Arts*. Abt Books, Cambridge, MA, pp. 114–120.
- Philbrick, N. (2003). *Sea of Glory: America's Voyage of Discovery: The US Exploring Expedition, 1838–1842*. Viking, New York.
- Poor, P.J., Smith, J.M. (2004). "Travel cost analysis of a cultural heritage site: The case of historic St. Mary's City of Maryland". *Journal of Cultural Economics* 28 (3), 217–229.
- Port Authority of New York and New Jersey (1993). *The Arts as an Industry: Their Economic Importance to the New York–New Jersey Region*. Port Authority of New York and New Jersey, New York.
- Port Authority of New York and New Jersey (1994). *Destination New York–New Jersey: Tourism and Travel to the Metropolitan Region*. Port Authority of New York and New Jersey, New York.
- Rushton, M. (2000). "Public funding of controversial art". *Journal of Cultural Economics* 24 (4), 267–282.
- Seaman, B.A. (1987). "Arts impact studies: A fashionable excess". In: Radich, A.J. (Ed.), *Economic Impact of the Arts: A Sourcebook*. National Conference of State Legislatures, Denver, pp. 43–75.
- Seaman, B.A. (2002). *National Investment in the Arts*. Center for Arts and Culture, Washington, DC.
- Schuster, J.M.D. (1989). "Determinants and correlates of arts support by states". In: Shaw, D.V., et al. (Eds.), *Cultural Economics, '88: An American Perspective*. Association for Cultural Economics, Akron, OH, pp. 211–224.
- Scribner, D. Jr. (1976). "Historic districts as an economic asset to cities". *The Real Estate Appraiser*, May–June, 7–12.
- Smithsonian Institution Libraries (2002). *From Smithson to Smithsonian: The Birth of an Institution* (an on-line exhibition).
- Smolensky, E. (1986). "Municipal financing of the US fine arts museum: A historical rationale". *Journal of Economic History* 46 (3), 757–768.
- Surrey, S.S. (1973). *Pathways to Tax Reform: The Concept of Tax Expenditures*. Harvard University Press, Cambridge, MA.
- Throsby, D. (1983). "Perceptions of quality in demand for the theatre". In: Shanahan, J.L., Hendon, W.S., Hilhörst, I.H., van Straalan, J. (Eds.), *Markets for the Arts*. Association for Cultural Economics, Akron, OH, pp. 6–21.
- Throsby, D. (1992). "Artist as workers". In: Towse, R., Khakee, A. (Eds.), *Cultural Economics*. Springer-Verlag, Berlin, pp. 201–208.
- Throsby, D. (1994). "A work-preference model of artist behaviour". In: Peacock, A., Rizzo, I. (Eds.), *Cultural Economics and Cultural Policies*. Kluwer Academic, Dordrecht, pp. 69–80.
- Throsby, D. (2003). "Determining the value of cultural goods: How much (or how little) does contingent valuation tell us?" *Journal of Cultural Economics* 27, 275–285.
- US Census Bureau (2001). *Statistical Abstract of the United States, 2000*. CD-COMP-ABSTR00. Also, *Statistical Abstract of the United States for 2001 and 2002*.
- Zimmer, A., Toepler, S. (1999). "The subsidized muse: Government and the arts in Western Europe and in the United States". *Journal of Cultural Economics* 23 (1–2), 33–49.

## TAX INCENTIVES IN CULTURAL POLICY\*

J. MARK SCHUSTER

*Massachusetts Institute of Technology, Cambridge, MA, USA*

### Contents

Abstract	1254
Keywords	1255
1. Introduction: Tax policy as cultural policy	1256
2. The literature on tax policy in cultural policy	1259
2.1. The early literature	1259
2.2. Tax expenditure analysis	1260
2.3. Other issues in tax policy	1263
2.4. Putting tax policy back on the cultural policy agenda	1265
3. The purpose of tax policy vis-à-vis the arts and culture	1266
3.1. Defining the tax base v. providing incentives	1267
3.2. Individual v. collective decision making	1269
4. The economic effects of tax exemption	1271
4.1. Introduction	1271
4.2. The price elasticity of individual giving	1272
4.3. Tax incentives as matching grants	1275
4.4. Price elasticities by charitable sector	1276
4.5. The income elasticity of giving	1276
4.6. The economics of corporate contributions	1277
4.7. Studies of charitable giving in other countries	1279
4.8. Who pays and who benefits? The question of incidence	1281
4.9. Crowding in or crowding out?	1283
4.10. Conclusion	1285
5. The international move toward tax-based incentives	1286
5.1. Chile	1287

\* In writing this chapter I have relied extensively on several excellent overview articles by Richard Steinberg [Steinberg (1990); Steinberg and Bilodeau (1999); Steinberg (2003)]; the chapter would not have been possible without the guidance of these papers. I also wish to recognize the contributions made by my two Research Assistants, Jason Schupbach and Susanne Seitinger, who cast a wide international research net to identify the many resources referenced here as well as the many others that did not make it into the final version of this chapter.

*Handbook of the Economics of Art and Culture, Volume 1*

*Edited by Victor A. Ginsburgh and David Throsby*

*Copyright © 2006 Elsevier B.V. All rights reserved*

*DOI: 10.1016/S1574-0676(06)01036-2*



5.2. Australia	1287
5.3. Singapore	1289
5.4. France	1290
5.5. United States	1291
5.6. Romania	1292
5.7. Conclusion	1293
6. Summary and concluding remarks	1293
References	1294

## Abstract

Cultural policy discussions are increasingly concerned with the creation and restructuring of tax incentives; thus, cultural policy and tax policy are becoming more and more intertwined. With the widely held perception that there has been a general decrease in the availability of direct public resources for culture, a search has begun for other sources of support and for ways to provide incentives for those other sources. Moreover, with the growth in the use of forms of decentralization, désétatisation, and devolution in cultural policy, increased attention has been paid to tax policy as a way of spreading decision making over public resources more broadly. Thus, there has been a rise in the use of tax policy to provide incentives for what is considered to be desirable behavior vis-à-vis the arts and culture, particularly though not exclusively with respect to its nonprofit component. It has been clearly documented that the indirect aid embedded in various taxes forgone by the various levels of American government are a much more important source of financial support than are the government programs of direct support and that foremost among these is the deductibility of charitable contributions. As a result, many countries have begun to pay more attention to the “American model” of cultural support with its high level of reliance on private donors and its attendant tax incentive structure.

Adopting a “tax expenditure” perspective, this chapter begins with a summary of the existing literatures on tax policy in cultural policy. It then proceeds to a consideration of the accumulated evidence with respect to the effects of tax exemption: the price elasticity of giving; the income elasticity of giving; the differences between individual and collective decision making with respect to distributing public resources to the arts and culture; the kinship between tax incentives and matching grants; the economics of tax incentives for corporate contributions; the incidence of tax incentives in comparison to the incidence of direct support; and the extent to which tax incentives promote crowding out or crowding in behavior with respect to other revenue sources. While most research has been focused on American tax incentives for charitable giving, some studies do exist for other countries, and their results are summarized where available.

For the moment, the econometric results depend substantially on the model specification and the type of dataset used. For those who would argue in favor of tax-based incentives, at least with respect to charitable donations, there is considerable evidence

in support of the view that such incentives would have the desired economic effects. But those who are dubious about the net effects can also find evidence to support their arguments. Despite the ambiguity of the econometric evidence in the United States, the use of tax-based incentives is proliferating. A set of contemporary international examples suggests (i) that indirect aid is the terrain in which many of the most interesting innovations and variations in the funding of the arts and culture are taking place, and (ii) that this proliferation is more likely due to the influence of politics and advocacy than to the influence of reasoned analysis.

### **Keywords**

cultural policy, tax policy, tax incentives, charitable donations, tax deductions, tax credits, nonprofit sector, matching grants, price elasticity of giving, income elasticity of giving, incidence analysis, crowding theory, tax expenditures, models of decision making

*JEL classification:* D12, H20, H22, H30, Z10

## 1. Introduction: Tax policy as cultural policy

Most analytic attention to cultural policy has been paid to direct policy, and rightfully so, since government's cultural agencies and the policies, programs, and projects that they oversee are a visible and important component of cultural policy. But over the past twenty-five years or so, considerable research in the cultural policy arena has demonstrated that indirect methods of government involvement also figure importantly in the cultural policy mix. Unemployment or other social welfare rules and how they interact with employment practices in the cultural sector are cases in point, but first among these is the contribution that tax policy makes to cultural policy.

Tax policy is of interest for four primary reasons:

- *The use of tax policy to provide incentives for what is considered to be desirable behavior in the arts and culture, particularly though not exclusively within its non-profit component.* Nonprofit arts and cultural organizations might benefit from a wide variety of favorable tax provisions: they might be exempted from property tax,<sup>1</sup> from sales taxes on purchases they make, or from paying income tax on any surplus revenue they realize in a taxable year (as long as that surplus is reinvested in the mission of the organization); a donor might be able to benefit from tax provisions that reduce the donor's cost of private giving to the cultural institution. Individual artists might also benefit from various tax provisions: they might be exempted from state income taxes or state sales taxes under certain circumstances; they might even benefit from a national income tax exemption, as is the case for creative artists in Ireland.<sup>2</sup> Corporations in the cultural industries might benefit from special tax breaks of their own: reduced VAT rates on the products they produce, tax credits to encourage certain types of investment, or the like.
- *The use of special dedicated taxes as a source of revenue for the arts and culture.* Recent debates in cultural policy have seen an upswing in the call for dedicated sources of public revenue for the arts and culture. In large part this has been fed by an increase in the number of state lotteries and a relatively successful attempt to make this new source of public revenue palatable to voters – to garner political support, lottery revenues are dedicated to favored “good causes”. But there are other, nonlottery examples as well. The creation of the Colorado Scientific and Cultural Facilities District, a six-county special taxation district within which an additional one-tenth-of-one-percent sales tax is collected to support scientific and cultural facilities, is a case in point [Martell (2004)].
- *A desire to ensure fairness of tax treatment within the arts and culture as well as across sectors.* Here the primary concern is the extent to which tax policy interacts in an unusual way with the sector under consideration. To take but one example, because of the sporadic and unpredictable nature of the income of authors,

<sup>1</sup> An excellent discussion of property tax exemptions is contained in Brody (2002).

<sup>2</sup> A comparative survey of the tax treatment of individual artists in various European countries is contained in Munnely (1986).

playwrights, or composers, it may be good public policy to introduce income averaging, allowing spikes in income to be spread over longer periods of time and thereby ameliorating what might otherwise be an unacceptable tax burden.

- *A turn away from direct forms of support to indirect forms of support as a political strategy.* As public budgets have become more restricted and as direct support for the arts and culture has experienced moments of controversy, less visible forms of government support have become more attractive to the sector as well as to politicians.

Moreover, tax policy in particular and indirect aid more generally are the terrains in which many of the most interesting innovations in the public funding of arts and culture are taking place.<sup>3</sup> Here more than in direct aid is where experimentation and variety can be found. At the biennial meetings of the Association for Cultural Economics and of the International Conference on Cultural Policy Research, as well as in the pages of their respective journals, the *Journal of Cultural Economics* and the *International Journal of Cultural Policy*, it is clear that all over the world policy discussion is increasingly concerned with the creation and restructuring of tax incentives (though ownership and operation of state cultural institutions and direct support through grants and operating subsidies continue to be favored modes of support in many countries). As a result, it has become quite difficult for individuals and organizations in the cultural sector to keep track of all of the tax incentives for which they might be eligible and of all of the tax provisions that might impact them.<sup>4</sup>

Even with this brief introduction, it is clear that the topic of “tax policy as cultural policy” is a vast one with many corners worthy of sustained exploration and analysis. Indeed, *any* tax regime might be the subject of a simple yet revealing analysis: to what extent does this current form of taxation or this tax provision advantage or disadvantage the cultural sector? And to what extent could this tax regime be modified to provide desirable benefits to the sector? At what cost? Much of the discussion concerning the intersection between tax policy and cultural policy is concerned with precisely the question of the extent to which tax policy affects cultural policy. Put another way, much of the discussion of the contribution that tax policy can make to cultural policy is focused on the first point above: To what extent can tax policy be used to provide effective incentives for desirable behavior vis-à-vis the arts and culture?

Taxes play a role in many corners of cultural policy:<sup>5</sup> tax provisions are used in an attempt to attract artists to become part of urban redevelopment efforts [Schuster (1999)] or more broadly at a national level to support a particular group of artists, as

<sup>3</sup> See, for example, Schuster (1999).

<sup>4</sup> This has given rise to taxation catalogues. Simmonds (2001) is an example of such a guide for the arts in the UK; Staines (2004) provides one for artists and promoters operating in Europe; and Simon (1987) and Facchina, Showell and Stone (1993) provide useful cross-sector catalogues for the United States.

<sup>5</sup> For a broad summary of tax incentives for nonprofit organizations in the United States, see Simon (1987), but note that this extensive summary does not include any discussion of tax incentives for profit-making organizations in the cultural field.

occurs in Ireland [Coffey (1998)]; they are used to lower the relative costs of certain cultural goods and services as compared to other goods and services – lower VAT rates on the purchase of books are a case in point [Rouet (1999)]; they are used in addition to regulation and outright purchase to provide an incentive to keep artworks considered part of the national patrimony within national borders [McAndrew and O’Hagan (2000); O’Hagan and McAndrew (2001, pp. 49–50)] or to make donations to national museums in lieu of taxes more attractive than sale [Council for Museums, Archives and Libraries (2002, p. 10)]; and special dedicated taxes are implemented to serve as a supplementary source of support for the arts [Schuster (1996, 2001); Martell (2004)]. These examples are only the tip of a very large iceberg.

But one area of tax policy has generated more interest in cultural policy – and in governmental policy vis-à-vis the nonprofit sector more generally – than any other: the tax treatment of donations to cultural and other nonprofit organizations. This topic has reached its preeminence because of the fact that, in an era of what appear to be increasingly limited public resources, many countries have begun to pay more attention to the “American model” of cultural support with its high level of reliance on private donors, particularly individuals but also corporations and private nonprofit foundations.<sup>6</sup> It has been clearly documented that the indirect aid embedded in various taxes forgone by the various levels of American government are a much more important source of financial support than are the government programs of direct support [Feld, O’Hare and Schuster (1983)], and that preeminent among these is the deductibility of such contributions.

Thus, in order to maintain some focus in the current chapter, I will concentrate my attention on tax provisions with respect to charitable donations and their incentive effects. I will make reference to examples drawn from other components of tax policy where appropriate, as well as to the broader issues that tax policy raises when it intersects with cultural policy, but I will not be able to provide a comprehensive overview of this very large field. Restricting my focus in this way has one further advantage: this is where the data are probably the best – though only in certain national contexts – and where, to date, the analysis has been the most complete.

But before I turn almost exclusively in this direction, let me turn first to a summary of the current state of the art with respect to the cultural-policy-based literature on tax policy.

<sup>6</sup> A general misimpression about the sources of revenue for American cultural organizations deserves to be clarified here. As a rough rule of thumb, American cultural organizations receive approximately 60 percent of their revenue from “earned” sources of revenue – ticket sales, subscription sales, memberships, and sales through shops, cafes, publishing operations, and parking facilities. The remaining 40 percent is in the form of “unearned” revenue. Typically three quarters of this unearned income comes from the private charitable contributions of individuals, with the remaining one-quarter coming from corporations, foundations, and the government. Thus, while there are always exceptions to such a generalization, the most salient characteristic of this revenue profile is the importance of private contributions from individual supporters of the organization. They provide an average of 30 percent of the total revenue of the organization. But, as we shall see, this 30 percent is made up of two portions: truly private after-tax donations and the taxes that have been forgone through public policy.

## 2. The literature on tax policy in cultural policy

### 2.1. The early literature

The first recognition of tax policy as an important complement to direct aid in cultural policy (at least in the United States) came at about the same time as the primary governmental cultural agencies – the National Endowment for the Arts and the National Endowment for the Humanities – were being created. Thus, from the very beginning a mix between direct and indirect aid was under discussion. August Heckscher's 1962 report in which he recommended that President Kennedy create an "Advisory Arts Council" made numerous recommendations for changes in tax law aimed at increasing government assistance to the arts [cited in [Netzer \(1978\)](#)]. Several reports that were released about the time that the National Endowment for the Arts was created also argued for an increased use of indirect subsidies, presumably to counteract the fears of overcentralization that accompanied the creation of the new agencies.<sup>7</sup> In each of these reports tax policy played only a minor role. To the extent that these documents paid attention to this area, they were all exhortatory; the fundamental lack of reliable data made it virtually impossible for them to be analytic in any meaningful way.

In 1973 the Commission on Private Philanthropy and Public Needs (colloquially known as "The Filer Commission" after its chairman), a privately funded citizens' panel, was formed to study the role of philanthropic giving and voluntary public-oriented activity in the United States. This effort which resulted in the publication of five volumes of working papers and reports [[Commission on Private Philanthropy and Public Needs \(1977\)](#)] is still the most ambitious attempt to understand the dynamics of the American nonprofit sector and its relationship to the state. Only one twenty-two page paper in its more than 3000 pages dealt with the nonprofit arts [[Hightower \(1977\)](#)]. While carefully documenting the importance of various sources of support to arts organizations, this paper was remarkably silent on the behind-the-scenes role played by tax law. This is especially striking since one entire volume of the commission's papers dealt directly with a myriad of questions concerning tax provisions. In retrospect, it seems clear that the arts sector was not yet focused on the actual contribution that tax policy was making to its financial health, even as others outside the sector were beginning to sense an accumulation of interesting research questions and policy issues surrounding both existing and proposed tax laws.

In the mid to late 1970s some sustained attention began to be paid to tax policy questions in Europe. Ignace Claeys Bouuaert was commissioned by the Commission of the European Communities to compile three comparative reports on taxation issues in cultural policy among the European Economic Community (EEC). These reports dealt with the tax treatment of cultural organizations and contributions to those organizations

<sup>7</sup> See [Toffler \(1965, pp. 182–208\)](#); [Rockefeller Panel Report \(1965, Chapters 4 and 5, pp. 120–121, 138–143\)](#); and [Baumol and Bowen \(1966, Chapter 13 and pp. 348–356\)](#).

[Bouuaert (1975)], the tax treatment of cultural workers in the EEC [Bouuaert (1977)], and the tax treatment of historic properties [Bouuaert (1979)]. Each of these reports is a cross-national catalogue describing the national legislation in force in the EEC countries at the time. In each case Bouuaert struggled with the fact that a satisfactory set of categories to describe various taxation regimes had not yet been developed within the arts and culture, and the result was an ambitious though incomplete survey of the state of the art at the time.

## 2.2. Tax expenditure analysis

The publication of Stanley Surrey's (1973) book *Pathways to Tax Reform* marked a clear methodological and analytic divide in the study of tax incentives. Surrey argued that indirect government action, particularly the indirect action embodied in tax incentives, could and should be compared to direct aid; forgone taxes should be analyzed as a government "expenditure" in much the same way that direct outlays would be analyzed, and the two approaches should be compared to understand their respective attributes and their respective advantages and disadvantages. He coined the phrase "tax expenditure analysis" to describe this approach. A decade later, one of the most comprehensive analytic studies of the relationship between tax policy and cultural policy was commissioned by the Twentieth Century Fund as part of a series of books on support for the arts and culture in the United States.<sup>8</sup> *Patrons Despite Themselves: Taxpayers and Arts Policy* [Feld, O'Hare and Schuster (1983)] was the first study to attempt a tax expenditure analysis for a single sector of the economy rather than across a taxation instrument, and its importance goes well beyond the arts and culture.<sup>9</sup> Perhaps still the best known and most widely cited treatment of tax questions in the arts and culture, *Patrons Despite Themselves* remains a standard reference work despite the fact that the American tax context within which it was written has changed significantly, particularly with respect to marginal tax rates.

Feld, O'Hare and Schuster demonstrated that in the United States the indirect aid to the arts and culture that was embodied in various tax exemptions and reductions dwarfed direct aid to the arts and culture. For 1973 – their reference year<sup>10</sup> – they estimated that

<sup>8</sup> This set of studies makes for interesting reading, particularly if one places each volume within its historical context. Feld, O'Hare and Schuster (1983), mentioned in the text above, provided the *indirect* aid complement to Netzer's (1978) Twentieth Century Fund study of *direct* government aid to the arts in the United States, which focused on the National Endowment for the Arts and the state arts councils. In addition, Meyer (1979) was asked to consider policy in relation to art museums; and Banfield (1984) took up the question of public policy vis-à-vis the visual arts.

<sup>9</sup> As far as its authors are aware, it remains the only such comprehensive study for a single sector.

<sup>10</sup> The reference year was 1973 because it was the year for which the Commission on Private Philanthropy and Public Needs conducted the National Survey of Philanthropy, systematically collecting data for charitable contributions that allowed for the separate identification of the sectors to which the surveyed donors had made their contributions. In the words of Christopher Jencks (1987, p. 324), this survey "was the largest and most

indirect aid to the arts and culture in the form of forgone taxes was nearly two-and-a-half times the amount of direct governmental aid.<sup>11</sup> But, more importantly, they raised a number of policy questions inherent in supporting the arts and culture through tax exemptions rather than directly. First, they demonstrated that there was an important interaction between donor tastes, their income levels and therefore their marginal tax rates, and their relative propensity to give to the arts and culture as opposed to other charitable sectors. This, they argued, has led to a distortion of the public policy elements of a tax-exemption-based support policy. Moreover, they pointed out the degree of influence that had been accorded individual *private* donors in the allocation of what could be construed as *public* resources.<sup>12</sup> They cited many instances in which the general taxpayer was being asked to pay, in part, for the rather idiosyncratic results of the decisions made by individual donors. The important analytic point here is, of course, not solely that the decisions of an individual donor might be idiosyncratic. The decisions made in a direct aid system might also be idiosyncratic in the sense that they too might not be in line with accepted public policy. The question is whether one system of providing public assistance achieves systematically better results than does the other when measured with respect to articulated policy.<sup>13</sup> Still, public sector idiosyncrasies in decision making arise because of bureaucratic or political failures that presumably can be corrected through vigilance and institutional design; the idiosyncrasy of private decisions is inherent and cannot be as easily corrected.<sup>14</sup>

With respect to the property tax, Feld, O'Hare and Schuster (1983) argued that the net results of existing tax incentives was to fuel capital intensity among nonprofit organizations, making buildings and land inexpensive relative to other uses of their resources.<sup>15</sup> They demonstrated how changes in tax law could have perverse effects on arts organizations – museums, for example, would find themselves opposing tax reform behind the scenes because it was in their interest to keep marginal tax rates as high as possible, thereby keeping the marginal price of charitable donations lower and more attractive to donors. They also outlined how, at the heart of many tax exemptions, private interests were continually in tension with public policy.

detailed survey of giving ever conducted in America". But note that Jencks has criticized this source of data for possibly overestimating charitable giving among high-income individuals who itemize their deductions and underestimating charitable giving among low-income itemizers.

<sup>11</sup> As we will see below, more recent estimates have arrived at even more dramatic ratios for the United States.

<sup>12</sup> Elsewhere in this volume Dick Netzer argues that the degree to which tax incentives are taken up by donors in the United States may provide the best available estimate of taxpayers willingness to pay for the arts and culture out of public resources. In this regard, he argues, this estimate may even be superior to contingent valuation studies.

<sup>13</sup> See King and Blaug (1976) for an interesting example of evaluating direct aid in this manner.

<sup>14</sup> I am grateful to Victor Ginsburgh and David Throsby who made these points in response to a draft of this chapter.

<sup>15</sup> This theme is further explored by Vladeck (1976).



Table 1  
Revenue sources for American art museums, 1988

	Revenue (\$ millions)	Percentage	Adjusted percentage*	Notes
Operating revenue (earned income)	122.4	14.0%	14.0%	
Private support (contributed income)	235.0	27.0%	15.0% (12.0%)	Private portion Tax expenditure
Value of art donated	77.3	8.9%	2.3% (6.6%)	Private portion Tax expenditure
Total direct federal support	95.7	11.0%	11.0%	
Total direct state and local support	168.7	19.3%	19.3%	
Total indirect government support			18.6%	
Endowment income	173.0	19.8%	19.8%	
Total	872.1	100.0%	100.0%	

Source: Fullerton (1991).

Notes: Calculations are based on data from 155 arts museums surveyed in 1989 by the Association of Art Museum Directors.

\*Adjusted percentage reports percentages after adjustments have been made for the appropriate accounting of the tax expenditures embedded in charitable contributions.

There have been other more recent attempts to estimate the relative role of forgone taxes in support of American cultural institutions. Some have looked at a specific set of cultural institutions, basing their estimates on the limited data available from those institutions. For example, Fullerton (1991, pp. 198–199), focusing on art museums and using 1988 data, estimated that 18.6 percent of the operating income of American art museums was comprised of the value of forgone taxes on contributions of cash and the value of capital gains taxes forgone on gifts of appreciated art work (Table 1). Thus, by his estimate indirect federal aid for museums was one and a half times direct federal aid (11.0 percent of total revenue on average). Thus viewed, public support for these museums actually comprised an average of 48.9 percent (18.6 + 11.0 + 19.3) of total revenues. Put another way, over a third of the benefit provided by the public sector was provided through the forgone taxes embodied in tax expenditures:  $(12.0 + 6.6)/48.9 = 38$  percent.

But even this analysis underestimated the overall benefit of federal tax exemptions because it included neither the nontaxation of net operating revenues nor tax exemptions related to restricted donations to endowment, nor did the analysis make any attempt to estimate state or local tax expenditures – for example, neither the value of property tax exemptions nor exemptions from sales taxes on museum purchases nor state tax incentives for charitable contributions were considered. And note further that the 155

museums considered here included five large government art museums, which are the exception rather than the rule in the United States. If they were removed, the implicit subsidy for private museums would be seen to be much larger than direct government spending.

Other studies have tried to look at the arts in the aggregate, deriving their estimates in a variety of ways using new partial datasets that have become available on aggregate charitable giving by sector. Most recently, Brooks has written several short papers that explore some of these themes with more recent data. He has estimated that in 1999 the ratio of indirect to direct funding at the federal level (calculated by estimating the forgone federal taxes in charitable giving to arts and cultural organizations and comparing that figure to the federal appropriation for the National Endowment for the Arts) was \$14 indirect : \$1 direct [Brooks (2003a)].<sup>16</sup>

While this estimated ratio can be criticized on a number of grounds – the numerator may be overestimated because it relies on data provided by the American Association of Fund Raising Counsel which has every reason to be generous in its figures, and the denominator is undoubtedly low because there is far more direct federal aid to the arts than that aid included in the budget of the National Endowment for the Arts – it does suggest that the relative importance of indirect aid in the form of forgone taxes has increased dramatically. Note also that, as was the case with the Fullerton study cited above, Brooks makes no attempt to include estimates of the influence of property tax (or other state and local) exemptions. Contrast this to Feld, O'Hare and Schuster (1983) who did include such estimates, though they are now quite outdated.

### 2.3. Other issues in tax policy

Even beyond strict tax expenditure analysis, Schuster has been the most consistent contributor to the (American) literature on the role of tax policy in cultural policy. In several papers he has taken a look at tax incentives through a comparative perspective: he compares the American charitable deduction to the traditional British deed of covenant system of providing a tax incentive for charitable contributions [Schuster (1985a, 1989a)]. These papers provide a defense of the merits of the traditional British system, even as it was being gradually moved toward the American system of incentives and its own policy merits were being watered down. Building on the earlier work of Bouuaert (1975, 1977, 1979), he demonstrates [Schuster (1986)] that, contrary to popular wisdom, American style incentives for charitable contributions had already been implemented in a number of Western European countries and that other countries

<sup>16</sup> In an unpublished analysis, Brooks estimates that "... [T]he ratio of private to public dollars to the arts varied over the 90s from about 12:1 to 19:1. In general, it's increasing. Assuming reasonable marginal federal tax rates, indirect federal aid outweighs direct federal aid by about 15:1. Getting the total indirect/direct ratio is harder because imputing state and local taxes is [difficult]. But I figure it's about 6:1." (Arthur Brooks, e-mail correspondence with the author, February 2, 2002.)

had gone far beyond the United States in inventing tax provisions to pursue various aspects of cultural policy, most particularly with respect to the treatment of both movable and immovable heritage resources. He also demonstrates that the take-up of these provisions by the citizens of these countries has been far less than has been the case for such exemptions in the United States.<sup>17</sup> He has also advanced a broad set of propositions concerning the use of tax incentives in the arts and culture [Schuster (1987)]; in this paper the focus is on the inherent dilemmas in supporting the arts and culture (as well as other sectors more generally) that occur when providing support through indirect rather than direct means.

Most recently Schuster (1999) has revisited the field of indirect aid to the arts as part of a commemoration of the twentieth anniversary of the publication of Dick Netzer's book, *The Subsidized Muse*. In this paper he points out that much has changed as other countries began to take up the cause of American-style tax incentives, but that what has changed most is governments' willingness to craft narrowly defined and targeted tax incentives. Moreover, while tax incentives have previously been used to provide a wide swath of benefits to broadly defined groups of recipients, more recent incentives have been more carefully crafted to target specific policy goals and to assure that those benefits are limited to intended beneficiaries. Such targeted incentives seem particularly well suited to overcoming resistance to behavior change, as will be seen further below.

While much of the discussion of tax policy that emanates from the cultural field is often based in advocacy more than in analysis, an occasional thoughtful contribution is made from within the field. Stephen Weil, a lawyer and museum administrator, is one such contributor. In a 1991 paper he explores the politics of tax reform in the United States and its relationship to the maintenance of indirect government aid to the arts [Weil (1991)] and mounts a spirited defense of tax incentives, particularly the charitable deduction, as a preferred way of structuring government support.

Another component of the tax literature in cultural policy focuses on the tax treatment of the donation of appreciated property, particularly the donation of paintings to museums in the United States. A lot of ink has been spilled on the comparative tax treatment of collectors and artists with respect to the donation of works of art, but most of this literature counts as straight advocacy rather than considered analysis.<sup>18</sup>

The particular problems related to heritage resources and the solutions that might be offered by various tax incentives have also commanded some sustained attention.

<sup>17</sup> Indeed, an increasingly large number of countries, particularly in Western Europe, have implemented American-style deductions for charitable contributions. [See also Weisbrod (1991); International Center for Not-for-Profit Law (2002); Inkei (2001); International Federation of Arts Councils and Culture Agencies (2003).] These tax regimes differ considerably with respect to floors, ceilings, and approved recipients of deductible donations, but the basic principle nearly always remains the same: a contribution for charitable purposes may be deducted from one's income before calculating the income tax that one owes. The exceptions are those countries that have opted for tax credits instead of contribution deductions (see below).

<sup>18</sup> Compare, for example, Hawkins (1988) with Feld, O'Hare and Schuster (1983, pp. 11–16, 44–48, and 220–223) and Fullerton (1991, pp. 206–207).

Alvarez (1988), Schuster (1986), and Netzer (1997) provide overviews of the relationship between tax policies of various types and preservation and conservation. Fogelman (1990) considers how the United States might adopt a set of tax rules that would foster preservation of the heritage based on the British system for protecting the national heritage. Schuster, de Monchaux and Riley (1997) discuss tax incentives in the context of a fuller presentation of the range of government tools used in preserving the built heritage. Interestingly, this cultural subfield, more than others, has led to the broad publication of information leaflets by governments seeking to explain the application and enforcement of the relevant tax rules and tax incentives. Finally, in his overview of economic policy issues in the arts O'Hagan (1998, pp. 104–130) devotes a chapter to questions of taxation. This chapter provides a summary and update of what is known about the role of taxation in cultural policy. Notably, it includes a section on the value added tax (VAT), which has been underemphasized in other work (including the current chapter).

#### 2.4. Putting tax policy back on the cultural policy agenda

Two factors have recently focused attention firmly back on tax policy. First, with the widely held perception that there has been a general decrease in the availability of direct public resources for culture, a search has begun for other sources of support and for ways to provide incentives for those other sources. Second, with the growth in the use of forms of decentralization, désétatisation, and devolution in cultural policy [Boorsma, Hemel and van der Wielen (1998)], increased attention has been paid to tax policy as a way of spreading decision making concerning public resources more broadly, since benefits in the form of forgone taxes follow the disaggregated decisions of individuals or organizations. Together, these factors have once again turned the field's attention toward cataloging and understanding the tax regimes currently in effect in various countries with one eye toward finding models of effective implementation and the other toward reconciling and integrating different regimes across countries.<sup>19</sup> Thus, the most recent contributions to the literature on tax policy and cultural policy are studies completed by the Regional Observatory on Financing Culture in East-Central Europe (the "Budapest Observatory") for the Council of Europe [Inkei (2001)] and a comparative study completed by the staff of the International Federation of Arts Councils and Culture Agencies (2003). These reports are, in turn, part of a broader literature that has focused on developing civil society with a strong nonprofit sector, particularly in Central and Eastern Europe. Thus, careful attention is being paid to the development of appropriate legal structures in these countries [International Center for Not-for-Profit Law (2002)].

Inkei (2001) captures nicely the current perspective of the field with respect to research on tax-based incentives in cultural policy. It is wary of individual and corporate influence, and laments the passing of high levels of central government support, but

<sup>19</sup> Much of the activity in this regard has been in the European Union, focused on facilitating the flow of artistic workers across national boundaries.

it also mixes optimism and advocacy with a healthy dose of skepticism. Inkei's report focuses on the tax treatment of corporate support with much less discussion of individual support, reflecting this report's emphasis on Central and Eastern Europe where it is believed (hoped?) (feared?) that corporations will provide the first important entrée to private support. Accordingly, the report carefully distinguishes between sponsorship – a commercial transaction – and a donation – a charitable gift – in the Anglo-American tradition, a distinction that is critical to maintain when trying to untangle current laws and procedures (sponsorship, for example, usually incurs VAT because it is construed as a purchase of a service, while donations do not; sponsorship may also cause a taxable “profit” for the organization in receipt of the sponsorship). In certain national contexts the tax treatment of corporate sponsorship looms large in cultural policy, as corporations move more and more sponsorship money into the arts and culture and as countries search for ways to get corporations to pick up an increasing proportion of cultural support.

Because private support for culture is a small part of private support to all kinds of charity and nonprofit causes, governments and social scientists tend to address the issue of private support to culture in the wider contexts of philanthropy, charity, the development of civil society, and the definition and regulation of nonprofit organizational forms. Inkei (2001) provides a rough estimate that internationally perhaps 10 percent of private philanthropy flows to culture. This estimate seems generous given that a much more reliable estimate for the United States calculates that in 2001 5.7 percent (\$12.14 billion) of total contributions to nonprofit organizations (\$212.0 billion) went to the arts, culture, and humanities [AAFRC Trust for Philanthropy (2002, p. 11)]. It is possible that because national laws restrict tax incentives to a narrower range of recipients and that cultural organizations are often designated recipients, the relative portion of tax-incentive-eligible donations flowing to the arts and culture is higher elsewhere. It is also possible that the arts and culture are used as the wedge sector to argue for tax incentives, thus becoming at least for a while one of the few sectors to benefit.

Whatever this proportion, one of its implications is that less analytic attention has been paid to the cultural sector (or, for that matter, to any particular sector) than has been paid to the economics of philanthropy as a whole.<sup>20</sup> Thus, much of the work that I am forced to rely upon later in this chapter will be based in the broader literature of tax policy with little reference to cultural policy. Moreover, much of what one sees in the arena of indirect aid can be better explained through political science than through economics. Nevertheless, economics remains a critical tool of analysis.

### 3. The purpose of tax policy vis-à-vis the arts and culture

Tracking the impact of tax policy on the arts and culture would be simplified if tax law had a single objective. But, as he points out in his remarkably succinct article on the

<sup>20</sup> See further Chapter 37 by Katz in this volume.

theory of the tax law treatment of nonprofit organizations in the United States (and, by extension, elsewhere), *Simon (1987)* identifies what he calls the four main functions of tax law: a financial support function providing assistance to the nonprofit sector; an equity function embodying the principles of fairness, redistributive equity and the pursuit of the public interest through the resulting incentives; a police function assuring that trustees, organizational managers, and donors are operating broadly within the public interest; and a border patrol function assuring that distinctions between the nonprofit sector and the for-profit and governmental sectors are clearly defined. But all of these are, of course, ancillary functions of tax law, the primary function being the raising of revenue for public purposes. Here is where the question of tax policy gets tricky because on the one hand there is a desire to make tax revenue collection as clean and straightforward as possible, yet on the other hand there is a desire that grows out of the concerns of various sectors of policy such as cultural policy to use tax policy as a locus for providing incentives or disincentives.

### *3.1. Defining the tax base v. providing incentives*

At the heart of any policy concern vis-à-vis tax law is a concern about incentives: what economic behavior has been given an impetus and how well has that impetus worked? But here lies a fundamental debate that has always affected – and continues to affect – the framing and analysis of tax policy. What appears to one person as a tax incentive, appears to another merely as a way of defining the base upon which taxes are to be collected. Where one comes down in this debate has shaped rather dramatically how one feels about the use of tax law in this manner.

In the United States the Second Revenue Act of 1917 provided for the deduction from income of donations made to charitable organizations. This was done for several intertwined reasons [*Wallace and Fisher (1977, pp. 2131–2132)*]. The act substantially increased income tax rates to fund America's involvement in World War I – the income tax itself had only been implemented a few years earlier. Yet there was a fear that this tax increase would curtail private support of public charities. It was argued that if private funding were to decrease considerably, the public sector would have to step in with increased support. At the same time it was also argued that charitable donations came from whatever surplus an individual felt that he or she had in income and that taxation would unduly limit that surplus, which might otherwise be deployed to public good through charitable contributions. Others argued that an income tax without a deduction would have the effect of imposing a tax upon the charitable recipients themselves, as donors would give less than would otherwise be the case because of the increased taxation on income. These arguments all amount to more or less the same thing, arguing in essence that the charitable deduction was actually a “hold harmless” provision, leaving charitable contributions untouched by income taxation. Some others went so far as to assert that individuals should not be taxed on that part of their income donated to charity, arguing in effect that income tax should only be imposed on consumable income.

These ideas about tax deductibility for charitable contributions persist in much of the discussion today. And they all lead to more or less the same place – the conclusion that the charitable deduction is best framed and best understood as *tax-base defining* rather than as an *incentive*, effectively taking questions of the actual incentive effect off the table.<sup>21</sup>

Still, this conceptual debate has raged for decades, particularly since the work of the [Commission on Private Philanthropy and Public Needs \(1977\)](#). Volume IV of its published research papers provides a useful summary of this debate, as does [Simon \(1987, pp. 73–76\)](#).<sup>22</sup> But the battle was most directly joined with the publication in 1973 of Surrey's book, *Pathways to Tax Reform*. As has already been mentioned, in this volume Surrey laid out the process of "tax expenditure analysis" in which he analyzed taxes forgone by the federal government for reasons unrelated to the purpose of the tax instrument itself as carrying the same fiscal and incentive effects as direct government expenditures. He also argued for the regular compilation of a "tax expenditure budget" which would detail the amount and the purposes of various forms of forgone taxes, a practice that has been adopted by the federal government as well as by many states.<sup>23</sup> This book, a seminal contribution to the analysis of the differences between direct and indirect government aid, shaped much of the later research, including [Feld, O'Hare and Schuster \(1983\)](#).

From an economic point of view, however, this philosophical debate is something of a diversionary tactic, obscuring the fact that changes in the various parameters of tax law – marginal tax rates, income brackets, income averaging provisions, and floors and ceilings on deductibility, not to mention outright changes in rules on deductibility – inevitably act to change the attractiveness of donations at the margin. In a system that allows the deduction of charitable donations from income before calculating the tax owed, an increase in marginal tax rates will make a charitable donation look less expensive (i.e., more attractive) to a donor. And a provision that would allow charitable deductions to donors who otherwise would only take the standard deduction (i.e., not itemize their deductions) would make charitable donations less expensive for taxpayers who take the standard deduction. On the other hand, a decrease in marginal tax rates would make donations more expensive (less attractive) for donors who itemize their deductions. Thus, whatever the origins of the tax provisions under consideration, there is considerable analytic leverage in understanding their incentive effects. One need look

<sup>21</sup> They also lead logically to the conclusion that a *deduction* from income rather than a *credit* against taxes is the appropriate way to structure this "hold-harmless" provision.

<sup>22</sup> See also the sources cited in [Feld, O'Hare and Schuster \(1983, footnote 8, p. 244\)](#).

<sup>23</sup> Tax expenditure budgets, where they exist and are done well, can be a boon for cultural policy researchers who are looking at the indirect side of government policy. In a recent report on the cultural policy of the State of Washington, for example, [Schuster et al. \(2003\)](#) benefited greatly from Washington's careful and detailed tax expenditure analyses [[Taylor \(2000\)](#)], but not all states make it their business to collect such detailed information. [Mikesell \(2002\)](#) provides a nice summary of the state of the art in tax expenditure budgets at the state level.

no further than to the furious lobbying emanating from the charitable nonprofit sector at moments of tax reform to understand that they see the question of receiving the benefit of forgone taxes primarily, if not exclusively, through the lens of incentives.

It is not then too large a step to evaluate the flow of forgone taxes in much the same way that one would evaluate direct subsidy; indeed, one can usefully consider forgone taxes as the functional equivalent of government subsidy. And, as [Simon \(1987, p. 75\)](#) points out, no less an authority than the US Supreme Court in a 1983 decision came down on the side of interpreting tax exemptions as subsidies: “[T]ax exemptions and tax deductibility are a form of subsidy that is administered through the tax system . . . [with] much the same effect as . . . [or being] similar to cash grants . . . [thereby constituting an expenditure of] public moneys.”<sup>24</sup> In more recent decisions, however, the Supreme Court has been inconsistent in this view, particularly when it comes to tax provisions affecting religious organizations, for which the constitutional principle of separation of church and state makes it more difficult to sustain the idea of tax provisions as incentives.<sup>25</sup>

Still, the debate between incentive provision and tax base definition has not entirely gone away.<sup>26</sup> And one would do well to remember that it is not yet true that “incentive” is seen as a neutral term in all national contexts. For example, [Inkei \(2001, p. 1\)](#) points out that while it is perfectly logical for the phrase “tax incentive” to be used to characterize such fiscal mechanisms in many countries, in some parts of Europe (particularly Central and Eastern Europe) this way of framing these policy mechanisms is somewhat controversial: “There is a feeling . . . that these approaches imply that the government is simply ‘bribing’ taxpayers to giving ‘protection money’ to culture.” Thus, the concept of tax forgiveness as an incentive as opposed to a base-defining provision remains controversial.

### 3.2. *Individual v. collective decision making*

Somewhat less evident in the debate about the effects of tax incentives is a question about the properties of various decision-making mechanisms in public policy. The trade-off between implementing policy through tax incentives as compared to more direct means can also be seen as one between individual and collective decision making. Few authors have addressed this trade-off head on [[Schuster \(1996, 2001\)](#)], and none except [Feld, O’Hare and Schuster \(1983, pp. 104–129\)](#) has brought any empirical analysis to bear on this question. Generally speaking, the use of tax incentives values individual over collective decision making: individual donors not only decide how their private net-of-tax contribution is to be allocated, but they are also accorded the right to make the same decision with respect to the forgone taxes in the transaction; museums and

<sup>24</sup> *Regan v. Taxation Without Representation of Washington*, 461 US 544-5 (1983).

<sup>25</sup> I am grateful to Alan Feld for pointing this out to me.

<sup>26</sup> See, for example, [Yarmolinsky \(2000\)](#).



other cultural institutions that own property in effect make their own decisions as to how much of a property tax exemption to take advantage of by their decisions concerning how to treat their land and buildings; and so on.

The policy-analytic task here is to compare the important attributes of one support system with another. The problem is disentangling the implications of the various decision-making mechanisms in use. Consider first direct aid. In the American context it is often assumed that the appropriate collective decision-making mechanism in direct funding of the arts and culture is peer review, in which a panel of one's peers evaluates and chooses among competitive applications for support.<sup>27</sup> But peer panel review is not the only form of collective decision making that is used. In some situations formulas are used to determine funding amounts; in others legislatures use line-item budgeting, thereby making funding decisions themselves; in still others the staff of cultural agencies make these decisions; and in yet others public referenda are used to determine the allocation of public resources [Frey and Pommerehne (1989); Frey (2000, pp. 115–129)]. Because of the intermingling of funds that arrive through a variety of mechanisms, in the end it becomes nearly impossible to separate the effects of one type of direct decision making from another. Nor are tax-incentive-based funding schemes the only form of individual decision making: publicly-funded vouchers by which individuals are given the ability to distribute collective resources according to their own preferences are another example,<sup>28</sup> matching grants are another, and self-directed allocations of a portion of one's tax payments yet another [Vajda and Kuti (2002)].

It may be that a society has an inherent preference for one type of decision making over all the others. Implicit in the move toward greater private influence in arts funding decisions is the assumption that the decentralized decisions made by a wide range of individuals would lead to results that would be superior to any of the candidate collective decision-making mechanisms. In other words, there may be a policy preference for a particular form of decision making, a preference that might outweigh any concern about economic effectiveness. Of course, one might also be concerned about exactly *what* decisions are made under each form of decision making. Do they align themselves with articulated public policy, or do they diverge from that policy, forcing public policy to compensate in some manner? Many of the idiosyncratic stories that are told about the eccentricities of individual donors to cultural institutions are evidence of this concern [Feld, O'Hare and Schuster (1983, pp. 130–178)].

Thus, as we turn to narrowly economic questions in the next section, it is important to remember that the *economic results* of any particular form of public support are not the

<sup>27</sup> Some attention has been paid to the attributes of peer panel review in cultural funding [Wyszomirski (1990); Independent Commission (1990); Galligan (1993); Brenson (2001)], though more attention has been paid to it in the field of science funding. There is much more work to be done here.

<sup>28</sup> With respect to vouchers, Schuster (1994) discusses the success of the Wintario Halfback Program, a voucher program in which losing lottery tickets in a dedicated lottery became vouchers at half their face value. With respect to matching grants – and their close affinity with tax incentive schemes – see Schuster (1989b).

only outcome of policy concern. We might value *particular forms* of decision making over others as a matter of policy, particularly if a comparison of *what* decisions are made under those various forms of decision making indicates that particular ways of arranging the system act more effectively to pursue articulated policy.

#### 4. The economic effects of tax exemption

##### 4.1. Introduction

The key economic question in considering the effects of tax deductions and credits lies in the cost–benefit analysis of that effect. Do the benefits to the public outweigh the costs? Once all of the other philosophical matters concerning direct versus indirect aid are resolved, this is the one question that remains. This section considers tax exemptions for charitable donations as but one example of tax exemptions vis-à-vis the nonprofit sector. The focus will be on charitable giving by individuals because, in the United States at least, individual charitable giving is a much more important source of support than are other sources of donations (Table 2); in 2001 individuals accounted for more than three-quarters of total charitable contributions (and an additional 8 percent through bequests).

These data reflect the importance of private initiative and the nonprofit sector in the provision of the arts and culture in the United States. Clotfelter (1985, p. 1) makes this point quite clearly:

“Although the interrelationships that have evolved between government, nonprofit institutions, and the legal structure are the result of hundreds of years of complex social development, it seems by no means accidental that this special reliance on nonprofit institutions and . . . favorable tax provisions have developed side by side.”

Table 2  
Total charitable contributions in the United States, 2001

	Contributions to all charitable organizations (\$ billion)	Percentage
Individual donors	\$160.72	75.8%
Bequests through estates	\$16.33	7.7%
Foundations	\$25.90	12.2%
Corporations	\$9.05	4.3%
Total	\$212.00	100.0%

Source: AAFRC Trust for Philanthropy (2002, p. 169).

Note: Unfortunately *Giving USA 2002* [AAFRC Trust for Philanthropy (2002)] does not provide sufficient information in its summary report to reconstruct comparable percentages for giving to the arts, culture, and humanities.

#### 4.2. *The price elasticity of individual giving*

The key to addressing the tax effect of an exemption for charitable donations is to notice that the tax incentive decreases the marginal price of giving: with no tax incentive it costs the private donor \$1.00 in after-tax income to get \$1.00 to a charity, but with a tax incentive it costs only \$1.00 times one minus the applicable marginal tax rate.<sup>29</sup> Seen from this perspective, the fundamental question is whether, in the presence of the tax incentive, total private net-of-tax giving goes up by more than the amount of forgone taxes. This reduces to a question of the price elasticity of giving: what percentage change in giving is engendered by a one percent decrease in the price of giving? In the literature this has been termed the “treasury efficiency” of the tax incentive.<sup>30</sup>

Economic theory would predict the sign of the price elasticity of giving to be negative – an increase in the price of giving at the margin is expected to result in a decrease in the total contribution.<sup>31</sup> But the question of whether or not that elasticity will be less than  $-1.0$ , the boundary between elasticity and inelasticity, is an empirical one. If the absolute value of the elasticity is greater than one, giving is price elastic – it more than responds to changes in the marginal response – and the total contribution increases by more than the tax expenditure alone. In other words, the tax expenditure has sufficiently leveraged increased private funds to justify the cost of the incentive. It is this potential that makes tax incentives for charitable donations an attractive proposition for governments: the possibility that a tax incentive will actually increase the revenues flowing to the favored sector in an amount that is greater (and, hopefully, much greater) than the forgone taxes.

As one might imagine, economists have attacked the estimation of the price elasticity of giving with considerable energy and ingenuity. Unfortunately, even after some forty years of research on this question a clear consensus has yet to emerge. In the earliest summary of the existing research of which I am aware, Clotfelter (1985, pp. 56–63) summarized sixteen economic analyses of charitable contributions in the United States. While the earliest of these, conducted in the 1960s, estimated low price elasticities – negative but with an absolute value of less than one – many later studies returned price elasticities that were negative and with an absolute value greater than one, suggesting an emerging consensus that charitable contributions were price elastic and did respond (at least in the United States) as had been predicted and hoped. Fullerton (1991), in his attempt to apply the accumulated research findings on tax incentives to arts museums, concurred:

<sup>29</sup> In a regime based on a tax deduction, this is simply one minus the donor’s marginal tax rate; in a tax credit system, it is one minus the rate at which the credit is being offered.

<sup>30</sup> A clear and concise summary of these and other issues is contained in Cordes (1999).

<sup>31</sup> One needs to be particularly careful in interpreting the accumulated literature on this point because some authors either have neglected to carefully record these negative signs or have chosen to represent the price elasticity with respect to marginal tax rate rather than with respect to price. Thus, negative signs in some studies have the same meaning as positive signs in others. In this chapter, I have chosen to maintain the use of negative signs by focusing on the price elasticity of giving.

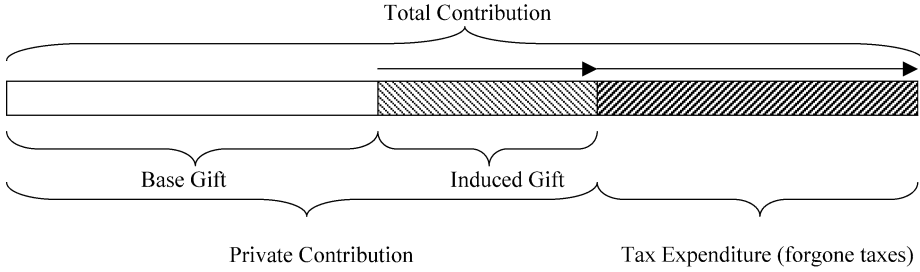


Figure 1. Components of the total contribution.

“Early research tended to find elasticities less than 1 [in absolute value] . . . but more recent work has found larger responses . . . These price elasticity estimates are all close to each other, especially considering the substantial differences in the nature of the data and the level of aggregation. More recent research concludes that the price elasticity of charitable giving is greater than 1 [in absolute value], though perhaps not much greater than 1. This implies that donees receive more than the government loses in tax revenue” (p. 212).

But then methodological questions were raised. In 1990 Steinberg summarized a new set of studies using panel data rather than cross-sectional or time series data in their estimates. These studies tended to conclude that giving was inelastic with respect to price, rather sobering news for the proponents of tax incentives. In a more recent summary Steinberg (2003) makes reference to ten more cross-sectional studies and eight more panel studies, concluding that the earlier pattern continues to hold, with cross-sectional data suggesting price elasticity, and panel data suggesting inelasticity.<sup>32</sup> In a recent paper not summarized by Steinberg, Tiehen (2001) uses an intermediate dataset: a cohort panel constructed from a series of household surveys on charitable giving. Her estimates of the price elasticity vary between  $-0.9$  (somewhat inelastic) and  $-1.1$  (moderately elastic). Thus, the debate continues.

To understand more fully the implications of these results, it is useful to return to Feld, O’Hare and Schuster (1983). They approached the analysis of charitable contributions in the presence of tax incentives by carefully disaggregating the total charitable contribution into its component parts (Figure 1): The “base gift” is the gift that a donor would have made in the absence of any tax incentive. The “total contribution” is the total gift that she would choose to make in the presence of the tax incentive. The components of the total contribution include the base gift plus any “tax expenditure” that has been realized in the transaction plus any increase that might have occurred in the

<sup>32</sup> Steinberg (2003) points out that there has been some suggestion that the disparity between these two approaches may be partially due to misspecification of the error distribution in panel studies rather than in analysts’ ability to control for unobservable individual-specific effects with panel data [Barrett and McGuirk (1992)].

Table 3  
The components of individual charitable giving, United States, 1973 (\$ millions)

	Private contribution		Tax expenditure	Total contribution
	Base gift	Induced gift		
Total charitable giving	\$12,027 69.0%	\$840 4.8%	\$4570 26.2%	\$17,437 100.0%
Giving to cultural institutions	\$118.56 36.6%	\$24.04 7.4%	\$181.6 6.0%	\$324.2 100.0%

Source: Feld, O'Hare and Schuster (1983, p. 43) analysis of data from the Commission on Private Philanthropy and Public Needs, National Study of Philanthropy, 1974.

Notes: Estimated tax expenditures do not include capital gains forgone on gifts of property.

A price elasticity of giving of  $-1.24$  as estimated by Feldstein (1975a) is used for all income classes of donors and all charitable sectors to arrive at these estimates.

donor's private net-of-tax gift; this last component can usefully be referred to as the "induced gift". The sum of the base gift plus the induced gift is the donor's total "private contribution".

What would be most desirable (and expected) would be for the induced gift to be positive, indicating that the tax exemption is working as intended, but it is also possible that the induced gift will turn out to be negative, indicating that the donor has taken advantage of the tax expenditure to reduce her private contribution below the base gift.<sup>33</sup>

Feld, O'Hare and Schuster (1983) estimated separately the base gift, the induced gift, and the tax expenditure components of total giving to all charities and total giving to cultural institutions. This is the only attempt of which I am aware in either the arts literature or the more general literature on charitable contributions to make such estimates. Using data from the 1974 National Study of Philanthropy and an estimated price elasticity of  $-1.24$  for all income classes and for all charitable sectors – an elasticity chosen to represent the apparent academic consensus on the price elasticity at the time in which they were working – they arrived at the set of estimates shown in Table 3. Note, however, that these estimates are for a year in which marginal tax rates were much higher than they are now in the United States and that they are based on a dataset that has been criticized [O'Hagan (1998); Jencks (1987)].

While too outdated to provide much insight into the working of American tax incentives some thirty years later, this table is at least helpful in highlighting an issue that has not received sufficient treatment in the literature on the use of tax incentives in public policy: the leveraging effect of the tax incentive. The estimates in Table 3 indicate that in 1973 it took a tax expenditure of nearly \$4.6 billion to induce additional private giving of \$840 million, a ratio of 5.44:1.00. In other words \$5.44 in tax payments

<sup>33</sup> The induced gift might also turn out to be zero, indicating that the donor has simply increased her contribution by the amount of the forgone taxes, leaving her with the same private net cost for the total contribution.

were forgone for every additional private donation of \$1.00 to charity. To be sure, the \$5.44 went to charity as well, directed by the donor within whose contribution it was included, but a critical question remains: Is it necessary to forgo \$5.44 to get \$1.00 to flow (and an additional \$6.44 to arrive), or are there other ways to structure an incentive that would accomplish the same result at lesser public cost?<sup>34</sup> Because donors to the arts and culture have higher incomes than the average donor to charity, these figures are even more extreme for culture. A tax expenditure of \$181.6 million resulted in an additional induced private gift of \$24.04 million, a ratio of 7.55:1.00. Thus, in the field of the arts and culture \$7.55 in taxes were forgone in order to get each additional \$1.00 to flow in private contributions.

#### 4.3. Tax incentives as matching grants

Framed in this way, a tax incentive for charitable contributions looks very much like a matching grant. Indeed, this is the whole premise of tax expenditure analysis: that tax breaks can be construed and evaluated as though they were direct aid programs. It is therefore instructive to look at the practice of matching grants to support the arts and culture, an attribute of cultural support that is also seen to be American in origin.<sup>35</sup> Without turning in another analytic direction entirely, suffice it to say that when government agencies (or legislatures themselves) have implemented matching grants, they have tended to design them with leveraging ratios such as \$1.00 public money to \$1.00 “new” private money or \$1.00 public money to \$2.00 “new” private money. I am unaware of any that have been designed at anything near the \$7.55 public money to \$1.00 “new” private money ratio implicit in the tax incentive system for the arts and culture as analyzed by Feld, O’Hare and Schuster summarized above. Such a program design, if explicit, would be open to criticism as a poorly designed use of public money if its actual goal were to leverage increased private resources.

An economist considering the design of matching grants would of course focus on many of the same questions addressed here. What is the level of private giving in the absence of the matching grant (the “base gift”)? How much “new” money is attracted from other sources to the arts and culture that would not otherwise have come in the absence of the offered match (the “induced gift”)? How much does the public sector have to give up – this time in direct expenditure – to get the change in giving (the “tax expenditure”)? And finally does the use of more decentralized individual decision making in

<sup>34</sup> A more recent study by Price Waterhouse, cited by Brody and Cordes (1999, pp. 146–147), simulates the changes to total charitable contributions that would occur if the tax deduction were removed completely. Under the assumption that donations are “fairly responsive to changes in the after-tax cost of giving”, they estimate that this would reduce private giving by 32 percent including both the induced gift and the tax expenditure. While they present estimates by subsector, their estimates are arrived at by simply applying this 32 percent decrease to each of the subsectors of charitable giving, a not entirely satisfactory approach.

<sup>35</sup> For a fuller discussion of matching grants than is possible in the current chapter, see Schuster (1989b, 2001). McDaniel (1977) provides a nice presentation of how a more explicit matching grant system might be designed to replace the charitable deduction.

the matching grant system make it preferable from the public policy point of view? On these questions, [Eckel and Grossman \(2003\)](#) have conducted an experiment to ascertain how individuals respond to a subsidy structured as a rebate as compared to a mathematically equivalent one structured as a matching grant; they find that “contributions are significantly higher with matching subsidies than with rebate subsidies” (p. 681). If these results can be replicated, they suggest yet another reason besides transparency for restructuring indirect aid as direct aid – such a switch may also be more effective from a treasury efficiency point of view.<sup>36</sup>

#### 4.4. *Price elasticities by charitable sector*

Do different charitable sectors face different price elasticities of giving? [Feldstein \(1975b\)](#) has estimated price elasticities for various sectors, though not for the arts and culture (column 1, [Table 4](#)). Interestingly, giving to religious organizations is inelastic with respect to price whereas giving to each of the other sectors is price elastic. This is sensible, as support of the church is considered more as an obligation than as a voluntary contribution in many sectors of society. Moreover, many donations to religious organizations are more concentrated among low-income individuals (including individuals who take the standard deduction rather than itemizing their deductions and for whom any charitable exemption provides no incentive) than are donations to other charitable sectors. Giving to cultural organizations, on the other hand, is likely to be most similar to giving to other sectors favored by higher income individuals such as educational institutions or hospitals. If this assumption is true, then [Feldstein’s](#) work suggests that a one percent decrease in the price of giving should result in more than a two percent increase in giving to cultural organizations.

#### 4.5. *The income elasticity of giving*

The price effect is accompanied by a second effect – the income effect – which also must be accounted for. A decrease in the price of giving is accomplished through a decrease in the donor’s marginal tax rate. If the donor’s marginal tax rate decreases, his taxes decrease and his net after-tax income therefore increases. As a result, the donor’s contribution may go up. Thus, one also needs to model the income effect on donations.

[Feldstein \(1975b\)](#) has estimated that giving is only elastic with respect to income for educational institutions and hospitals (column 2, [Table 4](#)). But because his data do not allow the separation of arts and cultural organizations, it is difficult to apply his results to this sector. I am unaware of any other attempts to disaggregate income elasticities by charitable sector, so there is little further information concerning what the impact of income changes might be on donations to the arts and culture.

<sup>36</sup> But note that in the United States, depending on the design of particular funding programs, charitable donations may be eligible for a tax incentive *and* attract a matching grant, thus providing an even greater incentive for giving.

Table 4  
Price elasticity of giving, various charitable sectors

Type of donee	Price elasticity of giving	Income elasticity of giving
Religious organizations	-0.49	0.63
Educational institutions	-2.23	1.22
Hospitals	-2.44	1.08
Health and welfare	-1.19	0.85
All others	-2.63	0.65

Source: Feldstein (1975b) cited in Fullerton (1991, p. 213).

Note: Negative signs have been added where necessary to be consistent with the mode of presentation in this chapter.

In a summary of sixteen economic analyses of individuals' total charitable contributions in the United States, Clotfelter shows that while the earliest (pre-1970) study of charitable giving estimated income elasticities of giving well over 1.00 – from 1.31 to 3.10 depending on income class – later studies have generally estimated positive income elasticities but ones that are less than 1.00 [Clotfelter (1985, pp. 56–63)]. Steinberg (1990) summarizes an additional 24 studies that benefit from new methodological approaches and the availability of new datasets, particularly panel data. In these studies, income elasticity is uniformly positive but nearly without exception less than +1.00, indicating that charitable giving is inelastic with respect to income.

Apart from price and income elasticities, the literature on tax incentives also includes studies of other independent variables such as the donor's income level, her educational level or her relationship to the recipient organization to determine what their effect is on charitable giving. Other studies focus on alternative dependent variables to look at other effects of tax provisions. One vein of this research is concerned with the impact of tax law on charitable bequests, another has looked at how changes in tax law affect the timing of donations, and yet another asks about the relationship between volunteering and making contributions – to what extent do changes in tax law lead to changes in the relative amount of time dedicated to volunteering, and how do those changes interact with changes in charitable donations? Steinberg (2003) provides a short summary of the relevant results on each of these questions.

#### 4.6. The economics of corporate contributions

Some econometric work has been done on corporate contributions. Though the impact of these contributions in the United States is considerably less than the impact of donations by individuals – as we have seen, corporate donations were only 4.3 percent of total charitable giving in 2001 as compared to 75.8 percent for individual giving – it is worth considering the available findings because of the general (false) impression that



Table 5  
Income elasticities for corporate contributions by charitable sector

	Art	Civic	Education	Health	Total
Linear model	+0.82	+0.58	+0.66	+0.24	+0.56
Double log model	+0.80	+0.75	+0.97	+0.82	+0.85
Full Box-Cox model	+0.86	+0.78	+0.92	*	+0.81
Single Box-Cox model	+0.80	+0.75	+0.94	+0.78	+0.84

Source: Navarro (1988, pp. 71–72).

\*The model failed to converge for this category.

corporate donations do (and must) play a critical role in a tax-incentive-based support system.

In the United States, charitable contributions by corporations are deductible from income prior to calculating taxes owed in much the same way that an individual's contributions are deductible. The primary differences are that corporate tax rates are much lower than the highest marginal tax rates for individuals and that ceilings on contributions are lower. Generally speaking, models of charitable corporate giving do not focus on the individual firm; rather, they use data aggregated by industry groups or by corporations as a whole. Once again, Clotfelter (1985, pp. 203–205) provides a summary of the results of a set of studies that have used a wide variety of mathematical forms and data sources. The two studies that are based on time-series data estimate price elasticities ranging from  $-1.03$  to  $-2.00$  depending on the model specification, suggesting that charitable organizations receive more in contributions from corporations than the government forgoes in taxes. These two time-series studies disagree on the question of the income elasticity of corporate giving: one estimates income elasticities in the  $+0.53$  to  $+0.63$  range, but the other estimates income elasticities that are greater than  $+1.00$ . Once again, cross-sectional data produce different results from time-series data. For a wide variety of cross-sectional studies with a wide range of model specifications, the income elasticity is estimated as being positive, but lower than  $+1.00$ .

Navarro (1988) also takes a look at the income elasticity of corporate contributions. First, he summarizes earlier studies – many of the same ones summarized by Clotfelter (1985) – that calculated income elasticities ranging from as low as  $+0.03$  to as high as a surprising  $+2.37$ .<sup>37</sup> Navarro's work is one of the few attempts to disaggregate elasticities by sector (Table 5).

According to his linear model, “only the arts category appears to be at all elastic” but with respect to each of the other model specifications, corporate contributions appear to be similarly moderately elastic across all charitable sectors. He concludes (p. 73):

<sup>37</sup> One study even reports a negative income elasticity in one of its ranges of  $-0.27$ !

“The finding that corporate contributions are moderately income-elastic has implications for both the philanthropic sector, which relies on corporate largess, and for the federal government, which has some responsibility to meet welfare needs. For the philanthropic sector, this finding means that charitable organizations and their recipients are as exposed to variations in the business cycle as for-profit institutions. This vulnerability underscores the need for aggressive fundraising efforts, particularly during recessionary times. For the federal government, this finding would appear to vitiate the argument that corporations, rather than the federal government, can be relied upon to meet social welfare needs in times of recession or depression.”

An analytic understanding of corporate charity is further complicated by the rise of corporate sponsorship, particularly in the field of the arts and culture. It is helpful to keep clear the distinction that is generally drawn between corporate *sponsorship* and corporate *philanthropy*. Corporate sponsorship is the expenditure of corporate funds to advance the image of the corporation; typically, it is not tied to the sale of particular products or services but is part of the corporation’s advertising expenditures (and therefore typically deductible without limit from the corporation’s income before calculating taxes). Corporate philanthropy on the other hand is given with no expectation of a return to the corporation other than increased goodwill; corporate philanthropic contributions are only deductible from income to the extent allowed by national laws (which vary widely on this question).

Any number of countries that have moved to introduce tax-based incentives for private support for the arts and culture have begun with incentives targeted at corporations rather than at individuals. This seems to have been done for two reasons: first, an attempt to model themselves after the “American system”, which is widely (mis)perceived to be primarily constructed around corporate support, and second, a belief that corporations provide the easiest place to begin to attract increased private support, a belief which in turn has led to an emphasis on sponsorship rather than philanthropy. These countries have concluded that such a system has to be built on clear self-interest – in this case the self-interest of the corporate donor. This has led to considerable confusion between the ideas of “sponsorship” and “philanthropy”, so much so that in many countries “sponsorship” has been adopted as the more general term to describe both types of support. Not surprisingly, this has created nagging analytic problems for researchers trying to untangle the economics of corporate support.<sup>38</sup>

#### 4.7. Studies of charitable giving in other countries

To what extent are the American results summarized here replicated in the other countries that now provide tax incentives for charitable contributions in one form or an-

<sup>38</sup> See Schuster (1985b) for a discussion of this issue.

other?<sup>39</sup> Most of these countries have adopted the income tax deduction following the American system, though often with more restrictive limits on the class of eligible recipients and on the minimum and maximum donation for which a tax benefit can be realized. This system allows a donor to deduct her charitable contribution from her income before calculating the income taxes that are owed. As has already been suggested above, in such a system the donor's incentive is a function of her marginal tax rate, which in turn is a function of her income; thus higher income donors benefit from higher price incentives. Other countries including Canada (which has systematically been replacing deductions with credits), Hungary [Vajda and Kuti (2002, p. 4)], and Israel make use of the tax-credit system of providing incentives. This system establishes a particular percentage at which the credit will be offered. In effect, the donor multiplies that percentage by her charitable contribution and then deducts that credit from her tax bill. In such a system, the price incentive is the same for all donors eligible for the credit. Finally, several countries including the UK, Ireland, and Denmark make use of a third system, the deed of covenant, which allows charities to recoup taxes already paid to the government by a donor when a donation is actually made [Schuster (1989a)].<sup>40</sup> Some specific examples of the international move towards tax-based incentives are discussed in Section 5 below.

Studies of the effectiveness of these systems have been undertaken in several countries.<sup>41</sup> In two early studies, Hood, Martin and Osbert (1977) analyzed Canadian giving, while Pacque (1982) analyzed giving in the Federal Republic of Germany. In both cases their price elasticity estimates were broadly consistent with those estimated for US data. But later studies for the UK and Canada have concluded that giving appears to be price inelastic in these countries [Steinberg (1990, p. 74)]. Using UK data Jones and Posnett (1990) found that the price of giving was a large and statistically significant determinant of *making* a charitable gift, but it was not determinant of the *size* of the gift.<sup>42</sup> Two recent studies of charitable giving in Singapore – Wong, Chua and Vasoo (1998) and Chua and Wong (1999) – report price elasticities that are quite high, especially if those donations are given to organizations that are members of the Community Chest, which itself

<sup>39</sup> A number of summaries of tax incentives by country have been published, though they are generally incomplete and often suffer from poor descriptions of the tax provisions actually in force. See, for example, Weisbrod (1991), International Center for Not-for-Profit Law (2002), Schuster (1986), and Inkei (2001); the last two are particularly concerned with tax incentives as they impact the arts and culture. The International Federation of Arts Councils and Culture Agencies (2003) has also undertaken a survey of research on the encouragement of philanthropy in the arts; this survey is available on their website ([http://ifacca.org/en/organisation/page04\\_research.html](http://ifacca.org/en/organisation/page04_research.html)) and is updated as further information becomes available.

<sup>40</sup> More recent changes in tax law in the UK include elements of the American charitable deduction, which have been implemented on top of and in addition to the more traditional structure of the British deed of covenant.

<sup>41</sup> I am grateful here to Steinberg (1990) who provides much of this summary.

<sup>42</sup> But note that this study was of a rather different system of offering tax incentives – the deed of covenant – so it is difficult to apply these results to a system reliant on charitable deductions. For a detailed discussion of the deed of covenant, see Schuster (1989a), but note once again that a variety of other tax-based incentives for charitable giving have been added to British tax law since 1989.

serves as an additional source of charitable contributions to nonprofit organizations. On the other hand, in these studies donations are estimated to be relatively inelastic with respect to income.

#### 4.8. *Who pays and who benefits? The question of incidence*

The question of who pays for and who benefits from government support for the arts has been the topic of considerable debate but little actual research. In the words of Feld, O'Hare and Schuster (1983, p. 71): "[T]wo inconsistent views concerning the arts have developed: (1) cultural institutions represent a gift from the rich to the rest of society, and (2) all of society pays to support an entertainment for the rich." As we will see, the limited research on government support for the arts more generally and on indirect support through tax incentives more narrowly suggests that neither is the case. Three sources inform this debate: Feld, O'Hare and Schuster (1983, pp. 71–103), Netzer (1992), and, in a rather different manner, Cordes (2004). What is surprising about the results of these studies is that, despite findings that seem to be beneficial to the nonprofit arts, the field has not picked up on them.

Feld, O'Hare and Schuster worked with four datasets that in retrospect appear to be rather primitive: the 1975 *Americans and the Arts* survey by the National Research Center of the Arts, a subsidiary of Louis Harris Associates; the 1974 *National Study of Philanthropy* commissioned by the Commission on Private Philanthropy and Public Needs; the 1974 Ford Foundation report, *The Finances of the Performing Arts*; and the 1975 report, *Museums USA: A Survey Report*, commissioned by the National Endowment for the Arts from the National Research Center of the Arts. Paying careful attention to a number of analytical matters – drawing a distinction between visitors and visits, handling students differently from other adults, distinguishing between direct aid and the three components of charitable contributions (the base gift, the induced gift, and the tax expenditure), and using a set of commonly held benchmark assumptions about the incidence of the various sources of support – they were able to compare the incidence of indirect aid to direct aid to total income and to construct what is arguably still the most complete look at the question of incidence in arts funding, albeit an outdated one.

To those who paid attention the findings were astonishing:

"The existing system is mildly redistributive, transferring relatively small amounts of benefits down the income scale . . . On balance, income to the arts is paid for disproportionately by the very wealthy and is enjoyed more by the moderately wealthy and the well educated" (pp. 102–103).

"[T]ax expenditures redistribute benefits in roughly the same way as does the entire arts institutions' support system" (p. 94).

"[M]oving even significant amounts of current art support from the charitable deduction to any of the alternatives would not significantly change the distribution effects of the whole system" (p. 102).

Taken together these findings suggested that the arts funding system, both in its entirety and with respect to its various components, could not be criticized for being regressive, the result that the field most feared would emerge from any incidence analysis. But note that Feld, O'Hare and Schuster were working at a time when the marginal tax rates were significantly higher than they have been since. It is quite possible that this snapshot of incidence is no longer accurate, yet at least one other study suggests that the main conclusions still pertain.

In the early 1990s Dick Netzer was invited to revisit the question of incidence in arts funding in an edited volume that was to focus on this question across a number of nonprofit sectors. The resulting paper [Netzer (1992)] was able to take advantage of newer, more reliable datasets: the 1985 Survey of Public Participation in the Arts, commissioned by the National Endowment for the Arts, provided higher quality data on the recipients of benefits, and the 1985 Census of Service Industries provided a steadily improving source of data on the income profile of nonprofit organizations. Like Feld, O'Hare and Schuster, Netzer was careful to distinguish between the incidence of direct aid and indirect aid, though his analysis was not quite as disaggregated as the earlier study.

Netzer pointed out that redistribution is seldom a major goal of arts funding, and he took a rather skeptical view as to how the historical evolution of cultural institutions might affect how they actually function in this regard:

“[T]he redistributive goal from the outset often was very secondary to the primary goal of establishing clubs to make possible cultural experiences that would be enjoyed, if not solely by the very rich who were the initial organizers and backers, then by the relatively educated and affluent upper-middle-class” (p. 174).

Though not quite willing to conclude that the system is mildly progressive, Netzer's findings agree with Feld, O'Hare and Schuster:

“[T]he output of the arts and culture subsector of the nonprofit sector is consumed mainly by individuals and households whose income, wealth, and social attributes are well above the national averages. In this, the subsector is like its close relative in the nonprofit world, higher education. Like higher education, the distributional effects of the subsidies to the subsector would be considered objectionable by many people, were it not for the positive external benefits the subsector is believed to generate” (pp. 186–187).

“The bulk of the benefits of the activities of nonprofit arts and cultural organizations are realized by people in the upper half of the income distribution, but the frequent allegation that support of the subsector from tax funds and tax-deductible gifts is a transfer from those in the middle of the income distribution to those at the top of the distribution is a caricature” (p. 202).

More recently Cordes (2004) has conducted an “incidence analysis”, but one of a somewhat narrower sort. Working with data from the 1997 Economic Census and data from Internal Revenue Service Form 990 (a form required of over 7000 nonprofit arts

organizations) as collected and analyzed by the National Center on Charitable Statistics, he disaggregates the various sources of income (though it appears that he neglects to separate out the induced gift) in order to construct a picture of what types of institutions receive what types of income and in what amounts. Thus his approach to the benefit side of incidence focuses on type of organizational recipient rather than going one step further to try to associate audience volumes and demographics with organizational type.<sup>43</sup>

In asking how the benefits of public support, both direct and indirect, are distributed among arts organizations – an interesting public policy question in its own right – Cordes focuses on the age, the budgetary size, and the type of arts organization. His conclusions include the following:

“In terms of breadth of coverage, and also effective subsidy value, the charitable tax deduction remains as the single most important source of subsidy for most nonprofit arts organizations . . . A disproportionate share of total direct and indirect arts subsidies are [sic] estimated to be received by some arts providers, which, apart from art museums, do not include organizations engaged in providing what some might describe as ‘high art or culture’ . . . On balance, the overall pattern of subsidies seems neither to favor nor to penalize the entry of new nonprofit arts providers” (p. 229).

“The system provides larger total effective subsidies to larger than to smaller organizations” (p. 231).

Taken together, these three studies suggest that the system of arts support in the United States, largely based on the forgone taxes embedded in various tax incentives, cannot be faulted for its incidence. To be sure, one would not design the current system to pursue a goal of redistribution. Still, a number of interesting attributes of the funding system are revealed by incidence analysis, attributes that various policies might find attractive. Nevertheless, there is still considerable work to be done to research fully the question of incidence in the funding of the arts and culture.

#### 4.9. *Crowding in or crowding out?*<sup>44</sup>

Finally, the question of whether or not government support for the arts “crowds in” or “crowds out” other sources of support is another question that is typically asked of direct support but that can just as easily be asked of indirect support through the tax system. One of the main arguments often advanced in favor of direct government support to the arts and culture in the United States is that it provides a “Good Housekeeping Seal

<sup>43</sup> This paper also contains more recent set of estimates of tax expenditures for the arts and culture: Cordes’ conclusion is that roughly two-thirds of estimated total support of nonprofit arts and cultural organizations is in the form of tax incentives and preferences.

<sup>44</sup> For further discussion see [Chapter 35](#) by Netzer in this volume.

of Approval” to the recipient organization;<sup>45</sup> such recipients, it is claimed, can then use that seal of approval to attract other donations, particularly private donations taking advantage of tax incentives. In this view, direct public support can “crowd in” private support. Others have argued that the more support the government provides through direct means, the less private donors will see the need for their own donations. In this view, direct public support can “crowd out” private support. Either way, the result is of importance in an era in which direct government support is generally being rolled back.

What is the evidence in either direction? Once again, Steinberg (1991; 2003, pp. 284–285) provides useful summaries.<sup>46</sup> In his 1991 survey article he found that the available evidence and the accumulated research results suggested that there was modest and statistically significant crowding out, at least with respect to governmental social service expenditures. In his 2003 article (p. 284), he concludes, “New studies generally concur with the consensus of earlier studies that donations will replace only a small fraction of governmental cutbacks.” In other words the crowding-in effect is limited, at best. But he also points out that evidence suggesting a crowding-in effect is appearing from time to time with respect to certain charitable industries.

Several studies have focused on the arts and culture, and their results have made important contributions to the broader literature on crowding in and crowding out. Using data from the unpublished American Association of Museum’s 1989 Survey of Museums, Hughes and Luksetich (1999) consider the relationship between funding sources for art and history museums; they find evidence of crowding in. Their initial specification suggests that each \$1.00 of support to art and history museums from federal government sources results in an increase in private support – including individual, corporate, and foundation support – of more than \$9.33. An alternative nonlinear specification also supports the conclusion of crowding in – in this model each additional \$1.00 of federal funding results in additional private support of \$10.88 – but the crowding-in effect only stimulates private giving up to a certain point at which crowding out begins to take over. They attribute much of the observed crowding-in effect to the heavy use of matching grants in federal funding, which ties federal support to co-financing from other sources of support. In addition they cite evidence that increased federal support may cause some displacement of state and local government support. Finally they show differences across museum subsectors, concluding that decreases in federal funding would be likely to have a greater impact on art museums than on history museums as lower levels of government and private contributors adjust their giving to such decreases. Focusing even more narrowly within the cultural sector, Kingma and McClelland (1995) report crowding out with respect to donations to (nonprofit) public radio stations. They estimate that for this cultural subsector a \$1.00 increase in government support results in a \$0.15 decrease in donations.<sup>47</sup>

<sup>45</sup> The reference here is to *Good Housekeeping*, a popular American magazine that awards its “Seal of Approval” to goods and services that it determines to be of particularly high quality.

<sup>46</sup> A broader survey of “motivation crowding theory” is provided by Frey and Jegen (2001).

<sup>47</sup> See also Kingma (1996).

Brooks has written a series of papers exploring these questions in various segments of the arts. Using 1995 data on 91 public radio stations in the United States, he finds that public funding to these radio stations has a positive impact on private giving, but that this impact rapidly decreases as the level of government subsidy increases [Brooks (2003b)]. This analysis also suggests that increases in state tax rates correspond to significantly higher donation levels.<sup>48</sup> Brooks has also explored these questions using data on American symphony orchestras [Brooks (1999, 2000b)]. The first of these papers, using data from five major American symphony orchestras over twelve seasons, concludes that there is no causality in either direction. In the other paper Brooks uses American Symphony Orchestra League data for 253 symphony orchestras over eight concert seasons (1984–1991). Here a different model specification leads to a more nuanced result (p. 461): “At low levels of government funding, philanthropy might be encouraged, but beyond a certain point crowding out begins.” In another paper Brooks (2000a) considers four different charitable sectors; using federal spending on arts and culture from 1966–1997 he fails to produce a statistically significant result, though the results do show modest crowding in. Brooks attributes the ambiguity of these results to the fact that the dataset included only federal funding and not state funding, thus introducing a possible omitted-variable bias.

For the moment, the most responsible conclusion with respect to the crowding-in/crowding-out debate is the observation by Brooks (2000a, p. 213) that “the relationship between government subsidies and private philanthropy is highly dependent on the subsector, the level of government involved, and the specific dataset used in the analysis”.

#### 4.10. Conclusion

Much the same might be said of the broader literature surveyed above concerning the economic effects of tax incentives for charitable contributions: the results depend substantially on the model specification and the type of dataset used. For those who would argue in favor of tax-based incentives, at least with respect to charitable donations, there is considerable evidence in support of the view that such incentives would have the desired economic effect. But those who are dubious about the net effect can also find evidence to support their arguments. It is also clear, though not explicitly explored in the summaries that have been provided here, that cultural proclivities tend to be very important. If one looks at behavior in the aggregate rather than behavior at the margin, it is quite clear that in some countries, particularly the United States, there is a much

<sup>48</sup> Note, however, that the number of states that explicitly provide tax incentives for charitable giving is relatively small and that the effect of a state incentive is diluted by the structure of the federal income tax law. According to one report, only five states provide explicit incentives for charitable contributions in state income tax, but twenty-eight additional states simply adopt the federal deductions, which include a deduction for charitable contributions [Russell (2001)]. The interaction between federal and state tax incentives for charitable contributions is discussed further below (cf. footnote 65).



stronger tradition of private support for the arts and culture. When studying the data for other countries, it is necessary to keep in mind that despite the existence of tax-based incentives, the degree to which those incentives are taken up is a function of the extent to which the provision of the arts and culture is seen to be a private responsibility or a mixed government-private responsibility. Future research might do well to try to model these differences – for example, for a country like Canada where tax law is national but where there are important differences in charitable behavior between the French-speaking province of Québec and the English-speaking provinces. (These differences are even mirrored in the actual structure of government arts funding across the provinces.)

Despite the relatively ambiguous evidence on the effectiveness of tax-based incentives for charitable behavior, it is clear that there has been a strong international movement toward the provision of such incentives, often based on the American system of so-called “income tax deductions”.<sup>49</sup> In the final section of this chapter, I turn to a discussion of the proliferation of tax-based incentive schemes for the arts and culture throughout the world.

## 5. The international move toward tax-based incentives

Despite the ambiguity of the econometric evidence in the United States, the use of tax-based incentives is proliferating in other countries, not only to provide an incentive for increased private support in one form or another, but also to target a wide variety of specific cultural policy goals [Schuster (1986)]. Many countries now have such tax provisions, mostly providing income deductions, but some structuring the incentive as a tax credit.<sup>50</sup> There is considerable experimentation with the parameters of these incentives; floors, ceilings, “superdeductions” (deductions that are set at more than 100 percent of the actual contribution), and various forms of administrative oversight have all been adopted. Many countries are experimenting with a wide variety of tax-based incentives that go well beyond the realm of private individual or corporate philanthropy.<sup>51</sup> I have argued elsewhere [Schuster (1999)] that indirect aid is the terrain in which many of the most interesting innovations in the funding of arts and culture are taking place, though because of their hidden nature in the tax code they are not often recognized as such. In this proliferation one more often sees the influence of politics and advocacy than the influence of reasoned analysis. These points are most clearly illustrated through contemporary international examples; several will suffice.

<sup>49</sup> The widespread use of the unfortunate phrase “income tax deduction” in the United States has led to considerable confusion in other countries as to how the American charitable deduction actually works. As we have seen, it is structured as a *deduction from income* prior to calculating one’s tax liability, not as a *deduction from one’s taxes* (which would be more precisely described as a tax credit).

<sup>50</sup> For an argument in favor of the replacement of deductions with credits, see Feld, O’Hare and Schuster (1983, pp. 216–220).

<sup>51</sup> For fuller though somewhat outdated discussions of these themes, see Schuster (1986, 1987, 1999).

### 5.1. Chile

The 1990 Chilean *Ley de Donaciones Culturales* (Law for Cultural Donations) is an example of tax legislation specifically designed to stimulate increased donations to the arts by private corporations;<sup>52</sup> it established a unique and limited form of tax incentive that might be described as a “monitored tax credit” (to distinguish it from an automatic tax credit). The intent is to provide an incentive for the private sector to make donations for cultural projects by having the state “share” the cost of approved projects. In order to take advantage of this cost sharing a potential donor, in collaboration with the proposed recipient cultural organization, must present a detailed project proposal to a committee created to certify private donations. The proposal must meet a set of criteria for acceptability. If a project is approved, the corporate donor is allowed a 50 percent tax credit on the corporate donation.<sup>53</sup> In this manner, the state contributes half of the project’s cost – but in the form of forgone taxes. The state never has to write a check or make a grant, but it is able to determine acceptability on a case-by-case basis. In all publications, posters and other publicity related to the project, both the donor and the recipient must mention that the project has benefited from tax benefits under the *Ley de Donaciones Culturales*, making explicit the fact that the project is being supported in the form of taxes forgone by the state. Moreover, when the project involves exhibits of artistic or historic objects or performances, the recipient organization must certify that the exhibit or performance will be open to the general public with free admission. Thus a rather high level of public benefit is insisted upon in order to take advantage of this form of co-financing. This is a public policy attribute that does not often appear in tax-based legislation for the arts and culture.

### 5.2. Australia

In Australia, the Department of Communications, Information Technology and the Arts administers three different tax incentive programs, all of them of relatively recent vintage: the Register of Cultural Organisations, the Cultural Gifts Program, and the Cultural Bequests Program [Penhallurick (1998)].<sup>54</sup>

The Register of Cultural Organisations allows qualifying cultural organizations to offer donors the incentive of a charitable deduction. Much as in the United States, an Australian donor is allowed to deduct from taxable income donations of cash or of property<sup>55</sup> that are made to nonprofit organizations appearing on the register.<sup>56</sup> The

<sup>52</sup> The text of the law is available in *Diario Oficial de la Republica de Chile*, No. 33,706, 28 June 1990, p. 7; and the regulations concerning its implementation are available in *Diario Oficial de la Republica de Chile*, 12 February 1991, p. 3.

<sup>53</sup> Total annual credits for a corporate donor are limited to 2 percent of the corporation’s taxable income.

<sup>54</sup> For current information on these programs, consult the Department’s website: [www.dcita.gov.au](http://www.dcita.gov.au).

<sup>55</sup> Under this scheme donations of property are limited to property that the donor acquired within 12 months of the donation, and the deduction is limited to the purchase price or the current market value, whichever is smaller.

<sup>56</sup> This principle applies equally to individual donors, corporate donors, and trusts and foundations.

interesting component of this scheme is the explicit registration of eligible cultural bodies. These organizations must have as a principal purpose the promotion of literature, music, the performing arts, the visual arts, craft, design, film, video, television, radio, community arts, Aboriginal arts, or the movable heritage. They must be organized in a manner that is considered appropriate under Australian law, have a main purpose that is cultural, and maintain a separate fund for the receipt of public donations that must be used exclusively for the cultural purposes of the organization. Finally, to be entered into the register, organizations must be approved by the Minister and by the Treasurer. The register, established in 1991, is primarily made up of music and other performing arts organizations; public art galleries, museums, and libraries are covered by the Cultural Gifts Program and are not part of this register.

The Cultural Gifts Program (formerly the Taxation Incentives for the Arts Scheme) provides an incentive for gifts of significant cultural items to public arts galleries, museums, and libraries. Under this program a donor is entitled to a deduction from taxable income for the market value of the gift. Two valuations by approved appraisers are required, and the donor can claim the average of the two values as the deduction. The Commissioner of Taxation can adjust or disallow the deduction if the gift is given with conditions that prevent or delay the receiving institution from having clear title, custody, and control over the item or that involve a material benefit to the donor. Materials describing the program make it very clear that the Australian Government considers this to be a tax expenditure and even estimate the extent of that expenditure: "... the policy objective of the program is to attract significant cultural material into public collections and ... gifts are funded by the Government through the donor's tax deduction (estimated to be about 40 percent of the value of the gift)".<sup>57</sup> Gifts of cash to these public collecting institutions can also be deducted from income prior to calculating one's taxes.

The Cultural Bequests Program operates as a supplement to the Cultural Gifts Program. It is designed to encourage private collectors to bequeath nationally significant cultural items to public art galleries, museums and libraries by allowing the value of the bequest to be deducted from one's estate and by exempting the bequest from capital gains tax. This program is unique in that it goes one step further in limiting the tax expenditure by placing a limit on the total value of bequests that can be authorized annually. This limit, set at 5 million Australian dollars per year, constrains the amount of forgone taxes that can be spent via this tax expenditure and creates a situation in which the government may have to consider and decide among competing applications.<sup>58</sup>

Taken together, these programs provide an unusually constrained set of tax-based incentives. The Australian government through its registration, approval, and review

<sup>57</sup> See the Department's website listed above.

<sup>58</sup> A more refined limitation scheme apparently exists in Italy. *Trupiano (2002)* suggests that in Italy the deductibility of donations is limited through a ceiling *for each taxpayer* – individual or corporate – *and in the aggregate*. If, in any year, total donations exceed the aggregate limit, a "tax" is charged on the recipients of deducted donations in proportion to the total donations received by the organization.

mechanisms exercises a relatively firm hand on the system to assure that it operates within the grain of public policy in a way that other more automatic tax incentives find it difficult to do.

### 5.3. Singapore

Surprisingly, the idea of a “superdeduction” – one that provides a deduction of more than 100 percent of the value of the contribution – seems to be gaining some traction in several countries, perhaps because in certain societal and cultural contexts it is felt that an extra incentive is required to prime the pump in a system that has not heretofore had a tradition of individual or corporate support, or perhaps because advocates on behalf of the cultural sector have been particularly effective in promoting their own self-interest. In Australia, the Victorian Arts Minister has called on the Federal Government to implement a 125 percent deduction for gifts of new contemporary art works to public collecting museums [Coslovich (2002)]; Iceland is considering a provision that would allow a 200 percent income deduction for charitable contributions [Einarsson (2004)]; and even in the United States at one point the President’s Committee on the Arts and Humanities proposed that donations to the arts and culture be afforded a 150 percent tax deduction. The justification in the latter case was little more than that the arts were particularly worthy.

In Singapore, the Ministry of Finance’s 2002 Budget contained a set of tax measures to promoted philanthropy, including a superdeduction:<sup>59</sup>

- all cash donations made to Institutions of a Public Character would be eligible for a double tax deduction;
- donations made to those Institutions of a Public Character which offer naming opportunities to donors in return for their financial support would be eligible for a single tax deduction but not for a double deduction;
- any excess donations above the annual ceiling made by individuals or corporations could be carried forward for up to five years so that the incentive is not lost;
- the amount that an organization may spend for fundraising purposes is capped at 20 percent of the funds raised.

At the same time, the National Arts Council set up a system of Special Accounts to assist arts organizations in receiving tax-exempted donations. Donations are to be deposited into these accounts for the use of the various recipient organizations. While the reason for the creation of such a middleman organization is unclear, several elements in the legislation suggest possible explanations: (i) funneling deductible donations through the National Arts Council may help assure that the donations would actually be used for the projects and purposes for which the donor intended them; or (ii) the fact that the

<sup>59</sup> For the information on Singapore, I am grateful to Siu-Yuin Pang, Senior Manager of the Singapore Symphony Orchestra. Details are available from Ministry of Finance (2002). Note that not all of the tax measures proposed in that budget are summarized here.

National Arts Council is responsible for issuing receipts for these donations relieves the organizations of that burden, but also helps monitor the system to be sure that deductible donations are being made within the parameters of the law.

The notion that a special institutional structure might be necessary to serve as the conduit for charitable contributions shows up in several of the above examples. Such a structure has also been proposed by those considering how American-style incentives might be applied in other national contexts.<sup>60</sup> The creation of such new institutional structures, which to American eyes appears wasteful of resources, is an attempt to balance the provision of incentives with the requirement for extraordinary measures to prevent abuse.

#### 5.4. France

In December of 2002 Prime Minister Jean-Pierre Raffarin announced twelve measures to reform the treatment of charitable contributions and foundations in France [Morel (2005)]. The proposed measures included “better” tax incentives for both individual and corporate donors, more favorable tax treatment for foundations, and simplification of the rules governing the creation and operation of “public utility foundations”. The tax proposals were swiftly reviewed and amended by Parliament, and the law<sup>61</sup> was adopted on August 1, 2003 [Ministère de la Culture et de la Communication (2004)]. The changes in tax law embodied in this legislation were made because of a general sense that the preexisting incentives embedded in French law were not sufficient, particularly those related to corporate patronage. Prior to this legislation corporate patronage was deductible from corporate income; because the corporate marginal tax rate was 33.3 percent at the time, this provision set the price of giving at 0.667. The new legislation has replaced the tax deduction with a tax credit at 60 percent, thus lowering the price of giving to 0.40.<sup>62</sup> In effect, this legislation created a superdeduction for corporate charitable contributions.

It is too soon to know what effect this change in incentives has had, but it would seem reasonable to suggest that, given the high levels of government support for the arts and culture in France, one would not expect to see the sort of dramatic change in corporate giving that the legislation hopes for. The incentive interacts with the culture of giving of the country in question, with the result that the same incentives will have very different effects in different places.

<sup>60</sup> See *Borgonovi and O’Hare (n.d.)* for an example based on a consideration of how American-style tax incentives might be applied in the Italian context.

<sup>61</sup> *Loi du 1er août 2003 relative au mécénat, aux associations et aux fondations.*

<sup>62</sup> The new legislation also increased the limits on corporate charitable contributions that could benefit from these tax incentives from 2.25 or 3.25 per thousand of the company’s turnover (depending on the type of beneficiary) to 5 per thousand of turnover. These limits were accompanied by carryover provisions so that donations beyond the limit would be eligible for preferential tax treatment in a future year.

### 5.5. United States

In the United States over the past ten years there has been a movement to create state arts endowments. For the most part these endowments are intended to be separate from the preexisting state arts council in the hope that they will provide additional public and private resources for the arts and culture. They have often been created to serve a broader range of artistic, cultural, and heritage initiatives. The fiscal mechanisms that have been used to fund these endowments vary widely [Rafool (2004)], but several make explicit use of state-level tax incentives. Here we consider three examples drawn from the states of Oregon, Wisconsin and New Jersey.

Donations to the Oregon Cultural Trust are eligible for a 100 percent tax credit up to an annual limit of \$500 for individuals and \$1000 for households. Corporations are eligible for an 80 percent tax credit on an annual maximum of \$2500 in contributions.<sup>63</sup> These incentives are available only when the donor contributes an equivalent amount to a qualifying Oregon nonprofit cultural organization in the same year. The result of a 100 percent tax credit, of course, is that a donor making a donation is simply directing that contribution from his or her tax bill to the Cultural Trust; put another way, within the limits on eligible contributions the price of giving is zero. What is the rationale behind such an offer? This tax incentive structure well illustrates the politics of such provisions. While eager to have more revenues available to them, cultural organizations are wary of the creation of such new institutional funding mechanisms because they fear that they will be placed into direct fund-raising competition with the new intermediary. The proposal for a 100 percent tax credit is intended to remove this fear and to remove political opposition. But a 100 percent tax credit can be an expensive proposition for the state, so it is accompanied by ceilings on its use.

While the 100 percent tax credit remains in place in Oregon, it has had a more difficult time in other states. When it was originally proposed, the Wisconsin Artistic Endowment Foundation was to be the beneficiary of a 100 percent state income tax credit up to a limit of \$50 on individual contributions and \$500 on corporate contributions.<sup>64</sup> When the Department of Revenue said that it would lose a lot of revenue because of this provision, the legislature knocked the credit down to 25 percent (though this was still considerably more generous than the tax incentive for giving to other charitable sectors, which was at the state's marginal income tax rate of 5 percent). Under pressure from the other charitable sectors as well as from state budgetary considerations, the governor ultimately took the effective credit down to the 5 percent level, and as a result the momentum for getting the Wisconsin Artistic Endowment Foundation up and running has evaporated.

At first glance, the funding model for the New Jersey Cultural Trust appears to be rather different. The State of New Jersey will match on a 1:1 basis private donations

<sup>63</sup> In subsequent years, the incentive for corporate donations was limited further.

<sup>64</sup> This account is based upon personal communication with George Tzougros, Executive Director of the Wisconsin Arts Board, 13 October 2004.

made directly to the Cultural Trust or private donations earmarked for the endowments of specific cultural groups. In either case the match is paid to the Cultural Trust and is used to build up its corpus. A floor of \$10 million per year in state match was established. Any organization that receives a donation of more than \$100,000 toward its endowment automatically qualifies for a grant of 20 percent of the match paid to the Cultural Trust. But remembering the similarity between matching grants and tax incentives, this funding model is the rough functional equivalent of a 50 percent tax credit against state income taxes. The question is the extent to which the potential donor actually perceives this program as essentially doubling his or her contribution. This depends a lot on how the program is marketed, though admittedly there must be a perceptual difference between writing the entire check (and knowing that a tax credit has made it possible/attractive for you to give more than otherwise would have been the case) and writing a check that is only half the value of what the Cultural Trust will receive.<sup>65</sup>

### 5.6. Romania

Legislation proposed in Romania has taken up a somewhat different issue in corporate giving. In 1992 the Parliament of Romania circulated a draft Law on Sponsorship for comment, as part of a process seeking to regularize the nonprofit sector and provide incentives for contributions to that sector. Although the legislation did not distinguish between sponsorship and philanthropy, the draft law proposed to distinguish between sponsorship undertaken anonymously from that undertaken with recognition provided to the sponsor.<sup>66</sup> The former would benefit from a more generous tax-incentive: both forms of sponsorship would be deductible from total income before calculating taxes owed, but in the case of anonymous sponsorships, an *additional* tax credit was proposed which would vary between 0.5 and 1 percent of the amount of the sponsorship, depending on characteristics of the sponsor. The principle behind the additional incentive was that purer motivation would be more highly rewarded, a result that reveals an underlying uncertainty and suspicion of tax-based incentives and how they might be captured to advance individual interests in a tax system in which they have not heretofore been a recognized component.

<sup>65</sup> There is one other issue pertaining to state tax law that deserves mention here. In the United States state tax payments are typically deductible from the taxpayer's income prior to calculating one's federal income taxes. The result is that a substantial portion of any tax incentive that is realized as part of one's state tax payments is clawed back by the federal government: taking advantage of a state income tax incentive lowers your state income tax, resulting in a decrease in the deduction you can take on your federal tax form for state income taxes; this raises your federal income taxes in the following year.

<sup>66</sup> The Parliament of Romania, draft, *Law on Sponsorship*, 1992(?): Article 5, ¶4; Article 6; and Article 7. One should be careful about attempting to divine too much intent from this proposal because of the confusion within the draft law itself among various types of "sponsorship". Apparently, the crafters of the law were trying to incorporate in the one term "sponsorship" three different types of financial transfers from both individual and corporate contributors – charitable contributions for ongoing, unspecified expenses; contributions intended to support a particular project; and sponsorship that was more narrowly intended to promote the name, trademark, or image of the sponsor.

### 5.7. Conclusion

Taken together, these examples demonstrate that many jurisdictions, whether national or sub-national, are looking at tax incentives that go well beyond what might be considered “natural” incentives linked directly to the marginal tax rates in force. They are all based on the conviction that tax incentives will actually work to increase the level of total contributions, but they also seem to be based on an assumption that a simple deduction will not prove to be a sufficient inducement. The question becomes how high does that inducement have to be to get the desired result and will that inducement ultimately prove too expensive?

## 6. Summary and concluding remarks

The point of concluding this chapter with these international examples is at once simple and complex. The general yet only ambiguously-supported belief in the effectiveness of American-style tax incentives has led to a wide proliferation of such mechanisms elsewhere. Often they are modeled on the American deduction, but not always. Some countries have opted for a credit rather than a deduction; some have implemented more restrictive maxima or minima; some have included superdeductions; and some have put in place institutional structures to monitor the system and prevent what they perceive as abuse or the potential for expensive over-use. But they are all linked in their desire to provide a certain type of incentive, an incentive for increased individual or corporate support which may well be seen as a long-run way to ease pressure on government support.

In addition to continuing research on the economic effects of tax-based incentives to refine further our understanding of those effects, there is also a need for comparative research on the forms of incentives and their relationship to various national contexts. Why have particular forms of incentives arisen in certain countries but not in others? What difference does variation in various parameters make in the operation of these tax laws? What are the comparative politics of tax-based incentives? And, finally, while researchers have much to learn from the broader research on tax-based incentives for nonprofit charitable organizations and their various sources of support, sector-based research would also be of interest. Feld, O’Hare and Schuster (1983) pointed out that changes in tax law will not affect all charitable sectors equally because the demographics of those who choose to fund the arts and culture are different from those who choose to fund other sectors. Is there an appropriate set of tax-based incentives for the arts and culture, one that would be justifiably different from the set available to another sector? Several of the examples cited above seem to be based on that premise. But the central analytic question remains: To what extent have they realized their promise?



## References

- AAFRC Trust for Philanthropy (2002). *Giving USA: The Annual Report on Philanthropy for the Year 2001*. AAFRC, Indianapolis.
- Alvarez, J.L. (1988). "Taxation and heritage development". In: Council of Europe, *New Ways of Funding the Restoration of the Architectural Heritage*, Architectural Heritage Reports and Studies, No. 13, Report of the Messina Colloquy. Council of Europe, Strasbourg, pp. 98–107.
- Banfield, E.C. (1984). *The Democratic Muse: Visual Arts and the Public Interest*. Basic Books, New York.
- Barrett, K.S., McGuirk, A. (1992). "Charitable giving and federal income tax policy: Additional evidence based on panel-data elasticity estimates: The case of the Student's-*t* distribution". Paper presented at 1992 annual conference of the Association for Research on Nonprofit Organizations and Voluntary Action, New Haven.
- Baumol, W.J., Bowen, W.G. (1966). *Performing Arts: The Economic Dilemma*. Twentieth Century Fund, New York.
- Boorsma, P.B., van Hemel, A., van der Wielen, N. (Eds.) (1998). *Privatization and Culture. Experiences in the Arts, Heritage and Cultural Industries in Europe*, CIRCLE Publication No. 10. Kluwer Academic, Dordrecht.
- Borgonovi, F., O'Hare, M. (n.d.). "New world wine in old world bottles: Adapting US indirect support of the arts to the Italian context". Unpublished manuscript, Goldman School of Public Policy, University of California Berkeley, Berkeley.
- Bouuaert, I.C. (1975). "Taxation of cultural foundations and of patronage of the arts in the member states of the European economic community". Report XII/670/75-EN, Commission of the European Communities, Ghent.
- Bouuaert, I.C. (1977). "Fiscal problems of cultural workers in the states of the European economic community". Report XII/1039/77-EN, Commission of the European Communities, Ghent.
- Bouuaert, I.C. (1979). "Tax problems of historic houses in the states of the European economic community". Report XII/728/79-EN, Commission of the European Communities, Ghent.
- Brenson, M. (2001). *Visionaries and Outcasts: The NEA, Congress, and the Place of the Visual Artist in America*. The New Press, New York.
- Brody, E. (Ed.) (2002). *Property-Tax Exemption for Charities*. Urban Institute Press, Washington, DC.
- Brody, E., Cordes, J.J. (1999). "Tax treatment of nonprofit organizations: A two-edged sword". In: Boris, E.T., Steuerle, C.E. (Eds.), *Nonprofits and Government: Collaboration and Conflict*. Urban Institute Press, Washington, DC.
- Brooks, A.C. (1999). "Do public subsidies leverage private philanthropy for the arts? Empirical evidence on symphony orchestras". *Nonprofit and Voluntary Sector Quarterly* 28, 32–45.
- Brooks, A.C. (2000a). "Is there a dark side to government support for nonprofits?" *Public Administration Review* 60, 211–218.
- Brooks, A.C. (2000b). "Public subsidies and charitable giving: Crowding out, crowding in, or both?" *Journal of Policy Analysis and Management* 10, 451–464.
- Brooks, A.C. (2003a). "In search of true public arts support". Unpublished manuscript, The Maxwell School of Syracuse University.
- Brooks, A.C. (2003b). "Taxes, subsidies, and listeners like you: Public policy and contributions to public radio". *Public Administration Review* 63, 554–561.
- Chua, V.C.H., Wong, C.M. (1999). "Tax incentives, individual characteristics and charitable giving in Singapore". *International Journal of Social Economics* 26, 1492–1504.
- Clotfelter, C.T. (1985). *Federal Tax Policy and Charitable Giving*. University of Chicago Press, Chicago.
- Coffey, S. (1998). "The social and fiscal status of artists in Ireland". In: *Profession Artist: Report on the Social and Fiscal Status of Artists – Part Three: Austria, Denmark, Ireland, Luxembourg*. European Commission, Brussels. Published in the framework of the Community Action Plan in the field of Art Training and Education, DGXXII Education, Training and Youth, Direction A, Unit 3.
- Commission on Private Philanthropy and Public Needs (1977). *Research Papers, Volumes I–V*. US Department of the Treasury, Washington, DC.

- Cordes, J. (1999). "The cost of giving: How do changes in tax deductions affect charitable contributions?" *Emerging Issues in Philanthropy*, newsletter of the Urban Institute and the Hauser Center for Nonprofit Organizations, Washington, DC.
- Cordes, J. (2004). "The partially subsidized muse: Estimating the value and incidence of public support received by nonprofit arts organizations". In: Schwartz, A.E. (Ed.), *City Taxes, City Spending: Essays in Honor of Dick Netzer*. Edward Elgar, Cheltenham, pp. 198–243.
- Coslovich, G. (2002). "Visual arts inquiry calls for \$15 m and tax incentives". *The Age*, <http://www.theage.com.au/articles/2002/09/06/1031115938517.html>.
- Council for Museums, Archives and Libraries (2002). "Acceptance in Lieu. Report 2000/02". *Re:source*.
- Eckel, C.C., Grossman, P.J. (2003). "Rebate versus matching: Does how we subsidize charitable contributions matter?" *Journal of Public Economics* 87, 681–701.
- Einarsson, A. (2004). "The economic contribution of SMEs within the cultural sector of a small society in a global context". Unpublished paper presented at the Third International Conference on Cultural Policy Research, Montreal, Canada.
- Facchina, B., Showell, E., Stone, J.E. (1993). *Privileges & Exemptions Enjoyed by Nonprofit Organizations: A Catalog and Some Thoughts on Nonprofit Policymaking*, Topics in Philanthropy 3. The Program on Philanthropy and the Law, New York University School of Law, New York.
- Feld, A., O'Hare, M., Schuster, J.M.D. (1983). *Patrons Despite Themselves: Taxpayers and Arts Policy*, A Twentieth Century Fund Report. New York University Press, New York.
- Feldstein, M. (1975a). "The income tax and charitable contributions: Part 1 – Aggregate and distributional effects". *National Tax Journal* 28, 81–100.
- Feldstein, M. (1975b). "The income tax and charitable contributions: Part 2 – The impact on religious, educational, and other organizations". *National Tax Journal* 28, 209–226.
- Fogelman, V.M. (1990). "A capital tax system to preserve America's heritage: A proposal based on the British national heritage capital tax system". *Vanderbilt Journal of Transnational Law* 23, 1–63.
- Frey, B.S. (2000). *Arts & Economics: Analysis & Cultural Policy*. Springer, Berlin.
- Frey, B.S., Jegen, R. (2001). "Motivation crowding theory: A survey of empirical evidence". *Journal of Economic Surveys* 15, 589–611.
- Frey, B.S., Pommerehne, W.W. (1989). *Muses & Markets: Explorations in the Economics of the Arts*. Basil Blackwell, Oxford.
- Fullerton, D. (1991). "Tax policy toward art museums". In: Feldstein, M. (Ed.), *The Economics of Art Museums*. University of Chicago Press, Chicago, pp. 195–235.
- Galligan, A.M. (1993). "The politicization of peer-review panels at the NEA". In: Balfé, J.H. (Ed.), *Paying the Piper: Causes and Consequences of Arts Patronage*. University of Illinois Press, Urbana.
- Hawkins, A. (1988). "Income taxes and museums in America". *Apollo* 128, 340–344.
- Hightower, C. (1977). "A report on the arts". In: *Commission on Private Philanthropy and Public Needs, Research Papers, Volume II – Philanthropic Fields of Interest*. US Department of the Treasury, Washington, DC, pp. 713–734.
- Hood, R.S., Martin, S.A., Osbert, L.S. (1977). "Economic determinants of individual charitable donations in Canada". *Canadian Journal of Economics* 10, 653–669.
- Hughes, P.N., Luksetich, W.A. (1999). "The relationship among funding sources for art and history museums". *Nonprofit Management & Leadership* 10, 21–37.
- Independent Commission (1990). *A Report to Congress on The National Endowment for the Arts*. The Independent Commission, Washington, DC.
- Inkei, P. (2001). *Tax Incentives for Private Support to Culture*, Council of Europe Document DGIV/CULT/MOSAIC(2001)13 (Regional Observatory on Financing Culture in East-Central Europe, Budapest).
- International Center for Not-for-Profit Law (2002). *Survey of Tax Laws Affecting Non-Governmental Organizations in Central and Eastern Europe*. International Center for Not-for-Profit Law, Washington, DC.
- International Federation of Arts Councils and Culture Agencies (2003). "Encouraging arts philanthropy: Selected resources". In: IFACCA Research Reports 2002–2003. IFACCA, Sydney, pp. 67–75.
- Jencks, C. (1987). "Who gives to what?" In: Powell, W. (Ed.), *The Nonprofit Sector: A Research Handbook*, Yale University Press, New Haven, pp. 321–339.

- Jones, A., Posnett, J. (1990.) "The impact of tax deductibility on charitable giving by covenant in the UK". Draft, Department of Economics and Related Studies, University of York, York.
- King, K., Blaug, M. (1976). "Does the Arts Council know what it is doing". In: Blaug, M. (Ed.), *The Economics of the Arts*. Westview Press, Boulder, pp. 101–125.
- Kingma, B.R. (1996). "Balancing the mix of revenues from profits, donations, and governments". *New Directions for Philanthropic Fundraising* 12, 143–156.
- Kingma, B.R., McClelland, R. (1995). "Public radio stations are really, really not public goods: Charitable contributions and impure altruism". *Annals of Public and Cooperative Economics* 66, 65–76.
- McAndrew, C., O'Hagan, J. (2000). "Export restrictions, tax incentives and the national artistic patrimony". *Cultural Trends* 37, 43–63.
- McDaniel, P.R. (1977). "Study of federal matching grants for charitable contributions". In: Commission on Private Philanthropy and Public Needs (1977), *Research Papers, Volume IV – Taxes*. US Department of the Treasury, Washington, DC, pp. 2417–2534.
- Martell, C. (2004). "Dedicated funding for arts, culture and science". *Public Finance and Management* 4, 50–74.
- Meyer, K. (1979). *The Art Museum: Power, Money, Ethics*. Morrow, New York.
- Mikesell, J.L. (2002). "Tax expenditure budgets, budget policy, and tax policy: Confusion in the states". *Public Budgeting & Finance* 22, 34–51.
- Ministère de la Culture et de la Communication (2004). *Un nouvel élan pour le mécénat culturel: Loi du 1er août 2003 relative au mécénat, aux associations et aux fondations*. Ministère de la Culture et de la Communication, Paris.
- Ministry of Finance (2002). "Budget statement 2002: A budget for a different world". Government of Singapore, source: <http://www.mof.gov.sg>.
- Morel, C. (2005). "Will businesses ever become legitimate partners in the financing of the arts in France?" *International Journal of Cultural Policy* 11, 199–213.
- Munnely, A. (Ed.) (1986). *Tax and the Artist: Survey of European Tax Codes as they affect Creative and Interpretative Artists*. The Arts Council, Dublin.
- Navarro, P. (1988). "The income elasticity of corporate contributions". *Quarterly Review of Economics and Business* 28, 66–75.
- Netzer, D. (1978). *The Subsidized Muse: Public Support for the Arts in the United States*. Cambridge University Press, New York.
- Netzer, D. (1992). "Arts and culture". In: Clotfelter, C. (Ed.), *Who Benefits from the Nonprofit Sector? The University of Chicago Press*, Chicago, pp. 174–206.
- Netzer, D. (1997). "Tax policies and cultural heritage". In: Hutter, M., Rizzo, I. (Eds.), *Economic Perspectives on Cultural Heritage*. St. Martin's Press, New York, pp. 170–182.
- O'Hagan, J. (1998). *The State and the Arts: An Analysis of Key Economic Policy Issues in Europe and the United States*. Edward Elgar, Cheltenham.
- O'Hagan, J., McAndrew, C. (2001). "Restricting international trade in the national artistic patrimony: Economic rationale and policy instruments". *International Journal of Cultural Property* 10, 32–54.
- Pacque, K.-H. (1982). "The efficiency of public support to private charity: An econometric analysis of the income tax treatment of charitable contributions in the Federal Republic of Germany". Working Paper No. 152, *Institute für Weltwirtschaft an der Universität Kiel*, Kiel.
- Penhallurick, L. (1998). "Tax breaks for philanthropy". *Smarts* 13, 6–7.
- Rafool, M. (2004). "State arts endowments". Unpublished document, National Conference of State Legislatures, Denver, Colorado.
- Rockefeller Panel Report (1965). *The Performing Arts: Problems and Prospects*. McGraw-Hill, New York.
- Rouet, F. (1999). *VAT and Book Policy: Impacts and Issues*, Policy Note No. 1. Cultural Policies Research and Development Unit, Council of Europe, Strasbourg.
- Russell, F. (2001). *Individual Income Tax Provisions in the States*. State of Wisconsin, Legislative Fiscal Bureau, Madison.
- Schuster, J.M.D. (1985a). "Tax incentives for charitable donations: Deeds of covenant and charitable contribution deductions". *University of San Francisco Law Review* 19, 329–376.

- Schuster, J.M.D. (1985b). "The non-fungibility of arts funding: Perspectives on corporation and foundation support". In: *The Proceedings from Arts Research Seminar #4, The Arts: Corporations and Foundations*. The Canada Council, Ottawa.
- Schuster, J.M.D. (1986). "Tax incentives as arts policy in Western Europe". In: DiMaggio, P. (Ed.), *Nonprofit Enterprise in the Arts: Studies in Mission and Constraint*. Oxford University Press, New York, pp. 320–360.
- Schuster, J.M.D. (1987). "Issues in supporting the arts through tax incentives". *Journal of Arts Management and Law* 16, 31–50.
- Schuster, J.M.D. (1989a). "The British deed of covenant as a tax incentive for charitable donations: Vestige or viable alternative?" In: James, E. (Ed.), *The Nonprofit Sector in International Perspective: Studies in Comparative Culture & Policy*. Oxford University Press, New York, pp. 267–285.
- Schuster, J.M.D. (1989b). "Government leverage of private support: Matching grants and the problem with 'new' money". In: Wyszomirski, M., Clubb, P. (Eds.), *The Cost of Culture: Patterns and Prospects of Private Arts Patronage*. American Council for the Arts, New York, pp. 63–97.
- Schuster, J.M. (1994). "Funding the arts and culture through dedicated state lotteries – Part II: Opening the way for alternative decision making and funding structures". *European Journal of Cultural Policy* 1, 329–354.
- Schuster, J.M. (1996). "Questions to ask of a cultural policy: Who should pay? Who should decide?" *Culture and Policy* 7, 115–139.
- Schuster, J.M. (1999). "The other side of the subsidized muse: Indirect aid revisited". *Journal of Cultural Economics* 23, 51–70.
- Schuster, J.M. (2001). "Who should pay (for the arts and culture)? Who should decide? And what difference should it make?" In: Rothfield, L. (Ed.), *Unsettling "Sensation": Arts Policy Lessons from the Brooklyn Museum of Art Controversy*. Rutgers University Press, Piscataway, pp. 72–89.
- Schuster, J.M., de Monchaux, J., Riley II, C. (1997). *Preserving the Built Heritage: Tools for Implementation*. University Press of New England, Hanover.
- Schuster, J.M., Karraker, D., Bonaiuto, S., Grogan, C., Rothfield, L., Smith, S.R. (2003). *Mapping State Cultural Policy: The State of Washington*. Cultural Policy Center, The University of Chicago, Chicago.
- Simmonds, J. (2001). *Art and Taxation: A Guide*. Institute of Art and Law, Leicester.
- Simon, J.G. (1987). "The tax treatment of nonprofit organizations: A review of federal and state policies". In: Powell, W.W. (Ed.), *The Nonprofit Sector: A Research Handbook*. Yale University Press, New Haven, pp. 67–98.
- Staines, J. (2004). *Tax and Social Security: A Basic Guide for Artists and Cultural Operators in Europe*. Informal European Theatre Meetings, Brussels.
- Steinberg, R. (1990). "Taxes and giving: New findings". *Voluntas* 1, 61–79.
- Steinberg, R. (1991). "Does government spending crowd out donations? Interpreting the evidence". *Annals of Public and Cooperative Economics* 62, 591–618.
- Steinberg, R. (2003). "Economic theories of nonprofit organizations: An evaluation". In: Anheier, H., Ben-Ner, A. (Eds.), *The Study of the Nonprofit Enterprise: Theories and Approaches*. Kluwer Academic/Plenum Publishers, New York, pp. 277–309.
- Steinberg, R., Bilodeau, M. (1999). *Should Nonprofit Organizations Pay Sales and Property Taxes?* National Council of Nonprofit Associations, Washington, DC.
- Surey, S. (1973). *Pathways to Tax Reform: The Concept of Tax Expenditures*. Harvard University Press, Cambridge, MA.
- Taylor, D. (Ed.) (2000). *Tax Exemptions – 2000: A Study of Tax Exemptions, Exclusions, Deductions, Deferrals, Differential Rates and Credits for Major State and Local Taxes in Washington*. Research Division, Washington State Department of Revenue, Olympia.
- Tiehen, L. (2001). "Tax policy and charitable contributions of money". *National Tax Journal* 54, 707–723.
- Toffler, A. (1965). *The Culture Consumers: Art and Affluence in America*. Penguin Books, Baltimore.
- Trupiano, G. (2002). "Public subsidies and tax incentives for culture in Italy". Unpublished paper presented at the 12th International Conference of the Association for Cultural Economics, Rotterdam.

- Vajda, Á, Kuti, É. (2002). "Citizens' votes for nonprofit activities in Hungary". Working Paper, Nonprofit Sector Research Fund, The Aspen Institute, Aspen.
- Vladeck, B.C. (1976). "Why non-profits go broke". *Public Interest* 42, 86–101.
- Wallace, J.A., Fisher, R.W. (1977). "The charitable deduction under section 170 of the Internal Revenue Code". In: Commission on Private Philanthropy and Public Needs, Research Papers, Volume IV – Taxes. US Department of the Treasury, Washington, DC, pp. 2131–2161.
- Weil, S. (1991). "Tax policy and private giving". In: Benedict, S. (Ed.), *Public Money and the Muse: Essays on Government Funding for the Arts*. W.W. Norton, New York.
- Weisbrod, B.A. (1991). "Tax policy toward non-profit organisations: An eleven country survey". *Voluntas* 2, 3–25.
- Wong, C.M., Chua, V.C.H., Vasoo, S. (1998). "Contributions to charitable organizations in a developing country: The case of Singapore". *International Journal of Social Economics* 25, 25–42.
- Wyszomirski, M.J. (1990). "The art and politics of peer review". *Vantage Point*, American Council for the Arts, insert in *Horizon*, spring 1990, pp. 12–13.
- Yarmolinsky, A. (2000). "The charitable deduction: subsidy or limitation?" *Nonprofit and Voluntary Sector Quarterly* 29, 173–178.

## PHILANTHROPY

STANLEY N. KATZ

*Princeton University, Princeton, NJ, USA*

### Contents

Abstract	1300
Keywords	1300
1. Philanthropy and the arts in the United States	1301
1.1. Origins: From charity to philanthropy	1301
1.2. The emergence of the philanthropic foundation	1303
1.3. The development of federal government “philanthropy”	1305
1.4. The impact of foundation philanthropy on art and culture	1306
1.5. Corporate philanthropy and art and culture	1307
1.6. Government cultural philanthropy: The endowments	1307
1.7. Philanthropy in the early 21st century	1309
1.8. Philanthropy, economy and government policy	1314
1.9. The American affinity for philanthropy	1315
2. Philanthropy outside the United States	1316
3. Conclusion	1319
References	1319

## **Abstract**

Charity on the English and early American model must be distinguished from philanthropy as it developed in America at the turn of the twentieth century. Philanthropy aims at the systemic eradication of social ills rather than, as does charity, at their amelioration. The general purpose philanthropic foundation became the standard vehicle for effecting this purpose, and made significant contributions to a variety of fields until the Great Depression, though funding for the arts was fairly limited. Much of the role of foundations was assumed by the federal government as it expanded in power and scope after World War II, often adopting the foundation *modus operandi* as its own. Partly as a result, foundations began to support the arts and culture, and when corporate philanthropy emerged in the 1950s it was also often aimed at the arts and culture. The National Endowments for the Arts and Humanities have been responsible since 1965 for the central government contribution to these fields. Recent changes in the foundation sector include the growth in the number of new, very well endowed foundations; the emergence of conservative foundations; a contraction in the scope of foundation funding, including fewer and smaller grants to the arts and culture; and the rising number of community and family foundations. American philanthropy is undoubtedly responsive to government policy and economic cycles, but there remains despite vicissitudes a unique affinity for philanthropy in the United States, perhaps explained in part by the relative weakness of the American state. In Europe, a different historical tradition and legal framework has given rise to different forms of support for the arts and culture, with a greater role for the state. Recent trends, however, suggest that non-American philanthropy is coming to resemble philanthropy in the United States.

## **Keywords**

philanthropy, arts, culture, foundations

*JEL classification:* A13, D31, D64, L31, L33, L82, Z11

## 1. Philanthropy and the arts in the United States

### 1.1. *Origins: From charity to philanthropy*

In English, and in the United States, “philanthropy” remains both a technical and a general term, and its technical meaning is imperfectly understood both by most practitioners and by the general public. The term is probably best understood in its historical context, and especially in opposition to the concept of “charity”, the duty to take care of the poor, the ill, and the otherwise unfortunate in society.

Charity in the American colonies was influenced in some respects by the practice of charity in England. For instance, England’s Poor Relief Act of 1601 had placed certain charitable responsibilities in the hands of municipal authorities; some early American charity took the same form, especially in New England, though municipalities would usually farm out the actual care of the needy to whoever could do it most cheaply. But at the same time, the Church of England (and the other Christian denominations), both played a charitable role itself and imposed upon its better-off members the duty to share their wealth with the unfortunate of the faith; so in some regions of America – like Anglican Virginia – parishes and their parishioners were charged with the care of the poor and needy.<sup>1</sup> But while Elizabeth’s Statute of Charitable Uses of 1601 (a statute now under revision in the United Kingdom) had begun to formalize the legal status of charity and especially of charitable trusts in England by delimiting what qualified as charitable activities, the instability and variability of colonial law provided an at best unfavorable environment for charity in America [Tomlins and Mann (2001)]. After the Revolutionary War that environment, especially outside New England turned openly hostile to charity for several decades, perhaps influenced by Jefferson’s warnings against “un-republican” organizations. In Virginia, the Anglican Church was disestablished and its assets seized, and Pennsylvania annulled the Statute of Charitable Uses.

But American impulses towards voluntarism and charity would not be thwarted, and the controversies that resulted required the intervention of the Supreme Court, which ruled, though not wholly unambiguously, in favor of those impulses in 1819 and 1844 (in the *Dartmouth College v. Woodward* and *Vidal et al. v. Philadelphia* cases). Still, the Court left enough unsaid for two divergent legal approaches to charity to develop. In New England, the law not only allowed charitable activities but encouraged them through tax exemptions [Wright (1992)]; but most other states limited the sorts of activities could pass as charitable, and did little to promote them [Zollmann (1929)]. As a result, charitable activity tended to thrive in the Northeast and upper Midwest, and to do less well in other areas.

Alongside this more permissive legal environment, a concatenation of developments was preparing the way for the dramatic expansion of the general idea and the legal definition of charity in the very late nineteenth century, with the development of the modern

<sup>1</sup> On early American charity, see Trattner (1999) and Katz (1996); on early English charity, see Jordan (1959) and Owen (1965).



conception of philanthropy.<sup>2</sup> The professions were developing associations to uphold occupational standards; large benevolent associations were coming to be run by salaried administrators instead of volunteers; national-scale issues like slavery, immigration, and black migration northwards were nationalizing public life, a process enhanced by the increased ease of communication that followed the federal government's reorganization of the postal service and its control over interstate commerce; the United States Sanitary Commission's efforts on behalf of public health on Civil War battlefields demonstrated the efficacy of scientific expertise, and then discharged into the medical community a body of scientifically minded experts concerned with dealing with the causes of disease and suffering instead of their symptoms; research universities were beginning to thrive; and the establishment of the National Conference of Charities and Correction emblemized, more generally, the professionalization of social welfare and the shift from the alleviation of social problems to their systematic solution.<sup>3</sup>

At the same time, the great entrepreneurs of the Era of the Robber Barons were amassing fortunes of unheard of sizes. These capitalists were all Christians, and most of them believed fervently that their Protestant faith required them to steward their wealth, and to use some considerable part of it for the benefit of others (though Carnegie's "Wealth" of 1889 is a remarkably pragmatic and secular document). They had always been charitable – Rockefeller had tithed the first dollar he ever earned as a boy. The difficulty for those captains of industry interested in social problems was that they were amassing liquid wealth more quickly than they were able to distribute it charitably. The charity they were able to practice was necessarily retail, and both their sense of social obligation and the amounts of money at their disposal required that they develop a wholesale process for distributing their rapidly accumulating riches.

The first instinct of these innovators was to imagine how they could expand and be more efficient in the charitable work they were already engaged in. But what really set them apart was their realization that the challenge was not merely or even primarily technical, but conceptual. For charity addressed the needs of individuals and responded to specific needs, whether of income, health or education. Even if charity could meet such needs on a continuing basis, it would do nothing to extirpate the underlying problems. Giving alms would not address the causes of poverty, nor would the provision of shelter and medicine eradicate disease. These modern businessmen (though the Russell Sage Foundation, one of the earliest and most ground-breaking foundations was established by a businessman's widow) wished to apply their wealth to the discovery of the underlying causes of personal distress, and to the formulation of strategies to rid their country and the world of such systemic scourges [Hammack and Wheeler (1994)].

<sup>2</sup> While charitable activities had sometimes been denominated "philanthropy", that term was widely and non-specifically used to describe generous or benevolent behavior; in the antebellum United States, for instance, those who opposed the institution of slavery were frequently described as "philanthropic".

<sup>3</sup> For reviews of these developments, see Hall (2003), Lubove (1965), Bremner (1996) and Karl and Katz (1981).

This was the origin of the modern idea of philanthropy. It emerged from a general context in which confidence was building that science and technology could provide tools to address social problems in radically new and effective ways. And it drew on the intellectual and business experience of the great capitalists, who had after all made their fortunes by innovating organizational strategies (especially the vertical integration of related extractive and manufacturing industries) and by employing the intellectual resources of the second scientific revolution (which was being so successfully managed by the emerging system of research universities). The insight was that these new techniques could be deployed to solve social, economic and medical problems just as effectively as they had served Standard Oil or United States Steel.

### 1.2. *The emergence of the philanthropic foundation*

The institutional embodiment of this philanthropy was the private philanthropic foundation. But the legal and political difficulties of establishing foundations were considerable. The rule against perpetuities militated against trusts (of which foundations were a type) enduring indefinitely, with the result that their capital never vested to their beneficiaries [Friedman (2000)]. And as Katz, Sullivan and Beach (1985) has shown, the laws of New York State, where much Robber Baron wealth was concentrated, were unfriendly to philanthropic foundations. Nonetheless, by the early 1900s several large foundations or foundation-like entities had successfully been established, and while Rockefeller's endowment of the Rockefeller Foundation with \$100 million and wide discretion to promote "the well-being of humankind throughout the world" (well, "mankind" in the early days) prompted a public outcry reminiscent of the populist 1890s, the Foundation was duly chartered by the state of New York in 1913 [Johnson and Harr (1988)]. The chartering cemented legal approval for a new sort of trust, with an indefinite class of beneficiaries and a very general purpose. In short, they created what Americans now call the private philanthropic foundation.

There was a calculated reason for the generality of purpose of these institutions. English law had always looked askance at over-specificity of purpose in trust, on the grounds that narrowly defined trust instruments would inevitably become impossible to carry out as social conditions changed over time (though the doctrine of *cy pres* helped resolve some such difficulties). The philanthropists understood the need for timeliness and flexibility from their experience in industrial management, and they sought to build it into their philanthropic foundations by appointing governing boards – trustees – whose principal duty would be constantly to refine the particular and immediate purposes of the foundation, consistent with its general commitment to the development of research-based solutions to the most urgent problems of the country and the world. They would do this through proactive grantmaking based on the payout of some substantial portion of the annual income of the trust. Such flexibility was less important for those foundations established with time limits for spending down their assets trusts, such as the Rosenwald Fund. Julius Rosenwald, the Sears Roebuck mail order magnate, endowed

the fund with more than \$20 million in 1927, on condition that it expend the funds within twenty-five years of his death. In the event, it took the Fund only sixteen years.

The model of the general purpose foundation, then, was established by the time the United States entered World War I, and it flourished during the 1920s. A significant number of private philanthropic foundations were established between the end of the Great War and the onset of the Great Depression. They were devoted to a wide variety of purposes, but the majority focused in various ways on education, medicine and public health, and international affairs. Considerable investment was made in addressing the “American Dilemma”, the role of race in modern democracy, and in fact Gunnar Myrdal’s book *An American Dilemma: The Negro Problem and Modern Democracy* was commissioned by the Carnegie Corporation.<sup>4</sup> There was relatively little focus on economics and the economy, since the new foundations were constantly subjected to the criticism that they were thinly disguised efforts by the rich, having sequestered their wealth from taxation, to influence public (and especially economic) policy. But some indirect investments were made by funding and or establishing intermediary organizations like the National Bureau of Economic Research, the Brookings Institution and the Social Science Research Council, by supporting the private natural resources planning efforts of both Herbert Hoover and Franklin Delano Roosevelt. Although the Carnegie Corporation did engage in some arts grantmaking (mostly in arts education), in the 1920s and 1930s, by and large the private foundations had little interest in supporting the arts and cultural sector. And then the problems of the Great Depression dwarfed the resources of the foundations and further de-prioritized cultural philanthropy, and World War II saw much of the expert staff members recruited into the war effort.

Also in this period, another sort of foundation emerged that would prove to be extremely important as the century wore on – the community foundation. As Howard (1963) documents, the first such foundation – the Cleveland Foundation – was established in Ohio in 1914. It was based on the traditional premise that “charity begins at home”, though the community foundation was as scientifically philanthropic as it was charitable, both in its methods of amassing funds and strategic approach to giving. The community foundation enjoyed the same legal status as a private foundation, but its assets were donated by wealthy members of its community and were intended to benefit that community specifically. The community foundation enabled local donors to gain the tax advantages and some of the leverage of private foundations without the expense and responsibility of establishing their own private foundations. Community foundations had their own boards, populated with a mix of local notables, mostly businessmen, but also with other sorts of prominent and knowledgeable citizens. The community foundations funded a wide range of community activities, and doubtless had a broader impact on local life, especially in cultural institutions and activities, than private foundations, at least until fairly recently. At about the same time, there developed (also in Ohio) a related but more conventionally charitable institution with a local

<sup>4</sup> See Lagemann (1992).

focus – the Community Chest. This institution has since had several forms (and it is now called the United Way), but the basic idea was to provide workers with the ability to “check off” (permit the employer to deduct regularly from the employee’s paycheck) amounts of money that a central organization in the community would distribute to a selected group of charitable, that is non-profit, organizations providing services to the local community [see Crane (1992); Brilliant (1992)].

### *1.3. The development of federal government “philanthropy”*

As is well known, the government of the United States emerged from the Revolutionary period as a designedly and adamantly weak state. The fundamental premise of the Constitution of 1787 was that powers not specifically granted to the national (“federal”) government were reserved to the governments of the states, or to the people. This was partly a clear principle of finance, and the budget of the federal government remained miniscule until the coming of Civil War in the mid-nineteenth century forced increases in revenue collection and military-industrial expenditure. Direct taxation was low at all levels, and was not an important federal source of revenue until after the passage of the Sixteenth Amendment in 1913. The division of functions between the federal government and the states meant that the latter had virtually sole responsibility for the police powers of the state – the health, safety and welfare of the people. Since charity was, as we have seen, very much a local issue until the late nineteenth century, both charity and (relatedly) cultural activity were either outside of any governmental jurisdiction at all or under the jurisdiction of state governments. As Alexis de Tocqueville pointed out so eloquently in the 1830s, most local cultural activity was supported by volunteerism and charitable giving. This is of course still the case to a remarkable extent.

But, of course, the power of the federal government has grown enormously, driven both by the demands of waging two wars (even before the United States declared that it was perpetually at war), and by the dramatically enlarged social and economic agendas that began during the Great Depression, reached a high point during the presidency of Lyndon Johnson with “The Great Society”. Despite the best efforts of successive late twentieth century Republican and Democratic administrations, it is still with us. The growth of the federal state (measured in terms of the proportion of GNP expended upon it) and the enlargement of the jurisdiction of the national government have been accompanied, not unsurprisingly, by its involvement in what look very much like philanthropic activities. Two points are worth noting. First, much of the thinking behind Roosevelt’s New Deal was heavily influenced by philanthropically funded research; the foundations, in other words, helped ease the federal government into their realm [Critchlow (1985)]. Second, New Deal activism, at least, actually encouraged philanthropy both because its very progressive tax rates encouraged charitable contributions, and because government policy depended for its implementation on the private infrastructure that philanthropy succored and in part constituted.

The practice began with the federal government’s response to the demands of World War II, when it became for the first time the dominant funder of scientific research.

The first step was federally supported military research, such as the Manhattan Project to develop atomic weapons. Subsequently the federal government became the funder of first resort in science and medicine in general; crucial in this regard, as Brownlee (1996) has shown, was the establishment of an almost universal mass income tax in 1943. At the end of the war, for the first time government “foundations” were established to drive the rapidly growing private research establishment: the National Science Foundation was established in 1950, and while the National Institutes of Health had been around in some form since 1897, it got its current name in 1948 and substantial financial support in the 1950s. These public agencies were dependent upon annual federal appropriations, and legally were quite unlike private foundations. What the government took from the “real” foundations was an approximation of their grantmaking procedures and their sectoral research planning strategies. Huge numbers of tax dollars (by historical standards) were invested in this enterprise, and one of the interesting effects of government “philanthropy” was the self-imposed necessity of private foundations, even the largest, to reposition their funding strategies in relation to what the government was and – especially – was not doing. Private foundations increasingly became niche players in relation to the total universe of research funding.

#### *1.4. The impact of foundation philanthropy on art and culture*

As already noted, American philanthropy paid little attention to art and culture in its first few decades. But changes in federal tax laws in the 1930s and 1940s stimulated a second wave of large private foundation creation – the Ford Foundation and Lilly Endowment, for instance – and in the post-war era several of these institutions identified culture as a promising area for philanthropic investment. This probably resulted in part from the need to move away from investment areas coming to be dominated by public funding, but also because of the emergence of new trustees with new interests and new views of the appropriateness of certain types of foundation funding. The arts and culture sector was democratizing itself in an era of rapidly expanding audiences, repertoires, and artists, and this made it easier for private foundations, always sensitive to allegations of elitism, to support it. By the 1960s and 1970s, first the Rockefeller and then, especially, the Ford Foundation had begun significant programs of cultural philanthropy. In the 1980s and 1990s, newer large foundations such as the Pew Charitable Trusts and the Wallace Funds entered the field, and there are now a number of newer foundations sufficient to support the creation of an umbrella organization of arts and culture funders, Grantmakers in the Arts, which organizes annual conferences and publishes their proceedings, and also produces research on philanthropic support for the arts and culture [Renz and Atlas (2000)]. The sum total of foundation support for the arts and culture, however, is quite small when compared either to private charitable donations or to state and local cultural funding, and it has fallen off dramatically in recent years [Focke (2001)].

### *1.5. Corporate philanthropy and art and culture*

The record of corporate philanthropy, another post-war American innovation, is similar. United States business corporations, like business firms everywhere, are institutions that seek to amass profits for productive re-investment or to distribute to shareholders if they have them. Milton Friedman (1997) and many other economic thinkers have gone out of their way to make clear that distributions of profits for other purposes, especially for philanthropic purposes, are ill-advised and inappropriate. It was only in 1935 that the Revenue Act made corporate donations to philanthropy tax-deductible, and since an important legal decision in 1948 they have become common practice [Muirhead (1999)]. The dominant rationale for corporate philanthropy, as Himmelstein (1997) puts it, has of course been that certain sorts of gifts “do well by doing good” – they benefit not only the donees, but also the business interests of the donor corporations. As corporate philanthropy became more and more common in the 1960s and 1970s, it tended to be directed at the communities in which a firm’s plants were located and its employees lived. Firms also developed matching gift programs through which charitable gifts of employees to universities and other non-profits were matched dollar for dollar by the corporation, both encouraging employee charitable behavior, and giving substantial satisfaction to charitable employees. The sums involved, however, are typically not large proportionate to earnings. Over the past fifty years they have generally ranged between 1 and 2% of pretax earnings – though for large firms such sums are non-trivial contributions. In the late 1970s industrialists in the upper Midwest led a movement to encourage corporations to boost their philanthropic giving to 5% of pretax profits, but the effort failed. In fact, in the 1980s and 1990s general levels of corporate philanthropy fell, and grant-making became more and more closely tied to corporate marketing strategies, and thus managed by the marketing and public relations arms rather than by corporate contributions offices of the firms [Porter and Kramer (2002)].

Much corporate philanthropy has gone to the cultural sector, mainly because support of the arts has appeared an attractive, uncontroversial way for industries with severe public relations problems to present themselves in a welcome positive light to the public. It is not surprising that the tobacco products industry led the way, with Philip Morris taking the sectoral lead. The first years of the twenty-first century have seen an unexpected rise in corporate giving: it amounted, according to Raymond (2003) to about 12 billion dollars in 2002, an increase of 10.5% over 2001 (in which an increase over 2000 occurred). But this was perhaps an anomalous development – the 1990s saw a fairly substantial decline in corporate philanthropy [Chronicle of Philanthropy (1999a, 1999b)].

### *1.6. Government cultural philanthropy: The endowments*

The federal government’s approach to the arts and culture has never flowed from a formed notion of cultural policy.<sup>5</sup> It took Congress more than ten years to decide

<sup>5</sup> See further Netzer’s Chapter 35 in this volume.

whether to accept James Smithson's gift to establish the Smithsonian Institution in Washington; and thereafter federal appropriations helped sustain and expand its vast (and arguably haphazard) complex of museums and research activities, far beyond what its donated trust funds could have provided by themselves [Oehser (1969); Burleigh (2003)]. A similar process took place with the Andrew Mellon endowment that began the National Gallery of Art in the capital city. The federal government encouraged individual Americans to support the arts and culture by means of the tax code, but it did little until the 1960s to provide directly for the sector (except, perhaps, for the Works Progress Administration's incorporation of visual and performing arts into its programs). But, to the surprise of many observers, years of lobbying by arts and cultural organizations (and individuals) resulted in the creation of federal cultural grantmaking agencies in 1965. In that year, in a characteristically Great Society mood, Congress passed an act establishing a National Foundation for the Arts and Humanities, comprised of two separate Congressionally appropriated "foundations", the National Endowment for the Arts, and the National Endowment for the Humanities. The authorizing legislation justified the two unendowed "Endowments" by reference to the importance of culture in strengthening democratic society, but the actual appropriations in the first years were very small [Netzer (1980)]. After all, the United States had never had a federal ministry for culture, for the very good historical reason that the Constitution never contemplated the possibility that the federal government would have responsibility for cultural activities of any kind; such activities and institutions were clearly intended to be within the jurisdiction of the states and localities.

The idea of having some sort of coordinating council to oversee them endowments (picking up on the original legislative plan for a National Foundation for the Arts and Humanities) was never implemented. Each Endowment has a presidentially-appointed Chair, subject to confirmation by the Senate, and the Chairs of each Endowment have significant discretion in grantmaking, although by statute each must work with a presidentially-appointed National Council. But the humanities are institutionally separated from the arts, mainly because "culture", the more general term, is considered such a dangerous concept in the United States when it comes to state action. Very few of the Chairs have been independently distinguished and politically strong, and the National Councils have frequently not been selected for the professional distinction of their members. Politics has usually been the rationale for making such appointments.

Under the leadership of the arts activist Nancy Hanks, the NEA began to receive substantially increased appropriations during the Nixon administrations in the early 1970s, and after: NEH was granted \$2.9 million in 1966, and less than \$8.5 million in 1969 when Hanks took the helm, but the NEA received about \$100 million in 1977, when she died [National Endowment for the Arts (2004)]. The NEH started off garnering about \$6 million in 1966, and also received about \$100 million in 1977 [Katz (2001)]. Each endowment increased the range of its funding, the NEA providing support both to individual artists and to leading performing arts organizations, while the NEH supported both individual scholarship and individual cultural institutions such as independent research libraries and historical societies. In real dollars, the Endowments reached their



highest appropriations levels in the very early 1980s, but neither ever broke the \$200 million mark even in the mid-1990s, the period of their greatest appropriations in nominal dollars. Support on this scale did not match to the annual total investments of private philanthropic foundations and individual donors, and they were even dwarfed by public funding at the state and local levels.

Furthermore, during the George H.W. Bush administration the worst fears of those who had opposed federal funding of culture on the grounds that it would politicize cultural affairs were confirmed. The so-called Culture Wars prominently involved the attacks of political and religious conservatives on the funding of individual artists. Congressmen and Senators wanted to know why NEA had supported allegedly homophobic and anti-Christian art, and comparable (though more restrained) conflicts emerged in the humanities. The result, as Zeigler (1994) and Marquis (1995) have shown, was the elimination of NEA individual grants to artists, and in general a movement away from the production of “high” art and support of traditional arts institutions to the funding of more popular and local forms of art. Partly in response to these crises, both endowments have come to serve in part as pass-through funders to state arts (and humanities) councils, which regrant these funds, mostly to local organizations.

But by the mid-1990s, and especially after the massive cutbacks of 1996, more and more federal money that might have gone to the NEA and NEH was passed directly through to state councils. Only a few of the humanities councils were simultaneously funded by state governments, but almost all of the arts councils received enormously more money from their states than from NEA. This was in part a continuation of the historic American preference for state rather than federal cultural policy, but of course it rendered the councils subject both to annual variations in state tax revenues, and to the vagaries of state politics, or both. For instance, state funding for the California Arts Council dropped from \$32 million in 2000–2001 and from \$29 million in 2001–2002, to \$1 million in 2002–2003 [Winn (2003)]. The message here is clear. As we enter the 21st century, philanthropy remains a significant source of the external support of cultural institutions and activities – although it trails far behind state and local public funding. Like all U.S. non-profit institutions since the Reagan revolution, cultural institutions increasingly will have to sustain themselves by means of earned revenues.

### *1.7. Philanthropy in the early 21st century*

Returning now to philanthropy proper in the United States, it is interesting to observe how little major institutional change has occurred, although there have been profound changes in broader philanthropic behavior. The fact is that the major institutional players remain private philanthropic foundations and community foundations, each of which has to be seen in the context of changing federal policies and priorities, since the philanthropic sector has always been an inherently reactive sector.

The number of private philanthropic foundations continues to grow, and the most remarkable change is the creation of new foundations with exceptionally large endowments. In 2004, according to the [National Center for Charitable Statistics \(2004\)](#), the



United States boasted almost 2000 private foundations with assets of \$25 million or more. This is a function of the dramatic creation of individual wealth in new industries, especially information technology, coinciding with federal tax policies favorable to the creation and retention of wealth. The most dramatic example has to be the Bill and Melinda Gates Foundation, founded only in 2000, and recently enhanced by a gift of more than \$3 billion by its founders. This makes the Gates Foundation, by any measure of historical economic value, the wealthiest foundation in the history of the United States. It has been fascinating to watch the Foundation, originally run in an intimate and informal fashion by Bill Gates' father and one of his business associates, swiftly take on the organizational form and grantmaking styles of older large philanthropic foundations. The same has been true of the Open Society Institute established in 1993 by the hedge fund manager George Soros: for many years Soros tried to direct the complex of foundations he established around the world by himself, but the major Soros foundation now looks more like the Ford Foundation than the donor seems to have intended.

The iron law of what, following DiMaggio and Powell (1983), might be called "philanthropic institutional isomorphism" has taken over for both the Gates and Soros foundations. The new foundations, accelerating a process begun in the 1970s, are changing the structure and dynamics of foundation giving, making for a more distributed philanthropic foundation system. Many of the new large foundations are in the Midwest (for instance, the Ewing and Marion Kaufman Foundation in Kansas City) or, more likely, on the west coast (for instance, the Hewlett and Packard Foundations). But the new foundations also take on new sorts of objectives, sometimes (but not always) reflecting the interests of their donors: the Gates Foundation seems to reflect very strongly the emerging philanthropic interests of Bill and Melinda Gates, and the Packard Foundation reflects primarily the interests of the late Lucille Packard; interestingly, the Hewlett Foundation seems to have developed a quite independent agenda of its own.

Another significant development, this one also originated largely in the 1970s, has been the emergence of self-consciously politically conservative foundations. Foundations such as Olin, Smith Richardson and Bradley have constructed grantmaking programs aimed at stimulating, institutionalizing and disseminating conservative ideology, both by funding university programs (for instance, in law and economics), colleges (conservative alternative student newspapers and organizations), publications (Liberty Fund) and think tanks (Heritage Foundation, American Enterprise Institute, Cato Institute). They have organized their own cross-institutional trade association (Philanthropic Roundtable) to compete with the establishment Council on Foundations and Independent Sector. They have criticized the mainstream liberalism of the foundation world, causing both surprise and consternation among the older foundations, which apparently did not realize that they were liberal, a bit like Moliere's *bourgeois gentilhomme*, who did not know that he was speaking prose. What has particularly characterized the management of the conservative foundations has been the exceptional clarity and focus of their programs, and their result orientation [see [National Committee for Responsive Philanthropy \(1997, 2004\)](#)].

The older and more mainstream foundations would deny it, but it is likely that the most important recent change in foundation behavior has been a movement to imitate the management strategies of the conservative foundations. Over the past decade a series of new foundation management rationales have developed, and along with them a consultant industry (mainly drawing on business school models and experience). The new calls are for “effective”, “strategic”, or “venture” philanthropy, and each of these is essentially a business management point of view. Of course philanthropic foundations, from the very start, have always had investment strategies, and they have always desired to be effective, but their styles were more patient, relaxed and generous. After all, by the 1920s foundations were already developing “programs” or investment areas, and appointing program officers to manage these areas. They were frequently proactive, and worked closely with their chosen grantees. According to Kiger (2000), they also tended to work with limited numbers of grantees, those in whom they had the most confidence, and whose work they could monitor closely.

I could argue that the recently emerging foundation management strategies are only nominally different from the historic model of foundation management, but in fact it seems that something quite distinctive is taking place. The changes are more in the attitudes of managers than in the techniques they use. What is now evident in many of the large U.S. philanthropic foundations is an impatience with what is said to have been the chaotic and languid styles of post World War II foundations – too many programs, too little monitoring and evaluation of grantees and these programs, too much forbearance with grantee hijacking of previously agreed upon goals and strategies, and too little measurable success. Success is now defined in a much more matter of fact fashion, related to a new preoccupation with short-term impact on specific (and sometimes narrow) policy goals. There is less tolerance for ambiguity and uncertainty, and less patience with longer-term projects. Corollary to these attitudes is impatience with research, especially basic research, into underlying causes, the signature insight of the original philanthropists, and a loss of confidence in universities as the most effective sites for carrying on philanthropic programs.

Today many foundations are narrowing their focus, limiting the range of their programs, demanding short-term, measurable deliverables contracted up-front with grantees, and holding grantees strictly accountable for what they do and do not do [see Porter and Kramer (1999); Rimel (1999); Center for Effective Philanthropy (2002)]. At the “venture” end of the new philanthropy, the entrepreneurial techniques of venture capital are being applied [Letts, Ryan and Grossman (1997)]. Donees are analogized to start-up firms, donors partner with them, establishing specific and measurable benchmarks, and continuing their investments only if periodic goals are met. A few foundations have carried these notions quite far, some of them restricting themselves to a single program, and most significantly altering the range of their substantive programs in line with these principles. The result appears to be a retreat from research, from long-term programming, from goals that cannot be quantitatively measured, and from programs that do not promise hard-edged social/economic results.

Given the changing nature and direction of the new governing attitudes in foundation management, it is not surprising that arts and culture funding is less and less prominent on the agendas of the large foundations. As [Renz, Lawrence and Smith \(2004\)](#) show, foundation support for the humanities, for instance, slipped as a proportion of overall foundation giving between 1992 and 2002, from about 2.5% to just over 2%, though more foundations made humanities-related grants. There is no longer a single major philanthropic foundation that identifies the funding of mainstream domestic arts and cultural institutions and activities as a significant investment area other than the Andrew W. Mellon Foundation. Where arts funding continues, it is normally thought of as subordinate to larger policy goals.

Consider what has happened at the Rockefeller Foundation. In the period after World War II, Rockefeller emerged as one of the largest funders of art and culture, operating through a Division of Arts and Humanities. The Division was always relatively poor in resources, compared to health and other more central Foundation concerns, but it leveraged its resources and had an important impact in the cultural sector. The Division was always precariously situated, but it was reinvigorated by a humanist Foundation President for a period of time, and survived a public-policy-oriented President. But recent President, a development economist, attempted to transform the entire organization into a vehicle for assisting disadvantaged groups, especially in the Third World. The Arts and Humanities Division was therefore reinvented as the Division of Creativity and Culture, and much of its thrust redirected in line with the new overall direction of the Foundation. It is not clear that the Division will survive the newest President. For example, the [Rockefeller Foundation \(2003\)](#) describes the desired result of its Creativity and Culture endeavors as “to enhance the well-being of societies and better equip them to interact in a global and dynamic world”. Relatively little of the Division’s resources are currently invested in any kind of conventional support and development of the domestic cultural sector. But despite these changes, Rockefeller is still an arts funder, and that is something that can now be said about only a handful of large private foundations. For the moment, there have not been major changes in corporate philanthropy with respect to the cultural sector. Expenditures continue at relatively low levels, programming remains tightly correlated to marketing goals, and corporate engagement with the sector is mostly reactive.

What has changed most in the philanthropic sector, however, is the remarkable expansion of both community and family foundations. There used to be a relatively small number of community foundations, mainly based in the major urban areas. But now even smaller towns and cities have established local community foundations, as have large and small regions of the United States; the total number of community foundations was estimated by the [C.S. Mott Foundation \(2001\)](#) to be 600 in 2000, and accelerating. This means that a much larger and more diverse population of donors, ranging from those of comparatively modest means to the superrich, now contributes funds for regranteeing in their communities. A significant recent innovation in management of community foundations, as in the United Way organizations, is in “donor-advised” funds. The pitch to new donors has been that they do not have to defer to the allocational

decisions of the trustees of community foundations, but can play an active role in determining the use of the funds they have contributed. Frequently this has meant that donors set up funds to be used for quite specific purposes, in health, social service, and sometimes in culture. This has helped to stimulate a tremendous surge of investment, much of it centrally managed in particular communities.

A significant recent development is the growth in popularity of donor-advised funds (DAFs). These philanthropic instruments have been available since the 1930s, but have only become prominent since the early 1990s. Some of them – those managed by commercial investment firms such as Fidelity Funds – compete with community foundations. It is too soon to evaluate their impact on broader patterns of philanthropic investment, but it is already apparent that they are becoming major players. [Luck and Feurt \(2002\)](#) estimated that in 2000, just the assets of DAFs managed by community foundations (that is, excluding those managed by for-profit firms) totaled more than \$5 billion. While we do not have the data to assess the range and extent of their sectoral investments, it seems that they have not yet had a major interest in the cultural sector.

This cannot be said of the other major change on the philanthropic scene at the turn of the twenty-first century – the explosion of family foundations. The “family” foundation is of course not new. It could be said that the Rockefeller Foundation and the Carnegie Corporation were among the first family foundations, but of course these institutions evolved into general purpose foundations managed primarily (or in the case of Carnegie, entirely) by others than relatives of the donor. These days when we talk of family foundations we refer to those whose trustees are entirely (or primarily) members of the donor’s family – and of course by the donor him or herself. Family foundations, many of them extremely small, number in the tens of thousands [[Foundation Center \(2005\)](#)]. In many cases they are really little more than unstaffed or thinly-staffed emanations of individual philanthropists. Given the remarkable creation of individual wealth in the United States in the 1990s, it is not surprising that thousands of family foundations have been newly-established, but it is quite surprising, given the earlier history of foundations, that so many extremely large family foundations have been established. Some of them, as already mentioned, like the Bill and Melinda Gates Foundation, quickly developed into more highly organized, fully-staffed and rationalized general (or broadly-purposed) foundations, but others – such as Gordon Moore’s Gordon E. and Betty I. Moore Foundation – have remained highly personalized donor-driven foundations, at least so far.

We are still, in effect, in the first decade of what many are calling the “new” philanthropy in the United States, and we do not yet have enough data to be sure what the impact of new institutions and new strategies is. I have suggested elsewhere that there is less new about the new philanthropy than the claims of its enthusiasts suggest [[Katz and Stacey \(2004\)](#)]. And it is worth noting that these claims come largely from the rapidly growing new field of for-profit philanthropic consultants, the bulk of them in or recently spun-off from business consulting firms such as McKinsey and Bain. But even I am pretty sure that there is something more than novel rhetoric at play.

My guess, without a full analysis of the available data, is that the combination of rapid expansion of the community foundation sector, the burgeoning of donor-advised

funds (both in the non- and for-profit sectors) and the proliferation of family foundations is substantially changing the dimensions of philanthropic giving, the composition of donee communities, and the relations between donors and donees. It is always difficult to analyze change while it is taking place, but my sense is that we are currently living through a truly significant transitional era in the history of philanthropy in the United States. If nothing else, there is considerably more philanthropic capital in play, and there is an equally significant increase in the number and character of players. And all of this is taking place on a truly continental basis, changing the map of philanthropy from one that twenty-five years ago was focused on the Northeast, the upper Midwest, California and the Pacific Northwest.

### *1.8. Philanthropy, economy and government policy*

I have just suggested the changes within the philanthropy, but I would also like to stress the obvious fact that the philanthropic sector has always developed interactively with major state policies and with the general economy. Other essays in this volume will deal with tax policy, for instance, but I hope it is obvious that both overall rates of income taxation and tax policies specific to both foundations and individual giving have specific and fairly predictable impacts upon philanthropy.<sup>6</sup> The current proposals for elimination of inheritance taxes, for instance, have stimulated both general and scholarly debate over what their legislation might do to the propensity of the wealthy to transfer wealth to their families rather than to the larger society. Likewise, the much heralded explosion in the intergenerational transfer of wealth, if true, is bound to have a measurable impact on philanthropic investment, and probably upon the styles of investment as well [see [Schervish and Havens \(2002, 2003\)](#)].

Moreover, changes in federal social policy have an impact (often hard to measure) on philanthropy. After the election of Ronald Reagan in 1980, Republicans and conservatives demanded (and achieved) steep reductions in the federal social budget, while at the same time calling for the private charitable sector to take up the slack left by the retreat of the state. That did not happen to the hoped-for extent, but cutbacks in federal funding certainly did redirect both individual and institutional philanthropic investment, particularly in the service-providing sector [[Salamon and Abramson \(1982\)](#)]. And, as I have already mentioned, it forced non-profit institutions to try to generate earned income. These were among the forces that compelled museums to expand their gift shops, to rent out their facilities for social occasions, and to mount more spectacular blockbuster exhibitions in order to augment dwindling philanthropic and municipal financial support. To this extent the cultural sector behaved much like those service-providing non-profits that began to charge clients for what had previously been free services.

The conservative political movement in American national politics that took off with the Reagan administration was continued by George H.W. Bush and his son – and it

<sup>6</sup> See further Schuster's [Chapter 36](#) in this volume.

was the elder Bush who tried to stimulate private giving by describing in a 1988 speech the “1000 Points of Light” making positive impacts in their local communities. But it has seldom been noticed that the emphasis on private charitable and philanthropic involvement was emphasized nearly as much by Bill Clinton and his centrist Democratic administration. It was on Clinton’s watch that the first ever White House conference focusing on charitable giving was held [[Chronicle of Philanthropy \(1999a, 1999b\)](#)]. But it is not only social politics that drive the interest of the state in the philanthropic sector. Roughly every ten years since World War I the federal Congress has launched noisy investigations into the behavior of the philanthropic sector, and we are currently at the beginning of such an episode. In 2004 the Senate Finance Committee investigated a wide range of alleged foundation abuses, ranging from excessive expenditures for foundation staff, trustees and administration to suggestions for legislating requirements for board meetings and oversight of foundation affairs [[Grassley \(2004\)](#)]. This should remind Americans that what they have taken to calling the Independent Sector (usually called the Third Sector in Europe and the rest of the world) is not so completely independent.

Such government policies and activities – aside from the policies restricting the size, scope and funding of the Endowments for the Arts and Humanities and changes in Internal Revenue Service regulations governing charitable donations – do not ordinarily have a direct, specifiable impact on the cultural sector. Mostly, the impact (or lack of impact) of philanthropy on culture comes from alterations in foundation and individual donor behaviors and preferences, and these alterations are mostly endogenously generated. Culture is truly a part of the larger philanthropic and non-profit sector. The problem, if there is one, is that it is not a very large part of the sector. My hunch, though, is that there is much more philanthropic support than arts and cultural advocates, such as Americans for the Arts, admit. It is hidden in individual and donor-advised gifts that are not adequately publicly reported.

### *1.9. The American affinity for philanthropy*

Why do Americans give? A great deal has been written on motivations for philanthropic behavior in the United States, but general explanations are not of much help in understanding the instinct to give to a particular cause. The interesting issue is the distinction between charitable motivations, presumably similar across national and cultural borders, and the propensity to philanthropic activity [see [Ilchman, Katz and Queen \(1998\)](#)].

Philanthropy seems to me a distinctly American phenomenon, although of course it is now being appropriated and adapted in other (mostly post-industrial) national cultures. The usual explanation, tax avoidance, does not take us very far in understanding the phenomenon. The earliest twentieth-century philanthropic foundations were after all created before the federal government was permitted to tax individual income, following the passage of the Income Tax Amendment in 1913. There had been state income taxes since the Civil War, but the rates of state taxation were very low. Indeed, when John D. Rockefeller, Sr. was considering forming his foundation during the first decade of the

twentieth century, he instructed his lawyers to investigate state rates of income taxation. Their response was that it did not matter in which state the foundation was chartered since state taxation was so low as not to be a factor in the decision as to where to incorporate. Taxation became more of an issue by the 1930s, with the passage of federal gift and estate taxes. Both the Ford Foundation and the Lilly Endowment were created at that time explicitly as tax avoidance measures, but neither was funded for many years. There is some evidence that declines in marginal rates of taxation produce declines philanthropic giving, but the effect is not strong. Taxation is clearly an incentive to engage in tax-deductible behavior, but it is hardly the only motivation and I doubt that it is the major factor.

Much more important, I would speculate, is what I would call the American philanthropic instinct. This is distinctly not a specifically altruistic instinct, but rather a desire to contribute to the improvement in the conditions of human existence. The philanthropist may or may not share the belief of Andrew Carnegie and John D. Rockefeller, Sr. that his religious duty was to discover the underlying causes of illness and misery in order systematically to address them. There is no doubt that Protestant religion was a strong component in the belief systems of many of the early philanthropists, and religion may well still play a motivational role in American philanthropy. But it is unlikely to be the dominant motivation, and it plays no role whatsoever for many givers. The deeper and more pervasive instinct is the one first clearly demonstrated by the first generation of American philanthropists, and it relates to a generalized commitment to make the world a better place for all.

Clearly urge to make the world a better place is not unique to one country. But historical conditions have given it a peculiar salience and form of expression in the United States [Katz (2003)]. It doubtless found its origins in a people living in a country with a weak state tradition, one in which there was little history of the state (especially the national state) “doing good”. The early philanthropists feared the state, and worried especially that the state in America would move in the direction of European welfare states of the late nineteenth and early twentieth centuries. The answer was for individuals to take up the responsibility that Europeans assigned to the states, and to retain the wellsprings of reform in the private sector [Karl and Katz (1981)]. Times have changed and government in America has grown, but the significant emphasis of the conservative movement of the last generation in American politics shows just how tied the rhetoric of American philanthropy is to state-replacing activities. I doubt most American philanthropists are self-consciously state-replacers in any explicit sense, but their instinct to imagine and support the non-profit engines of human betterment derive from the original Carnegie–Rockefeller tradition.

## **2. Philanthropy outside the United States**

I want to conclude with a few remarks about the arts, culture and philanthropy elsewhere in the world. I have focused on the United States both because American philanthropy



is what I study, but also because private cultural philanthropy is vastly larger and more significant in the U.S. than elsewhere in the world. Though I do not think there are statistics to rely upon, it is hard to imagine that the sum total of private financial investment in culture in the U.S. is not larger than that in all other countries combined. Even if that assertion is too extreme, it is absolutely clear that no other single nation comes close to the level of private philanthropic investment in culture currently existing in the U.S. [Salamon and Sokolowski \(2004\)](#) have shown that the philanthropic giving of the United States outstrips that of most rich countries (even when donations to religious bodies are excluded), and the tradition of state support for the arts and culture in European countries has tended to depress private giving to the sector. Of course, what is interesting is the extent to which so many other nations are moving in the direction of the United States, frequently dramatically, as the size of their state sector declines and as they pursue self-conscious policies of privatization in their cultural sector.

History has made the difference, and here I will use Europe as the example. As art and culture developed from, say, the Renaissance, they took root in and were supported by patronage – the gifts and protection of the wealthy and powerful. These included the church and the politically powerful, but much early patronage was provided by individuals – it came from kings and princes and barons and cardinals and state ministers and merchants. There was of course a coincidence of private wealth and public power in pre-modern times, and at least from the fifteenth century the wealthy and powerful felt it both their duty and their prerogative to support artists, writers, craftsmen and other producers of culture and cultural institutions such as schools and cathedrals. The aim of the patrons was to glorify themselves and their families. But of course the maintenance and glorification of the Catholic Church was an equally important motivation. And this complex system of private patronage remained in place well into the nineteenth century.

But it had begun to be supplanted by the state and the state church from the sixteenth century forward. An obvious example was abolition of the English monasteries by Henry VIII, and the consequent necessity of the Crown to provide for institutions of religion and education [[Jordan \(1959\)](#)]. Such tendencies increased with the rise of parliaments in the Netherlands and England in the seventeenth century, and by the time of the French Revolution the emerging nation-state (with or without a king) had tended to take over church and crown responsibilities for cultural patronage. This was not entirely a self-conscious or explicit decision on the part of the state, but rather a reflection of democratizing nations espousing strong notions of state power to do publicly what had previously been done by the nobility, the crown and the church [see [Adam \(2004\)](#)]. The patron-state emerged over the course of the nineteenth century, assuming the dominant responsibility for education, state religion and arts institutions such as libraries and museums. In federal states, such as Germany, these patronage functions were frequently distributed regionally, according to local traditions of federalism, but they were ordinarily conceived of as state functions by Europeans [see [D'Angelo \(2000\)](#)].

As the twentieth century began, these forms of state support persisted, though they were less frequently conceptualized as patronage proper. They became institutionalized



in (state and federal) ministries of culture, the official expression of the responsibility of the state for cultural institutions and cultural policy. And both of these developments were quite different from the emerging tradition of the privatization of culture in America. If the state was to collect tax monies for culture, it asserted the right to set the policies according to which they would be given. And across Europe, but especially on the continent, the state was a generous funder of culture, sometimes with a light imposition of policy and other times with a heavy hand (the national socialist movements, for instance). On the other hand, American museums and other major cultural institutions were founded almost entirely by private wealth in the late nineteenth century, and that pattern continues – with some important exceptions – to this day.

But by the 1970s the attack on the welfare state in Europe, led by Margaret Thatcher in the United Kingdom, focused the attention of government on methods of shifting the burden off the backs of taxpayers. Privatization emerged as the obvious solution, along with a stiff dose of neo-classical economics and economic policy. In effect, Britain began a policy of Americanization of the political economy that still drives most change in state policy across the world, and certainly in Europe. The trend has apparently hit hard in the cultural sector there. To some extent non-tax alternatives, especially lotteries, have emerged to take up the space left by a receding patron state, but the larger move has been to encourage the privatization of cultural institutions, whether by their severing their ties with the state, or by finding additional non-state sources of revenue; the current debate in the United Kingdom and in most of the Commonwealth countries about the raising of student tuition fees is just one instance.

But, perhaps independently, something like a U.S. philanthropic sector has been emerging on the continent (and also in Latin America, Japan, and elsewhere). The movement has been aided in some countries by legislative action to provide tax incentives for individual charitable giving, but also by policies favorable to the establishment of non-governmental organizations (analogues to American non-profits) and to the creation of legal novelties, philanthropic foundations [*The Economist* (2004)]. There are now considerable numbers of both private and corporate foundations in Europe, though the financial levels at which they operate are modest by American standards. Still they are part of what appears to be an inexorable rollback of the state and an expansion of the use of private wealth for public purposes.

I am not sure that we can tell with any precision what the specific impacts on this move to private support have meant for the cultural sector in Europe. Certainly there have been cries of distress from universities, opera companies, museums and other cultural institutions that sound a good deal like similar noises being heard in the U.S. But it is not clear, at least to me, what degree of financial impact, sorts of structural modifications or alterations in cultural content are coming into existence. It is hard, though, to imagine that there will not be some important implications for the cultural sector.

### 3. Conclusion

What seems obvious, when all is said and done, is that the winds of change are blowing. Although in vastly different ways, the state is retreating from culture everywhere – although at the same time it remains everywhere a significant player on the cultural scene. In the United States we are in a transitional phase in which levels of real and potential philanthropy are rising dramatically, the state has abandoned even the modest policy claims it briefly asserted in the 1960s and 1970s, and yet the cultural sector feels that it is caught in a situation of severe shortage of funding. In Europe, the philanthropic sector expands as privatization proceeds apace, yet the state is not prepared to abandon its responsibility for cultural policy. We are probably in Act III of a five-act play, and it is much too soon to know how the plot will unfold.

### References

- Adam, T. (Ed.) (2004). *Philanthropy, Patronage and Civil Society: Experiences From Germany, Great Britain and the United States*. Indiana University Press, Bloomington.
- Bremner, R. (1996). *From the Depths: The Discovery of Poverty in the United States*. New York University Press, New York.
- Brilliant, E. (1992). *The United Way*. Columbia University Press, New York.
- Brownlee, W.E. (1996). *Federal Taxation in America: A Short History*. Cambridge University Press, Cambridge.
- Burleigh, N. (2003). *The Stranger and the Statesman: James Smithson, John Quincy Adams and the Making of America's Greatest Museum – The Smithsonian*. William Morrow, New York.
- Center for Effective Philanthropy (2002). *Toward a Common Language: Listening to Foundation CEOs and Other Experts Talk about Performance Measurement in Philanthropy*. The Center for Effective Philanthropy, Cambridge, MA.
- Chronicle of Philanthropy (1999a). "Philanthropy takes the bully pulpit: Charities hope that conference at White House stimulates giving". 21 October.
- Chronicle of Philanthropy (1999b). "Share of company profits given to charity drops from 2.4% to 1.1% in 11 years". 16 December.
- Crane, C. (1992). *Before United Way: Community Chest*. John Appleyard Agency, Pensacola.
- Critchlow, D. (1985). *The Brookings Institution, 1916–1952: Expertise and the Public Interest in a Democratic Society*. Northern Illinois University Press, DeKalb.
- C.S. Mott Foundation (2001). *Two Decades in the Field of Community Foundations*. The C.S. Mott Foundation, Flint.
- D'Angelo, M. (2000). *Cultural Policies in Europe: Regions and Cultural Decentralization*. Council of Europe Publishing, Strasbourg.
- DiMaggio, P., Powell, W. (1983). "The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields". *American Sociological Review* 48 (2), 147–160.
- Focke, A. (2001). "A snapshot: Foundation grants to arts and culture". *GIA Reader* 12/3.
- Foundation Center (2005). *Key Facts on Family Foundations*. The Foundation Center, New York.
- Friedman, M. (1997). "The business of business is business". *Advancing Philanthropy*, Winter, 30–33.
- Friedman, L. (2000). *American Law in the Twentieth Century*. Yale University Press, New Haven.
- Grassley, C. (2004). "Opening statement". U.S. Senate Committee on Finance Hearing, June 22, 2004, available at <http://www.finance.senate.gov/hearings/statements/062204cg.pdf>. Accessed March 21, 2005.
- Hall, P.D. (2003). "A historical overview of philanthropy, voluntary associations and non-profit organizations in the United States, 1600–2000". Unpublished paper, Hauser Center for Nonprofit Organizations.

- Hammack, D., Wheeler, S. (1994). *Social Science in the Making: Essays on the Russell Sage Foundation, 1907–1972*. Russell Sage Foundation, New York.
- Himmelstein, J. (1997). *Looking Good and Doing Good: Corporate Philanthropy and Corporate Power*. Indian University Press, Bloomington.
- Howard, N. (1963). *Trust For All Time: The Story of the Cleveland Foundation and the Community Trust Movement*. Cleveland Foundation, Cleveland.
- Ichman, W., Katz, S., Queen, E. (Eds.) (1998). *Philanthropy in the World's Traditions*. Indiana University Press, Bloomington.
- Johnson, P., Harr, J. (1988). *The Rockefeller Century*. Charles Scribner's Sons, New York.
- Jordan, W.K. (1959). *Philanthropy in England, 1480–1660: A Study of the Changing Pattern of English Social Aspirations*. Russell Sage, New York.
- Karl, B., Katz, S.N. (1981). "The American private philanthropic foundation and the public sphere, 1890–1930". *Minerva* 19 (2), 236–270.
- Katz, M. (1996). *In the Shadow of the Poorhouse: A Social History of Welfare in America*. Basic Books, New York.
- Katz, S. (2001). "Rethinking the humanities endowment". *Chronicle of Higher Education*, January 5.
- Katz, S. (2003). "Carnegie and Rockefeller: Their most lasting contribution". In: Halberstam, D. (Ed.), *Defining a Nation: Our America and the Sources of Its Strength*. National Geographic, Washington, DC, pp. 212–217.
- Katz, S., Stacey, S. (2004). "The philanthropist's new clothes?" Paper presented at the International Society for Third Sector Research conference, Toronto, Canada, July.
- Katz, S.N., Sullivan, B., Beach, C.P. (1985). "Legal change and legal autonomy: Charitable trusts in New York, 1777–1893". *Law and History Review* 3, 51–90.
- Kiger, J. (2000). *Philanthropic Foundations in the Twentieth Century*. Greenwood Press, Westport.
- Lagemann, E. (1992). *The Politics of Knowledge: The Carnegie Corporation, Philanthropy, and Public Policy*. University of Chicago Press, Chicago.
- Letts, C., Ryan, W., Grossman, A. (1997). "Virtuous capital: What foundations can learn from venture capitalists". *Harvard Business Review* 75, 36–44.
- Lubove, R. (1965). *The Professional Altruist: The Emergence of Social Work as a Career, 1880–1930*. Harvard University Press, Cambridge, MA.
- Luck, J., Feurt, S. (2002). *A Flexible and Growing Service to Donors: Donor-Advised Funds in Community Foundations*. Council on Foundations, Washington, DC.
- Marquis, A. (1995). *Art Lessons: Learning from the Rise and Fall of Public Arts Funding*. Basic Books, New York.
- Muirhead, S. (1999). *Corporate Contributions: The View from 50 Years*. Conference Board, New York.
- National Center for Charitable Statistics (2004). "Number of private foundations in the United States, 2004". Available at <http://nccsdataweb.urban.org/PubApps/profileDrillDown.php?state=US&rpt=PF>. Accessed March 20, 2005.
- National Committee for Responsive Philanthropy (1997). *Moving a Public Policy Agenda: The Strategic Philanthropy of Conservative Foundations*. National Committee for Responsive Philanthropy, Washington, DC.
- National Committee for Responsive Philanthropy (2004). *Axis of Ideology: Conservative Foundations and Public Policy*. National Committee for Responsive Philanthropy, Washington, DC.
- National Endowment for the Arts (2004). *2003 Annual Report*. The National Endowment for the Arts, Washington, DC.
- Netzer, D. (1980). *The Subsidized Muse: Public Support for the Arts in the United States*. Cambridge University Press, Cambridge.
- Oehser, P. (1969). *Sons of Science: The Story of the Smithsonian Institution*. Greenwood Press, Westport.
- Owen, D. (1965). *English Philanthropy 1660–1960*. Harvard University Press, Cambridge, MA.
- Porter, M., Kramer, M. (1999). "Philanthropy's new agenda: Creating value". *Harvard Business Review* 77, 121–130. 216.

- Porter, M., Kramer, M. (2002). "The competitive advantage of philanthropy". *Harvard Business Review* 80, 56–68. 133.
- Raymond, S. (2003). "Taking the measure of corporate giving". On philanthropy, available at [http://www.onphilanthropy.com/tre\\_comm/tc2003-10-10b.html](http://www.onphilanthropy.com/tre_comm/tc2003-10-10b.html). Accessed May 10, 2005.
- Renz, L., Atlas, C. (2000). *Arts Funding 2000: Funder Perspectives on Current and Future Trends*. Foundation Center in Cooperation with Grantmakers in the Arts, New York.
- Renz, L., Lawrence, S., Smith, J. (2004). *Foundation Funding for the Humanities: An Overview of Current and Historical Trends*. The Foundation Center, New York.
- Rimel, R. (1999). "Strategic philanthropy: Pew's approach to matching needs with resources". *Health Affairs* 18, 229–230.
- Rockefeller Foundation (2003). *Annual Report: 2002*. The Rockefeller Foundation, New York.
- Salamon, L., Abramson, A. (1982). *The Federal Budget and the Nonprofit Sector*. Urban Institute, Washington, DC.
- Salamon, L., Sokolowski, W. (Eds.) (2004). *Global Civil Society: Dimensions of the Nonprofit Sector*. Kumarian Press, Bloomfield.
- Schervish, P., Havens, J. (2002). "New findings on the patterns of wealth and philanthropy". Center on Wealth and Philanthropy working paper. Center on Wealth and Philanthropy, Boston.
- Schervish, P., Havens, J. (2003). "Gifts and bequests: Family or philanthropic organizations? and How do people leave bequests: Family or philanthropic organizations?" In: Munnell, A., Sunden, A. (Eds.), *Death and Dollars*. Brookings Press, Washington, DC.
- The Economist (2004). "Doing well and doing good". 29 July.
- Tomlins, C., Mann, B. (Eds.) (2001). *The Many Legalities of Early America*. University of North Carolina Press, Chapel Hill.
- Trattner, W. (1999). *From Poor Law to Welfare State: A History of Social Welfare in America*. Free Press, New York.
- Winn, S. (2003). "New budget paints ugly picture for state's art programs". *San Francisco Chronicle*, July 30.
- Wright, C.E. (1992). *The Transformation of Charity in Postrevolutionary New England*. Northeastern University Press, Boston.
- Zeigler, J. (1994). *Arts in Crisis: The National Endowment for the Arts versus America*. Chicago Review Press, Chicago.
- Zollmann, C. (1929). *The American Law of Charities*. The Bruce Publishing Company, Milwaukee.

## ACKNOWLEDGEMENTS

In September 2004 an authors' conference was held at Princeton University at which most of the chapters contained in this volume were presented. Contributors benefited enormously from the interchange and discussion at this conference and this assisted greatly in the process of revising chapters for publication. We wish to express our thanks to the Industrial Relations Section at Princeton, and especially to Orley Ashenfelter and Alan Krueger, for their invaluable support in making this meeting happen. We are also grateful to the Belgian Fonds National de la Recherche Scientifique, the Belgian Central Bank (Banque Nationale de Belgique) and the Department of Economics at Macquarie University, Sydney, for their financial assistance.

We also wish to thank Laura Billington for her help in the preparation and editing of this volume.

Victor Ginsburgh  
David Throsby  
*November 2005*

# CONTENTS OF THE HANDBOOK

## VOLUME 1

### PART 1: INTRODUCTION

#### *Chapter 1*

Introduction and Overview

DAVID THROSBY

### PART 2: HISTORY

#### *Chapter 2*

Art and Culture in the History of Economic Thought

CRAUFURD GOODWIN

#### *Chapter 3*

The History of Art Markets

NEIL DE MARCHI AND HANS J. VAN MIEGROET

#### *Chapter 4*

The Evolution of Music Markets

F.M. SCHERER

### PART 3: VALUE

#### *Chapter 5*

Defining Cultural and Artistic Goods

ROGER MCCAIN

#### *Chapter 6*

Value and the Valuation of Art in Economic and Aesthetic Theory

MICHAEL HUTTER AND RICHARD SHUSTERMAN

### PART 4: LEGAL ISSUES

#### *Chapter 7*

The Economic Analysis of Art Law

WILLIAM M. LANDES AND DANIEL B. LEVINE

*Chapter 8*

Copies of Artworks: The Case of Paintings and Prints

FRANÇOISE BENHAMOU AND VICTOR GINSBURGH

*Chapter 9*

Copyright, Art and Internet: Blessing the Curse?

PATRICK LEGROS

*Chapter 10*

Censorship versus Freedom of Expression in the Arts

TUN-JEN CHIANG AND RICHARD A. POSNER

## PART 5: CULTURE AND THE ECONOMY

*Chapter 11*

The Arts in the “New Economy”

WILLIAM J. BAUMOL

*Chapter 12*

Culture and Economic Performance

MARK CASSON

*Chapter 13*

Culture and Economic Development

PAUL STREETEN

## PART 6: CONSUMPTION AND PRODUCTION

*Chapter 14*

Empirical Studies of Demand for the Performing Arts

BRUCE A. SEAMAN

*Chapter 15*

Nonprofit Firms in the Performing Arts

ARTHUR C. BROOKS

*Chapter 16*

Creativity and the Behavior of Artists

WILLIAM D.A. BRYANT AND DAVID THROSBY

## PART 7: CULTURAL INDUSTRIES

*Chapter 17*

Organization of Arts and Entertainment Industries

RICHARD E. CAVES

*Chapter 18*

**The Media and Advertising: A Tale of Two-Sided Markets**

SIMON P. ANDERSON AND JEAN J. GABSZEWICZ

*Chapter 19*

**The Movies**

ARTHUR DE VANY

*Chapter 20*

**Rockonomics: The Economics of Popular Music**

MARIE CONNOLLY AND ALAN B. KRUEGER

*Chapter 21*

**The Economics of Books**

MARCEL CANOY, JAN C. VAN OURS AND FREDERICK VAN DER PLOEG

**PART 8: LABOR MARKETS**

*Chapter 22*

**Artistic Labor Markets: Contingent Work, Excess Supply and Occupational Risk Management**

PIERRE-MICHEL MENGER

*Chapter 23*

**Artists' Careers and Their Labor Markets**

NEIL O. ALPER AND GREGORY H. WASSALL

*Chapter 24*

**Human Capital and Artists' Labor Markets**

RUTH TOWSE

*Chapter 25*

**Stardom and Talent**

MOSHE ADLER

**PART 9: AUCTIONS AND PRICES**

*Chapter 26*

**Art Auctions**

ORLEY ASHENFELTER AND KATHRYN GRADDY

*Chapter 27*

**The Computation of Prices Indices**

VICTOR GINSBURGH, JIANPING MEI AND MICHAEL MOSES



**PART 10: HERITAGE***Chapter 28***Cultural Heritage: Economic Analysis and Public Policy**

ILDE RIZZO AND DAVID THROSBY

*Chapter 29***The Economics of Museums**

BRUNO S. FREY AND STEPHAN MEIER

**PART 11: LOCATION***Chapter 30***Culture in Urban and Regional Development**

TRINE BILLE AND GÜNTHER G. SCHULZE

*Chapter 31***Cultural Districts and Their Role in Developed and Developing Countries**

WALTER SANTAGATA

**PART 12: POLICY***Chapter 32***The Arts and Economic Policy**

ALAN PEACOCK

*Chapter 33***Culture in International Trade**

KEITH ACHESON AND CHRISTOPHER MAULE

*Chapter 34***The Making of Cultural Policy: A European Perspective**

FREDERICK VAN DER PLOEG

*Chapter 35***Cultural Policy: An American View**

DICK NETZER

*Chapter 36***Tax Incentives in Cultural Policy**

J. MARK SCHUSTER

*Chapter 37***Philanthropy**

STANLEY N. KATZ

## AUTHOR INDEX

n indicates citation in a footnote.

- AAFRC, 495, 1266, 1271  
Abbé-Decarroux, F. 425, 426, 431, 432, 441,  
443, 443n, 451n, 453, 456, 457, 460, 464,  
486, 1207  
Abbinante, C. 249  
Abbing, H. 879, 882, 883n, 1131n  
Abbott, A. 774, 796  
Abele, H. 1059, 1083  
Abowd, J. 931, 934  
Abraham, K. 952n  
Abrams, B.A. 1243  
Abramson, A., *see* Salamon, L. 1314  
Acemoglu, D. 876, 891  
Acheson, K. 988n, 1154n, 1155n, 1157n,  
1158n, 1173n–1175n  
Adam, T. 1317  
Adamowicz, W. 1075n  
Adamowicz, W., *see* Boxall, P. 1076n  
Addison, J. 204  
Adler, M. 695, 835, 897, 898, 899n  
Adorno, T.W. 190  
Aghion, P. 353n, 801  
Agnello, R.J. 973  
Agnew, J.-C. 178  
Akerlof, G. 92n, 746  
Aksoy, A. 773  
Alberini, A. 1077  
Albert, M., *see* Hahnel, R. 365n  
Albert, R.S. 511  
Alesina, A. 1002, 1006, 1192, 1210, 1211  
Alexander, J.R. 326  
Alexander, P.J. 313  
Alexander, P.J., *see* Brown, K. 589  
Alexander, P.J., *see* Cunningham, B.M. 581n,  
589  
Allen, W. 735  
Allerston, P. 76n  
Allmendinger, J. 771n, 779  
Alper, N. 164n, 769, 771n, 774, 776, 791, 794,  
795, 816, 817, 821, 827n, 829, 847, 848  
Alper, N., *see* Wassall, G. 770, 794, 816, 829,  
829n, 835n, 841, 842, 842n, 873, 902  
Alpers, S. 260n  
Alvarez, J.L. 1265  
Amabile, T. 156, 158, 158n  
American Association of Museums, 1236  
Americans for the Arts, 1228, 1228n  
Ames, E.K. 218n  
Amirault, T. 794  
Andeoni, J. 1244  
Anderson, A.B., *see* Faulkner, R.R. 553, 773n  
Anderson, B. 1150n  
Anderson, J.E., *see* Colonna, C.M. 492n  
Anderson, R.C. 14, 918–920, 932n, 949, 950,  
973  
Anderson, S.P. 580, 581n, 583n, 584, 587,  
589n, 591n–594n, 595, 595n, 605, 608n, 609,  
609n  
Andersson, C. 314  
Andreasen, A. 423, 459, 460, 463  
Andreoni, J. 480, 480n, 495  
Angela, B., *see* Cancian, M. 573  
Angrist, J. 842n  
Anheier, H.K. 1035, 1035n  
Annalin, 1215n  
Anon., 1024n, 1035n, 1041  
Ansburg, P.I. 157  
Anzieu, D. 775  
Appelbaum, E. 804  
Appelman, M.D. 736n, 737, 737n, 741, 743n,  
751n  
Arendt, H. 190, 792  
Arian, E. 777  
Aristotle, 173  
Armstrong, M. 580, 580n, 591n, 594, 594n,  
605  
Arnheim, R. 256, 275n  
Arnold, D. 141n  
Arnold, M. 49, 482n  
Arnold, T.W. 316  
Aronson, R.L. 775

- Arrow, K.J. 304, 513, 885, 997  
 Art Newspaper, 1024n  
 Asabere, P.K. 1246  
 Asheim, G.B. 495  
 Ashenfelter, O. 690, 916, 927, 928, 930, 931, 931n, 933–935, 937–940, 950, 966, 973, 974  
 Ashenfelter, O., *see* Abowd, J. 931, 934  
 Ashworth, J. 1022n  
 Asimow, M. 323  
 Association of College, University and Community Arts Administrators, Inc., 421n  
 Atkinson, A.B. 798n, 1201  
 Atkinson, G.D., *see* Pearce, D.W. 990, 994  
 Atlas, C., *see* Renz, L. 1306  
 Auerbach, A. 1243  
 Austen-Smith, D. 498  
 Austen-Smith, D., *see* Jenkins, S. 423, 425, 426, 441, 453, 455, 457  
 Austen-Smith, J. 498  
 Australia ICOMOS, 987  
 Ayer, A.J. 201  
 Ayrton, M. 261, 262n
- Bach, S. 647  
 Bagnasco, A. 1105, 1107  
 Bagwell, K. 583  
 Bailey, M. 918n, 954  
 Bailey, S.J. 1021, 1025, 1033n  
 Bailey, T., *see* Appelbaum, E. 804  
 Bainbridge, W.S., *see* Stark, R. 375n  
 Bajic, V. 440n, 441, 452n  
 Baker, A.J. 537n  
 Baker, W.E. 781, 796  
 Bakos, Y. 296  
 Bamossy, G. 460  
 Banerjee, A.V. 1156n  
 Banfield, E.C. 1260n  
 Banos Piño, J., *see* Fernández-Blanco, V. 430n, 436n, 450n, 1073  
 Baquir, R., *see* Alesina, A. 1006  
 Barber, B.R. 406, 1150  
 Barbier, E.B. 990  
 Barbier, P. 142n  
 Bardhan, P. 392  
 Barnatt, C., *see* Starkey, K. 551n  
 Baron, D.P. 548n  
 Baron, J. 151n, 801, 804  
 Barrett, K.S. 1273n  
 Barro, R. 393n, 753  
 Barros, C., *see* Brito, P. 443, 447  
 Barros, P.P. 581n
- Barten, A.P. 417  
 Bartik, T. 958, 959n, 1067n  
 Bateman, I.J. 1077n  
 Bauer, H., *see* Abele, H. 1059, 1083  
 Baumgarten, A. 181  
 Baumol, H. 490–492, 1081n, 1151  
 Baumol, H., *see* Baumol, W.J. 128, 356, 782  
 Baumol, H., *see* Rubinstein, A.J. 485, 494  
 Baumol, W.J. 4, 14, 28, 128, 195, 341n, 347n, 352n, 354n, 356, 419, 419n, 420, 464, 490–492, 496, 685, 775, 779, 782, 794, 801, 815, 918, 920, 922, 949, 950, 957, 973, 1019n, 1072n, 1073n, 1079–1081, 1081n, 1198, 1227, 1259n  
 Baumol, W.J., *see* Baumol, H. 490–492, 1081n, 1151  
 Baumol, W.J., *see* Rubinstein, A.J. 485, 494  
 Bauwens, L. 932, 934, 973  
 Baxter, J.L. 365n  
 Bayart, D. 1019n, 1039n  
 Bazin, G. 314, 315  
 Beach, C.P., *see* Katz, S.N. 1303  
 Beardsley, M.C. 192, 202, 203  
 Becattini, G. 1105, 1107  
 Becattini, G., *see* Pyke, F. 1105  
 Beck, J. 734, 748, 750n  
 Becker, G.S. 216, 240n, 297n, 434, 447n, 448n, 513, 586, 688, 868, 986n, 1020n  
 Becker, G.S., *see* Michael, R.T. 440  
 Becker, G.S., *see* Stigler, G.J. 162, 188, 447n, 448n, 545, 586, 898, 1074n, 1083n  
 Becker, H. 789n, 961  
 Beebe, J. 573, 573n, 581, 604  
 Beggs, A. 931, 932n, 933–935, 938–940, 964n, 973  
 Behrman, S.N. 1152  
 Belcher, M. 1038  
 Belinfante, A. 556  
 Belk, R., *see* Andreasen, A. 459  
 Belk, R.W., *see* Andreasen, A. 423, 459, 460, 463  
 Bell, C. 62  
 Belting, H. 271  
 Beltrami, D. 115  
 Benedict, P. 85  
 Benghozi, P.-J., *see* Bayart, D. 1019n, 1039n  
 Benhamou, F. 253n, 485, 835, 883, 985, 1002, 1003n, 1009, 1009n, 1019n, 1131n, 1174n, 1175n, 1188, 1245  
 Bénichou, P. 791

- Benjamin, W. 190, 271  
 Bennett, O. 50  
 Bentham, J. 42, 43, 161  
 Benyovszky, K. 138  
 Beresford, J., *see* Ellis, D. 831n, 832  
 Berg, M. 179n  
 Berg, P., *see* Appelbaum, E. 804  
 Berger, D.G., *see* Peterson, R.A. 557  
 Berger, M., *see* Hummel, M. 1059n  
 Bergstrom, J. 1165  
 Bergstrom, T. 480, 481n, 499, 1214  
 Bergstrom, T., *see* Cancian, M. 573  
 Berlinger, L., *see* Weick, K. 775  
 Berlioz, H. 134  
 Bernhardt, D. 939, 940  
 Berry, S. 557, 605, 958  
 Bertarelli, S. 736  
 Besanko, D., *see* Baron, D.P. 548n  
 Besen, S.M. 296, 699n, 700  
 Bessen, J. 293  
 Bettig, R. 889  
 Bhattacharjee, S., *see* Gopal, R.D. 710n, 714  
 Bianchini, F. 1093n  
 Bianchini, F., *see* Landry, C. 1063, 1069, 1090n  
 Bianconi, L. 130n  
 Bielby, D. 823  
 Bielby, D., *see* Bielby, W. 773, 774, 816, 822  
 Bielby, W. 773, 774, 816, 822  
 Bielby, W., *see* Bielby, D. 823  
 Biey, M.L. 918  
 Biglaiser, G. 545  
 Bikhchandani, S. 1156n  
 Bille, T. 1064n, 1073, 1073n, 1075, 1077, 1077n, 1191, 1198  
 Bilodeau, M. 481, 485  
 Bilodeau, M., *see* Steinberg, R. 1253n  
 Bina, V. 1194  
 Birdzell, L.E., *see* Rosenberg, N. 392  
 Bishop, R.-C. 1078n  
 Bittlingmayer, G. 737n  
 Black, J. 938, 940  
 Blackman, S.A.B., *see* Baumol, W.J. 347n, 492  
 Blair, R.D. 599  
 Blanchard, I. 87n  
 Blass, A.A. 905  
 Blattberg, R.C. 1040n  
 Blau, J.R. 562, 772, 785, 1020n, 1039n  
 Blau, J.R., *see* Foster, A.W. 1020n  
 Blaug, M. 4, 5, 339n, 494, 869, 871, 875, 883, 884, 1019n  
 Blaug, M., *see* King, K. 1261n  
 Blockmans, W. 79n, 80n, 88n, 98n, 115  
 Blockmans, W., *see* Previer, W. 115  
 Blogie, J. 973  
 Blokland, H. 1185n  
 Blondé, B. 85  
 Blondé, B., *see* Stabel, P. 77n  
 Blum, J. 129n  
 Blume, A.C. 75n  
 Blume, L., *see* Bergstrom, T. 480, 481n, 499  
 Blundell, R. 426n  
 Boden, M.A. 157, 159  
 Bodin, J. 30, 31  
 Boehlert, E. 705  
 Boers, M. 94n  
 Boin, J.G. 773  
 Boldrin, M. 290, 292n, 296, 300, 708  
 Bolnick, B.R. 368  
 Bolton, P. 559n  
 Bonaiuto, S., *see* Schuster, J.M. 1268n  
 Bonato, L. 423, 425, 441, 449  
 Bonet, L. 1063  
 Bonfait, O. 116  
 Bonus, H. 899n, 961  
 Book, S.H. 419n  
 Book, S.H., *see* Globerman, S. 423, 425, 429n, 439–441, 443, 447, 464, 466n, 489n, 1026n  
 Boorsma, P.B. 1265  
 Boorstin, E.S. 712  
 Borean, L. 100n  
 Borges, J. 272  
 Borghans, L. 695, 899, 899n, 900n  
 Borgonovi, F. 1244, 1290n  
 Borrione, P. 1115  
 Börsch-Supan, H. 1040n, 1041n  
 Boskin, M. 952n, 959  
 Boulding, K. 160n, 572, 573  
 Bourdieu, P. 193, 776, 782, 986n, 1020n, 1198  
 Bourdon, J. 277n  
 Bourreau, M. 1208  
 Bouuaert, I.C. 1260, 1263  
 Bouvaist, J.M., *see* Boin, J.G. 773  
 Bowen, W.G. 562  
 Bowen, W.G., *see* Baumol, W.J. 4, 28, 419, 419n, 420, 464, 490, 496, 685, 794, 815, 1019n, 1072n, 1073n, 1080, 1081n, 1198, 1227, 1259n  
 Bowles, S. 885  
 Boxall, P. 1076n  
 Boxall, P., *see* Adamowicz, V. 1075n  
 Braden, J.B. 996n, 1077n

- Brau, R. 1005  
 Bremner, R. 1302n  
 Brenson, M. 1270n  
 Brent, W. 326  
 Bridgeman, P.W. 149n  
 Bridgman, B. 1006  
 Brilliant, E. 1305  
 Brito, P. 443, 447  
 Brockwell, P.H. 319n  
 Broderick, C.J., *see* Blattberg, R.C. 1040n  
 Brody, E. 1256n, 1275n  
 Brooks, A.C. 482, 490n, 493, 495, 495n, 496, 498–500, 1088n, 1263, 1285  
 Brooks, A.C., *see* Kushner, R.J. 495  
 Brooks, A.C., *see* McCarthy, K. 480, 491, 492, 496, 998n, 1239  
 Brown, E. 485  
 Brown, J. 261, 958  
 Brown, K. 589  
 Brown, T.J. 599  
 Brownlee, W.E. 1306  
 Bruner, J. 156  
 Bryant, T. 161n  
 Brynjolfsson, E. 290n, 306n  
 Brynjolfsson, E., *see* Bakos, Y. 296  
 Bryson, B. 423n  
 Buccola, S. 939, 940  
 Budd, M. 198  
 Buclens, N. 918, 920, 950, 974  
 Bullard, J., *see* Lange, M.D. 1025n  
 Burckhardt, J. 270n  
 Burke, A.E. 773, 774  
 Burke, E. 40, 175n  
 Burke, P. 1163  
 Burleigh, N. 1308  
 Burns, P. 939, 940  
 Burns, T. 803  
 Burton, B.J. 974, 974n  
 Buruna, I. 361  
 Busenhardt, I., *see* Frey, B.S. 1035n, 1153, 1166  
 Butcher, S., *see* Alper, N. 769, 771n, 774, 795  
 Butters, G. 585  
  
 C.S. Mott Foundation, 1312  
 Cabizza, M.M. 572n  
 Cable, D.M., *see* Voss, G.B. 561n  
 Cahn, S. 1170, 1173  
 Caillaud, B. 571n, 580, 580n, 594n, 602  
 Caillois, R. 271  
 Calkins, M.L. 326  
 Calvert, C., *see* Richards, R.D. 322  
  
 Camerer, C. 785, 793  
 Cameron, S. 429n, 430n, 447, 447n  
 Campbell, J. 922n  
 Campbell, L. 81n, 88n  
 Campos, L., *see* Castañer, X. 878  
 Cancian, M. 573  
 Candela, G. 950, 965, 973, 974  
 Cannon-Brookes, P. 985n, 1153  
 Canoy, M.F.M., *see* Appelman, M.D. 743n  
 Cantor, J.E. 1033  
 Cappelli, P. 801n  
 Cardell, N. 973  
 Carlsson, F. 1076n  
 Carmilly-Weinberger, M. 313, 314, 314n, 316–320  
 Carmody, C. 1170  
 Carpenter, D. 1212  
 Carson, A. 425, 432n, 442, 451  
 Carson, R.T. 1078, 1199  
 Carson, R.T., *see* Bateman, I.J. 1077n  
 Carson, R.T., *see* Mitchell, R.C. 1077n, 1078n  
 Carter, H.O., *see* Halter, A.N. 486  
 Carter Hill, R. 918  
 Casagrandi, R., *see* Rinaldi, S. 515  
 Case, K.E. 918, 918n, 963, 964, 967, 968n, 973  
 Cassady Jr., R. 631, 636  
 Casson, M.C. 362, 366n, 369, 374, 375  
 Castañer, X. 878  
 Castelnuovo, E. 256  
 Cavazzini, P. 102n  
 Caves, R.E. 160, 162, 535n, 560n, 621, 675, 723, 734, 737–739, 741, 780n, 781, 784, 787, 788, 876, 878, 882, 889, 1033, 1086n, 1154n  
 Caygill, H. 179, 180n  
 Cecchini, I. 99n–101n  
 Cellini, R. 516, 1005, 1113, 1114n  
 Censolo, R., *see* Bertarelli, S. 736  
 Center for Effective Philanthropy, 1311  
 Chae, S. 1208  
 Chambers, F.P. 311, 312  
 Chamoux, F. 256n, 260, 262, 262n, 273  
 Champy, F. 772  
 Chanel, O. 916, 918–920, 932, 934, 939, 940, 950, 967, 968, 973, 974, 996  
 Chartrand, H., *see* Alper, N. 769, 771n, 774, 795  
 Chastel, A. 985n  
 Chastel, A., *see* Chamoux, F. 256n, 260, 262, 262n, 273  
 Chevalier, J., *see* Goolsbee, A. 740n  
 Chiang, T.J. 213n

- Chisholm, D.C. 552, 657  
 Choi, J.P. 587, 595, 608n  
 Chomsky, N. 157  
 Christoffersen, H. 1059  
 Christopherson, S. 551, 773, 774, 778, 796n, 1109  
 Christopherson, S., *see* Storper, M. 551  
 Chronicle of Philanthropy, 1307, 1315  
 Chua, V.C.H. 1280  
 Chua, V.C.H., *see* Wong, C.M. 1280  
 Chung, K. 653, 904  
 Citro, C. 829n, 832  
 Clapp, J. 311–316, 319, 320, 964n  
 Clark, D.E. 1023, 1066n, 1076n, 1093n  
 Clark, K. 62  
 Clawson, M. 1247  
 Clay, H. 51  
 Clerides, S.K. 732n  
 Clotfelter, C.T. 1271, 1272, 1277, 1278  
 Coase, R.H. 535, 635, 704, 705, 781  
 Coate, S., *see* Anderson, S.P. 580, 583n, 584, 587, 589n, 591n–594n, 595, 595n, 608n  
 Coates, D. 429  
 Cocq, E., *see* Messerlin, P. 1169  
 Coen, P. 102n  
 Coffey, S. 1258  
 Cohen, G.A. 793  
 Cohen, H. 156–158  
 Colbert, F. 419  
 Cole, J.R., *see* Kingston, P. 829, 829n  
 Coleman, J.S. 1112  
 Collier, P. 404  
 Collingwood, R.G. 197  
 Collins, A. 621  
 Collins, J. 62  
 Colombet, C. 276n, 279  
 Colonna, C.M. 492n  
 Commanducci, R. 74n  
 Commission on Private Philanthropy and Public Needs, 1259, 1268  
 Conant, M. 551, 620, 632n, 633n  
 Conaway, M.C., *see* Carson, R.T. 1199  
 Conner, K. 292n  
 Contractor, N.S., *see* Waterman, D. 452  
 Cook, P.J., *see* Frank, R.H. 786, 791, 835, 901  
 Cooke, A., *see* Paton, D. 431n  
 Cooper, C. 1058  
 Cooper, S. 319, 321  
 Corbett, A.R. 291  
 Cordes, J.J. 1228n, 1236, 1237, 1272n, 1281, 1282  
 Cordes, J.J., *see* Brody, E. 1275n  
 Cordone, R., *see* Rinaldi, S. 515  
 Corneo, G. 575, 575n  
 Cornes, R. 481  
 Corning, J. 417, 423, 425, 427, 432, 449, 453, 457  
 Coser, L.A. 739, 784n  
 Coslovich, G. 1289  
 Costanza, R. 988  
 Coulsen, N.E. 1246  
 Council for Museums, Archives and Libraries, 1258  
 Court, A.T. 952  
 Courty, P. 676n, 689  
 Cowen, T. 492, 493, 516n, 736, 736n, 758, 759, 786, 875, 890, 891n, 1024n, 1081n, 1150, 1154n, 1155, 1186  
 Cox, R., *see* Chung, K. 653, 904  
 Crampes, C. 587, 607n  
 Crane, C. 1305  
 Creemers, M. 740n  
 Creigh-Tyte, S. 1009n  
 Critchlow, D. 1305  
 Croce, B. 190, 197  
 Crone, T. 968n  
 Cropper, M. 959  
 Csikszentmihalyi, M., *see* Getzels, J.W. 536n  
 Cuccia, T. 1110, 1113, 1114n, 1126n  
 Cuccia, T., *see* Cellini, R. 516, 1113, 1114n  
 Cummings Jr., M.C. 742  
 Cunningham, B.M. 581n, 589  
 Curtis, M.K. 324  
 Curwen, P., *see* Allen, W. 735  
 Cuzin, J.-P. 262n  
 Cuzin, J.-P., *see* Laclotte, M. 259  
 Cwi, D. 419, 1057, 1059  
 Czujack, C. 918, 920, 961n, 973, 974  
 Dacey, J.S. 511  
 Dale, M. 1156n  
 Daly, H., *see* Costanza, R. 988  
 Danaher, P.J. 599  
 D'Angelo, M. 1317  
 Danto, A.C. 273n, 288, 288n, 289, 961  
 Darnell, A.C. 1021, 1034  
 Dasgupta, P.S. 989  
 Dasgupta, S. 158n  
 d'Aspremont, C. 596  
 David, P.A. 297  
 Davidson, B. 1236  
 Davidson, C., *see* Deneckere, R. 608n

- Davison, R., *see* Wassall, G.H. 794, 816, 829, 829n, 835n  
 Day, B., *see* Bateman, I.J. 1077n  
 DCMS, 1193  
 de Grauwe, P. 737n  
 de Haan, J., *see* van den Broek, A. 1217n  
 de Kam, C.A., *see* Goudriaan, R. 422n, 423, 425, 439n, 441, 442, 450  
 de la Barre, M. 918, 920, 925, 927, 928, 962, 966, 974  
 de Limburg-Stirum, M. 974  
 De Long, B. 295n  
 De Marchi, N. 15, 71n, 75n, 82n, 90n–92n, 96n, 98n, 99n, 104n, 107n, 108n, 173n, 179n, 261, 265, 266, 279, 912n, 960, 960n  
 de Meza, D., *see* Black, J. 938, 940  
 de Monchaux, J., *see* Schuster, J.M. 985n, 1265  
 de Montebello, P. 1040  
 de Piles, R. 173n, 960  
 de Saint-Simon, C. 260  
 De Swaan, A. 1196n  
 De Vany, A. 620, 621, 623, 623n, 627, 629n, 630n, 633, 633n, 634n, 635, 636, 638–643, 643n, 644, 645, 651, 652n, 653–659, 684, 685, 685n, 1156n  
 De Vries, J. 75n, 86n, 94n  
 Debeauvais, R. 778  
 Debreu, G. 188  
 Deck, L., *see* Cropper, M. 959  
 Degand, J.C. 973  
 Delberger, E., *see* Boskin, M. 952n, 959  
 Delluc, B. 272n  
 Delluc, G., *see* Delluc, B. 272n  
 Delon, M. 129n  
 Deltas, G. 939, 940  
 Demortier, G. 962, 974  
 Demsetz, H. 513  
 Deneckere, R. 608n, 747  
 Denucé, J. 82n  
 Dertouzos, J.N. 599n  
 Desbois, H. 276  
 Dewatripont, M. 289n, 1210  
 Dewenter, R. 1073, 1083  
 Dewey, J. 171, 192, 198  
 Dexter, G., *see* Lord, B. 1032n  
 di Vittorio, A. 939, 940  
 Diamond, P.A. 1078n  
 Diamond, S. 1240  
 Dickenson, V. 1022n  
 Dickie, G. 193  
 Diewert, E. 958n, 959  
 DiMaggio, P. 419n, 421, 422, 423n, 461, 461n, 462, 463, 476n, 482, 483n, 561n, 774, 800, 802, 1020n, 1026n, 1039n, 1040n, 1310  
 DiNardo, J. 423n  
 Dixit, A. 517n, 585, 736, 1146n  
 Dobson, L. 441, 453, 456, 457, 461  
 Docclo, S., *see* Chanel, O. 950  
 Docclo, S., *see* de la Barre, M. 918, 920, 925, 927, 928, 962, 966, 974  
 Donahue, S.M. 551  
 Donald, S. 939, 940  
 Dorchy, L. 973  
 Doyle, C. 1208  
 Doyle, G. 1161n  
 Drazen, A., *see* Alesina, A. 1192  
 Drèze, J. 794, 798  
 DRM Watch Staff, 292n  
 Dubus, P. 261  
 Duchamp, M. 287–289, 297, 298n  
 Duffy, E.A., *see* Bowen, W.G. 562  
 Dukes, A.J. 584n, 586, 589, 589n, 590n, 595, 608n  
 Dukes, A.J., *see* Gal-Or, E. 581n, 589, 597n, 608n  
 Durlauf, S. 885  
 Duverger, E. 82n, 83n  
 Earl, P.E. 374  
 Easterlin, R.A. 364  
 Easterly, W. 1055n  
 Easterly, W., *see* Alesina, A. 1006  
 Eaton, B.C. 573n, 593n, 608n  
 Eberts, J. 1168  
 Ecalle, F. 737n  
 Echeverría, I. 96n, 97n  
 Eckel, C.C. 1276  
 Eckert, R., *see* De Vany, A. 620, 621, 623n, 629n, 630n, 633, 633n, 634n, 635  
 Economides, N. 579n  
 Ehrensaft, P. 417, 417n  
 Ehrlich, C. 139, 771n, 782  
 Eichenberger, R., *see* Frey, B.S. 923, 974  
 Einarsson, A. 1289  
 Einav, L. 620n  
 Ekelund, R. 950, 974  
 Ekelund Jr., R.B. 374, 425, 436, 437, 450  
 El Sarafy, S. 988  
 Elias, N. 128  
 Eliot, M. 675, 699  
 Eliot, T.S. 203

- Ellickson, R.C. 395  
 Ellis, A. 998n  
 Ellis, D. 831n, 832  
 Ellison, G. 306n  
 Elsen, A.E. 1039n  
 Elsen, A.E., *see* Merryman, J.H. 273, 273n, 276, 276n, 278  
 Elster, J. 792  
 Engelbrecht-Wiggans, R. 939, 940  
 Engers, M.P., *see* Anderson, S.P. 593n  
 Engle, R. 965  
 Englin, J., *see* Boxall, P. 1076n  
 Engwall, L. 599  
 Epple, 958  
 Epstein, D. 1211  
 Epstein, J. 740  
 Epstein, R.A. 1243  
 Erekson, H., *see* Wittke, A. 958  
 Ernst, R., *see* Glück, J. 510  
 Escalera, J. 452  
 Etlin, R.A. 512n  
 Eurostat, 728  
 Evans, G. 1069n, 1093n  
 Ewing, D. 87n
- Facchina, B. 1257n  
 Faine, H.R. 489n  
 Falconer, P., *see* Bailey, S.J. 1021, 1025, 1033n  
 Falomir, M. 95n, 97n, 115  
 Fantel, H. 139, 139n  
 Fantoni, M. 74n  
 Faulkner, R.R. 553, 554, 771n, 773n, 779, 781, 790, 795n  
 Faulkner, R.R., *see* Baker, W.E. 781, 796  
 Favaro, E. 99n, 100n, 115  
 Fazioli, R. 489, 492  
 Federal Committee on Statistical Methodology, 817  
 Federal Trade Commission, 620, 636  
 Federico, G., *see* Foreman-Peck, J. 378  
 Federico, R. 779  
 Fee, C.E. 542n  
 Feenstra, R. 958n  
 Feigenbaum, S. 1009  
 Feist, A. 767  
 Feist, A., *see* O'Brien, J. 769, 770  
 Feld, A. 1241n, 1258, 1260, 1260n, 1261, 1263, 1264n, 1268, 1268n, 1269, 1270, 1273, 1274, 1281, 1286n, 1293  
 Feld, A., *see* O'Hare, M. 1033
- Feld, L., *see* Pommerehne, W.W. 924, 974, 1032n  
 Feldstein, M. 1019n, 1033, 1040n, 1274, 1276, 1277  
 Feller, G. 497  
 Felton, M.V. 423, 425, 427, 428, 431, 450, 450n, 453, 456, 464, 492, 829n, 1073, 1231n, 1242n  
 Felton, M.V., *see* Greckel, F. 423, 425, 431, 450n, 453, 455  
 Fernández, Y., *see* García, M.I. 1193, 1194  
 Fernández-Blanco, V. 430n, 436n, 450n, 452, 1073  
 Fernández-Blanco, V., *see* Prieto-Rodríguez, J. 452, 1029n, 1198  
 Ferrando, J. 580, 602n  
 Feuillolay, M. 950, 973  
 Feurt, S., *see* Luck, J. 1313  
 Fiedler, J. 136n  
 Figini, P., *see* Candela, G. 965, 974  
 Filer, R.K. 775, 790, 816, 830, 832, 838, 841, 842, 847, 847n, 873, 874, 1036n  
 Filippini, M., *see* Fazioli, R. 489, 492  
 Fine, B. 885  
 Fine, M. 711  
 Finn, A., *see* Hoskins, C. 1156  
 Finney, S., *see* Whitehead, J. 1072n  
 Fiorina, M. 1211  
 Fisher, C.G. 423n  
 Fisher, R.W., *see* Wallace, J.A. 1267  
 Fisher, W.W.I. 714  
 Fishwick, F. 734, 737n, 743n  
 Fitzpatrick, J. 322  
 Fitzsimons, S., *see* Fishwick, F. 737n, 743n  
 Fletcher, J., *see* Cooper, C. 1058  
 Flores, D., *see* Chae, S. 1208  
 Flores, N.E., *see* Carson, R.T. 1078  
 Flores, R. 974  
 Flores, R., *see* Czujack, C. 973  
 Florida, R. 803, 1067–1069  
 Flynn, L.J. 699  
 Focillon, H. 259  
 Focke, A. 1306  
 Fogelman, V.M. 1265  
 Fohrbeck, K. 777  
 Foley, M., *see* Bailey, S.J. 1021  
 Fondazione Eni Enrico Mattei, 1006  
 Ford Foundation, 419n, 421, 422, 441, 464, 464n, 1073n  
 Foreman-Peck, J. 378  
 Forrest, D. 423, 425, 427, 452, 465



- Fort, R. 429n  
 Foster, A.W. 1020n  
 Foster, T., *see* Maddison, D. 1019n, 1024, 1196  
 Foucart, A., *see* Chamoux, F. 256n, 260, 262, 262n, 273  
 Foucart, L. 271n  
 Foundation Center, 1313  
 Fox, R.F. 605n  
 Francastel, P. 260  
 Francois, P. 1149  
 Frank, B. 1079n  
 Frank, R.H. 371, 430, 485, 786, 791, 835, 901  
 Fredericksen, B.B., *see* Peronnet, B. 265, 281  
 Frederiks, J.G. 116  
 Frederiks, P.J., *see* Frederiks, J.G. 116  
 Freeman, A.M. 996n  
 Freeman III, A.M. 1077n  
 Freidson, E. 774, 776, 792, 796  
 Frey, B.S. 339n, 482, 515, 770, 787, 878, 887, 918, 920, 923, 950, 957, 974, 997n, 999, 1002n, 1019n, 1020n, 1023, 1029n, 1030, 1032, 1034, 1035n, 1036n, 1072n, 1075, 1078, 1078n, 1079n, 1081n, 1125n, 1136n, 1153, 1166, 1192, 1198n, 1199, 1240, 1242, 1247, 1270, 1284n  
 Frey, B.S., *see* Meier, S. 1019n, 1029n  
 Fridberg, T., *see* Bille, T. 1073n  
 Friedman, L. 1303  
 Friedman, M. 1307  
 Fronville, C.L. 1040n  
 Froomkin, M., *see* De Long, B. 295n  
 Fry, R. 62  
 Fujita, M. 1087n  
 Fullerton, D. 1262, 1264n, 1272, 1277  
 Furhammar, L. 325  
 Furhoff, L. 599, 602  
  
 Gabszewicz, J.J. 573n, 580n, 582, 583, 587, 596, 597, 597n, 601, 601n  
 Gabszewicz, J.J., *see* d'Aspremont, C. 596  
 Gabszewicz, J.J., *see* Ferrando, J. 580, 602n  
 Gadamer, H.G. 190  
 Gage, N. 311  
 Gagliardi, F., *see* Bonato, L. 423, 425, 441, 449  
 Gal-Or, E. 581n, 589, 597n, 608n  
 Gal-Or, E., *see* Dukes, A.J. 589, 590n  
 Galassi, M.C. 83n  
 Galbraith, J.K. 60  
 Gale, D.E. 1246  
 Gale, I. 939, 940  
 Galenson, D.W. 816, 818, 924–926, 926n  
  
 Galiani, F. 32  
 Galligan, A. 1270n  
 Galligan, A., *see* Alper, N. 791  
 Gandal, N. 939, 940  
 Gange, M. 605n  
 Ganzeboom, H.B.G. 422n, 441, 460  
 Gapinski, J.H. 422n, 423, 425, 426, 428n, 432, 436, 436n, 440, 441, 450, 451, 451n, 464n, 484, 487, 489, 492, 1021, 1021n, 1073, 1073n  
 Gaquin, D., *see* Citro, C. 829n, 832  
 García, J. 428n, 429n  
 García, M.I. 1193, 1194  
 García Álvarez, E.G., *see* Lopéz Sintas, J. 423n, 1198  
 García Fuentes, P.G. 97n  
 Gardner, H. 158  
 Garreau, J. 1089, 1109  
 Garrison, L.C., *see* Kaali-Nagy, C. 449n, 460  
 Gatzlaff, D. 966  
 Gauthier, M.M., *see* Chamoux, F. 256n, 260, 262, 262n, 273  
 Gautier, P.Y. 276  
 Gayer, A. 291, 291n, 714  
 Gazel, R. 1060  
 Gedo, J.E. 512  
 Gedo, M.M., *see* Gedo, J.E. 512  
 Gelderblom, O. 80n, 90n, 92n  
 Gemser, G., *see* Wijnberg, N.M. 540n  
 Genesove, D. 933, 936  
 Genesove, D., *see* Ashenfelter, O. 939, 940  
 George-Warren, H. 677n  
 Gerard, K., *see* Ryan, M. 1076n  
 Gérard-Varet, L., *see* Chanel, O. 916, 918–920, 932, 934, 939, 940, 950, 967, 968, 973, 974, 996  
 Getz, D. 1039n  
 Getzels, J.W. 536n  
 Getzner, M. 1079n, 1192  
 Ghemawat, P. 305n  
 Ghirdelli, P., *see* Shook, J.R. 151n  
 Ghoneim, A.F. 1167n  
 Ghosh, A. 638  
 Giacotto, C., *see* Clapp, J. 964n  
 Giardina, E. 1003  
 Gibson, L. 1070  
 Gielens, G., *see* de Grauwe, P. 737n  
 Gilbert, C.E. 74n, 115  
 Gilbert, D., *see* Cooper, C. 1058  
 Gilhooly, K.J. 511  
 Gilmore, J.H., *see* Pine, B.J. 1075n, 1086, 1200

- Gimpel, J. 313  
 Ginsburg, J.C. 222  
 Ginsburgh, V. 5, 173n, 213n, 232, 328n, 365n, 781, 785n, 904, 918, 924, 925, 930, 939, 940, 960n, 962, 973, 974, 1013–1015, 1019n, 1034n, 1178  
 Ginsburgh, V., *see* Ashenfelter, O. 930  
 Ginsburgh, V., *see* Bauwens, L. 932, 934, 973  
 Ginsburgh, V., *see* Benhamou, F. 253n  
 Ginsburgh, V., *see* Buelens, N. 918, 920, 950, 974  
 Ginsburgh, V., *see* Chanel, O. 916, 918–920, 950, 967, 968, 974, 996  
 Ginsburgh, V., *see* Czujack, C. 973  
 Ginsburgh, V., *see* de la Barre, M. 918, 920, 925, 927, 928, 962, 966, 974  
 Ginsburgh, V., *see* de Limburg-Stirum, M. 974  
 Ginsburgh, V., *see* di Vittorio, A. 939, 940  
 Ginsburgh, V., *see* Dorchy, L. 973  
 Ginsburgh, V., *see* Flores, R. 974  
 Gintis, H., *see* Bowles, S. 885  
 Ginzburg, C. 262n  
 Givon, M. 292n  
 Glaeser, E. 886  
 Glaeser, E.L. 784, 1027n, 1031n, 1090n  
 Glasser, H. 72n  
 Glazer, A. 1031  
 Gliberman, S. 421n, 423, 425, 429n, 439–441, 443, 447, 464, 466n, 489n, 495, 1026n  
 Gliberman, S., *see* Book, S.H. 419n  
 Glück, J. 510  
 Godfrey, C., *see* Peacock, A.T. 1019n  
 Goettler, R. 575n, 604, 606  
 Goetzmann, W.N. 15n, 915, 915n, 916, 918, 918n, 919–922, 924, 925, 950, 955n, 963, 966–968, 973, 974  
 Goetzmann, W.N., *see* Oster, S. 1022, 1031n  
 Gold, S. 438  
 Goldberg, M.A., *see* Mark, J.H. 964, 967  
 Goldberg, P. 958  
 Goldberg, V.P. 552n  
 Goldman, W. 536, 621, 647  
 Goldstone, L. 1075n, 1177  
 Goldthwaite, R.A. 74n, 79n, 98n, 115  
 Gombrich, E. 960, 1134n  
 Gómez, M.V. 1037n, 1063, 1070  
 Goodale, G. 497  
 Goodman, N. 192, 271  
 Goodman, R.P., *see* Bergstrom, T.C. 1214  
 Goodwin, C.D. 62–65  
 Goolsbee, A. 740n  
 Gopal, R.D. 710n, 714  
 Gopnik, A. 240  
 Gordon, K., *see* Leclair, M.S. 1012n  
 Gordon, R., *see* Boskin, M. 952n, 959  
 Gorelli, S., *see* Bonato, L. 423, 425, 441, 449  
 Goudriaan, R. 422n, 423, 425, 439n, 441, 442, 450, 487, 1021  
 Gozzano, N. 103n  
 Graddy, K., *see* Ashenfelter, O. 916, 930, 931, 933–935, 950, 966, 973, 974  
 Graddy, K., *see* Beggs, A. 931, 932n, 933–935, 938–940, 964n, 973  
 Graeser, P. 919, 920, 922, 973  
 Grafton, R.Q. 996n  
 Graham, J.L. 375  
 Graham, M., *see* Bailey, S.J. 1021  
 Grampp, W.D. 48, 49n, 221, 240n, 961, 1003n, 1025, 1032, 1186  
 Graña, C. 782  
 Grassley, C. 1315  
 Graves, F.L. 1064n  
 Gray, C.M. 419, 421, 422n, 441, 461, 489, 1022n, 1073n, 1074  
 Gray, C.M., *see* Heilbrun, J. 421, 1019n, 1033n, 1035, 1035n, 1062n, 1084n, 1089, 1234  
 Gray, J. 1150  
 Gray, L.S. 797n  
 Greckel, F. 423, 425, 431, 450n, 453, 455  
 Greco, A.N. 556n, 732n  
 Green, J., *see* Kohut, A. 393  
 Green, J., *see* Scotchmer, S. 293n  
 Green, J.R. 293n  
 Green, R. 465n  
 Greenblatt, R., *see* Alper, N. 769, 771n, 774, 795  
 Greenless, J., *see* Abraham, K. 952n  
 Greffe, X. 1116  
 Gregersen, M., *see* Bille, T. 1075, 1191  
 Greve, A. 77n  
 Greve, A., *see* Stabel, P. 77n  
 Grief, A. 71n  
 Griffiths, A. 257n, 263n  
 Griliches, Z. 842n, 952, 958  
 Griliches, Z., *see* Boskin, M. 952n, 959  
 Grimes, K., *see* Forrest, D. 423, 425, 427, 452, 465  
 Grin, F., *see* Abbé-Decarroux, F. 432, 441, 443, 443n, 464  
 Grodecki, L., *see* Chamoux, F. 256n, 260, 262, 262n, 273

- Grogan, C., *see* Schuster, J.M. 1268n  
 Groot, L., *see* Borghans, L. 695, 899, 899n, 900n  
 Gross, J.M., *see* Krasilovsky, M.W. 697, 698, 699n, 700, 706  
 Grossman, A., *see* Letts, C. 1311  
 Grossman, G. 255n, 584n, 585, 589, 590n  
 Grossman, P.J., *see* Eckel, C.C. 1276  
 Groves, T. 635  
 Gruenberg, B. 422n  
 Guerrieri, P. 1177n  
 Guerzoni, G. 950, 985  
 Guillory, J. 178, 179  
 Gurgand, M. 880  
 Gurgand, M., *see* Menger, P.-M. 798, 836n, 902  
 Gurney, A. 1226n  
 Gustafsson, K.E. 599, 602  
 Gwartney, J. 393
- Haack, S. 151n  
 Habermas, J. 792  
 Hackman, R., *see* Allmendinger, J. 771n, 779  
 Hahn, R.W. 293  
 Hahnel, R. 365n  
 Hall, P.D. 475, 1302n  
 Hall, R.E. 784  
 Haloman, D.K. 139n  
 Halter, A.N. 486  
 Halvorsen, R. 953n, 960n  
 Hamilton, J. 570  
 Hamlen Jr., W.A. 141n, 652, 696, 780, 903  
 Hammack, D. 1302  
 Hampden-Turner, C. 377n  
 Hand, C. 621, 638  
 Hand, C., *see* Collins, A. 621  
 Handy, C. 795n  
 Handy, F. 484  
 Hanemann, M. 1078n  
 Hanemann, M., *see* Bateman, I.J. 1077n  
 Hanley, N., *see* Bateman, I.J. 1077n  
 Hansen, C. 583  
 Hansmann, H. 227, 481, 483, 486, 494n, 543n, 561n, 1026, 1028, 1031  
 Harbaugh, W.T. 1031  
 Harburg, E., *see* Rosenberg, B. 549  
 Harhoff, D., *see* Scherer, F.M. 619n  
 Haritchabalet, C., *see* Crampes, C. 587, 607n  
 Harr, J., *see* Johnson, P. 1303  
 Harris, C. 1197  
 Harris, M. 1105
- Harrison, B., *see* Storper, M. 1107  
 Harrison, T., *see* Coates, D. 429  
 Hart, O. 540n, 541, 1210  
 Hart, P. 562n  
 Hart, T. 497  
 Hartwick, J.M. 989, 989n  
 Harvey, D., *see* O'Hagan, J. 1012n  
 Haskell, F. 102n  
 Hassan, Z.A., *see* Green, R. 465n  
 Haurin, D., *see* Gatzlaff, D. 966  
 Hausch, D., *see* Gale, I. 939, 940  
 Hausman, J.A., *see* Diamond, P.A. 1078n  
 Havens, J., *see* Schervish, P. 1314  
 Hawkins, A. 1264n  
 Hayes, N. 510  
 Hazlett, T. 605  
 Heal, G.M., *see* Dasgupta, P.S. 989  
 Heberlein, T., *see* Bishop, R.-C. 1078n  
 Heckman, J. 966  
 Hegel, G.W.F. 185  
 Heidegger, M. 190  
 Heifetz, A., *see* Hvide, H.K. 544  
 Heilbrun, J. 420, 421, 435n, 438, 449, 450, 452, 484, 493n, 1019n, 1033n, 1035, 1035n, 1062n, 1081n, 1084n, 1089, 1234, 1239n  
 Heinen, M.L. 973  
 Hellwig, M. 296n  
 Henley, J. 325  
 Hepburn, J. 544  
 Heravi, S., *see* Silver, M. 959  
 Herlihy, D. 115  
 Herrick, B., *see* Kindleberger, C. 1055  
 Hesmondhalgh, D. 773n, 784n, 796n  
 Hett, T., *see* Bateman, I.J. 1077n  
 Heunks, F.J. 514  
 Hibbert, C. 103n  
 Hicks, J.R. 188  
 Hightower, C. 1259  
 Hill, K., *see* Ansbury, P.I. 157  
 Hillage, J., *see* Jackson, C. 791  
 Himmelstein, J. 1307  
 Hirsch, E.D. 201  
 Hirsch, P. 784n  
 Hirschman, A.O. 378n, 404, 792  
 Hirschleifer, D., *see* Bikhchandani, S. 1156n  
 Hirschleifer, J. 1020n  
 Hjorth-Andersen, C. 455n, 737n, 743n, 1205  
 Hjorth-Andersen, C., *see* Bille, T. 1075, 1191  
 Hocking, J.G., *see* Halter, A.N. 486  
 Hoefler, A., *see* Collier, P. 404  
 Hoet, G. 265, 280

- Hofstede, G. 375  
 Hol, B. 256n  
 Holahan, W.L. 747  
 Holden, B.R., *see* Roberts, B. 391n  
 Holden, J. 998  
 Holland, B. 141n  
 Hollister, V., *see* Throsby, D. 522, 769, 775, 817, 824, 829  
 Holmes, M. 74n  
 Holmes, S. 375  
 Holtzmann, B. 259, 279  
 Honey, S., *see* Jackson, C. 791  
 Hood, R.S. 1280  
 Hope, E. 514  
 Horowitz, H. 417n  
 Horstmann, I. 938  
 Horvth, C., *see* Yetkiner, I.H. 740n  
 Hosios, A. 918n  
 Hoskins, C. 1156  
 Hotelling, H. 572  
 Hough, G. 202  
 House, J. 319  
 Housman, A.E. 510n  
 Houthakker, H.S. 424, 425, 430, 442, 442n, 443, 464  
 Hoving, T. 256, 256n, 260  
 Howard, N. 1304  
 Howitt, P., *see* Aghion, P. 353n, 801  
 Howson, S. 58  
 Hrycak, A., *see* Abbott, A. 796  
 Huffman, F.E., *see* Asabere, P.K. 1246  
 Hughes, A. 271  
 Hughes, P.N. 1284  
 Hui, K. 290n, 709n, 710n  
 Huisman, A. 281  
 Hull, P.C., *see* Peterson, R.A. 419, 420, 422, 441, 466n  
 Hume, D. 33, 34, 176, 1134  
 Hummel, M. 1059n  
 Hunt, R.M., *see* Bessen, J. 293  
 Huntingdon, S.P. 404  
 Huntington, P. 427n, 494  
 Hutcheson, F. 37, 40, 177, 178, 178n  
 Hutter, M. 8n, 985n, 998n, 1014, 1015  
 Hvide, H.K. 544  
 Iapadre, L. 1174n, 1175n  
 Iapadre, L., *see* Guerrieri, P. 1177n  
 ICARE, International Centre of Studies on the Economics of the Arts, 1040  
 Ihlanfeldt, K. 1090, 1090n  
 Ilchman, W. 1315  
 Ilott, T., *see* Eberts, J. 1168  
 In Re Auction Houses Antitrust Litigation, 930  
 Independent Commission, 1270n  
 Independent Sector, 475, 477, 485, 495  
 Inkei, P. 1264n, 1265, 1266, 1269, 1280n  
 International Center for Not-for-Profit Law, 1264n, 1265, 1280n  
 International Federation of Arts Councils and Culture Agencies, 1264n, 1265, 1280n  
 International Intellectual Property Alliance, 345  
 Irmen, A., *see* Hellwig, M. 296n  
 ITC, *see* WIPO, 1111  
 Ito, T., *see* Kurabayashi, Y. 420  
 Jackendoff, R. 157, 157n  
 Jackson, C. 791  
 Jackson, K. 311, 1227n  
 Jackson, R. 1025  
 Jacobs, P., *see* Lange, M. 437n, 1025n  
 Jacobsen, J.P., *see* Burton, B.J. 974, 974n  
 Jalil Moreno, Y. 1106  
 James, E. 487, 1029n  
 Jana, R. 497  
 Jansson, A.M. 988  
 Jeancolas, J.-P. 1147n  
 Jeanfils, P., *see* Flores, R. 974  
 Jeanfils, P., *see* Ginsburgh, V. 924, 925, 974  
 Jeffri, J. 476n, 776, 816, 883, 902  
 Jeffri, J., *see* Alper, N. 769, 771n, 774, 795  
 Jeffri, J., *see* Baumol, W.J. 775, 779  
 Jegen, R., *see* Frey, B.S. 1284n  
 Jencks, C. 1260n, 1274  
 Jenkins, S. 423, 425, 426, 441, 453, 455, 457  
 Jenkinson, T., *see* Feigenbaum, S. 1009  
 Jennings, M., *see* O'Hagan, J. 1207n  
 Jensen, R. 540n  
 Jeromack, P. 154  
 Jevons, W.S. 56–58  
 Jewitt, I., *see* Dewatripont, M. 1210  
 Johnson, P. 1019n, 1059, 1135n, 1303  
 Johnson, P., *see* Ashworth, J. 1022n  
 Johnson, P., *see* Darnell, A.C. 1021  
 Johnson, R.L., *see* Belinfante, A. 556  
 Johnson, S. 202  
 Johnson, S., *see* Green, R. 465n  
 Johnson, W.R. 255n, 296, 700, 790  
 Jones, A. 1280  
 Jones, C. 939, 940  
 Jones, D. 312, 313, 315, 316, 318, 323, 325

- Jones, E.L. 29, 372n, 392  
 Jones, S.R.G. 391  
 Jones, W.R. 313  
 Jones-Lee, M., *see* Bateman, I.J. 1077n  
 Jordan, W.K. 1301n, 1317  
 Jorgenson, D., *see* Boskin, M. 952n, 959  
 Jovanovic, B. 790  
 Jullien, B., *see* Caillaud, B. 571n, 580, 580n, 594n, 602  
 Jullien, B., *see* Crampes, C. 587, 607n  
  
 Kaali-Nagy, C. 449n, 460  
 Kadushin, C., *see* Coser, L.A. 739, 784n  
 Kafka, P. 673n  
 Kahin, B. 290n  
 Kahn, C., *see* Engelbrecht-Wiggans, R. 939, 940  
 Kahn, J.R., *see* Clark, D.E. 1023, 1066n, 1076n, 1093n  
 Kahneman, D. 689, 933n  
 Kaiser, U. 605  
 Kaitatzi-Whitlock, S. 599n  
 Kaldor, N. 1088  
 Kalleberg, A. 801n  
 Kalleberg, A., *see* Appelbaum, E. 804  
 Kane, D. 460  
 Kant, I. 181  
 Kanter, R.M. 774, 801n  
 Kaplan, P.H.D. 315  
 Karl, B. 1302n, 1316  
 Karpinski, C. 261n  
 Karraker, D., *see* Schuster, J.M. 1268n  
 Karttunen, S. 816  
 Katz, E., *see* Handy, F. 484  
 Katz, M. 1301n  
 Katz, R.S., *see* Cummings Jr., M.C. 742  
 Katz, S. 1303, 1308, 1313, 1316  
 Katz, S., *see* Ilchman, W. 1315  
 Katz, S., *see* Karl, B. 1302n, 1316  
 Kay, A., *see* Alper, N. 769, 771n, 774, 795  
 Kealy, E. 796n  
 Keegan, C., *see* Orend, R.J. 439  
 Keeter, S., *see* Kohut, A. 393  
 Kendall, E. 1237n  
 Keren, M. 490n  
 Kern, R.M., *see* Peterson, R.A. 419, 420, 422, 441, 466n  
 Kerns, P.M., *see* Colonna, C.M. 492n  
 Kesenne, S. 493  
 Keser, C. 939, 940  
 Keynes, J.M. 61, 62, 64  
  
 Khakee, A. 438  
 Kieshnick, B.D., *see* Wildman, S.S. 575n  
 Kiger, J. 1311  
 Killingsworth, M. 791  
 Kind, H.J., *see* Barros, P.P. 581n  
 Kindermann, C. 1019n  
 Kindleberger, C. 1055  
 King, A.E., *see* Kushner, R.J. 482  
 King, K. 1261n  
 King, S. 219  
 Kingma, B.R. 499, 1284, 1284n  
 Kingsbury, H. 536n  
 Kingston, P. 829, 829n  
 Kinkead, D. 97n  
 Kirby, S.N., *see* Besen, S.M. 296, 699n, 700  
 Kirchberg, V. 427n, 1011  
 Kirchgässner, G. 1020n  
 Kirchgässner, G., *see* Pommerehne, W. 418n, 423–426, 431, 432n, 449, 1073  
 Klaic, D. 498  
 Klamer, A. 152, 153, 997n, 1209  
 Klapisch-Zuber, C., *see* Herlihy, D. 115  
 Klein, B. 296, 708, 709n  
 Klein, D.C. 740n  
 Klein, H.J. 1039  
 Kleingartner, A., *see* Paul, A. 797  
 Kleit, A.N. 699–701  
 Klemperer, P. 929n  
 Kling, J., *see* Cardell, N. 973  
 Klopffleisch, R., *see* Rürup, B. 743n  
 Kluckhohn, C., *see* Kroeber, A.L. 150  
 Knetsch, J., *see* Kahneman, D. 689  
 Knight, F.H. 380  
 Knight, J.R., *see* Carter Hill, R. 918  
 Knott, C. 775  
 Koestler, A. 256n, 276n  
 Kohut, A. 393  
 Kolb, B.M. 429n, 460, 495  
 Kolstad, C.D., *see* Braden, J.B. 996n, 1077n  
 Kominski, R. 832n  
 Konrad, K.A., *see* Glazer, A. 1031  
 Koopmann, G., *see* Guerrieri, P. 1177n  
 Korevaar, G., *see* Romein, E. 93n  
 Korski, T. 326  
 Kosmopoulou, G., *see* Deltas, G. 939, 940  
 Köster, P.R. 427n  
 Kotkin, J. 1067  
 Kraakman, R., *see* Hansmann, H. 543n  
 Kramer, M., *see* Porter, M. 1307, 1311  
 Krasilovsky, M.W. 697, 698, 699n, 700, 706

- Krebs, S. 424, 425, 442, 442n, 453, 456, 1073n, 1192  
 Kremer, M. 537, 781  
 Kreps, D., *see* Baron, J. 801, 804  
 Krider, R. 620n  
 Kris, E. 775, 776  
 Kroeber, A.L. 150  
 Krowitz, E. 1139n  
 Krueger, A.B. 676, 678–680, 682–686, 688, 691, 696, 696n, 786, 1082n  
 Krueger, A.B., *see* Angrist, J. 842n  
 Krugman, P., *see* Fujita, M. 1087n  
 Kuan, J. 482  
 Kubersky-Piredda, S. 74n, 75, 75n, 76n  
 Kuh, E. 439n  
 Kukies, J., *see* Scherer, F.M. 619n  
 Kurabayashi, Y. 420  
 Kurz, O., *see* Kris, E. 775, 776  
 Kushner, R.J. 482, 495  
 Kushner, R.J., *see* Brooks, A.C. 482, 1088n  
 Kuti, É., *see* Vajda, Á 1270, 1280  
 Kuhl, S., *see* Hansen, C. 583
- LaCasse, C., *see* Horstmann, I. 938  
 Laclotte, M. 259  
 Laddie, J. 270n  
 Lafont-Couturier, H. 260n  
 LaFranco, R. 671  
 Lagemann, E. 1304n  
 Lahr, M., *see* Listokin, D. 1240, 1245, 1246  
 Laibson, D., *see* Glaeser, E. 886  
 Lancaster, K. 108n, 297n, 786, 789n, 957, 960, 1076n, 1198  
 Landes, D.S. 361  
 Landes, W.M. 213n, 214n, 224n, 234n, 237n, 292n, 739, 888n, 890, 923, 974, 1111  
 Landes, W.M., *see* Becker, G.S. 216, 240n  
 Landry, C. 1063, 1069, 1090n  
 Lange, M. 423, 425, 427, 430, 437n, 465, 489, 1025n  
 Lange, M., *see* Luksetich, W.A. 423, 425, 431, 433, 437n, 438, 439n, 453, 456, 463n, 483, 495  
 Lanza, A., *see* Brau, R. 1005  
 Laplante, B. 779  
 Laplante, B., *see* Debeauvais, R. 778  
 Larsen, J.K., *see* Rogers, E.M. 548  
 Latcovich, S. 732, 734  
 Lauderdale Earl of, J.M. 45  
 Laussel, D., *see* Ferrando, J. 580, 602n  
 Laussel, D., *see* Gabszewicz, J.J. 573n, 583, 587, 596, 597, 597n, 601n  
 Lavand’homme, P. 973  
 Lawall, S.N. 1163n  
 Lawrence, D.H. 201  
 Lawrence, S., *see* Renz, L. 1312  
 Lawson, R., *see* Gwartney, J. 393  
 Layard, R. 364  
 Lazarus, P.N. 551  
 Lazear, E.P. 781  
 Lazzaro, E. 263, 264, 961n  
 Leacock, S. 364n  
 Leavis, F.R. 204  
 Lebrun, D. 973  
 Leclair, M.S. 1012n  
 Lee, C. 621, 638, 639  
 Lee, C., *see* De Vany, A. 620, 644, 645, 656, 657  
 Lee, J.-W., *see* Barro, R.J. 753  
 Lee, M., *see* Zuzanek, J. 452  
 Leggat, R. 287n  
 Legros, P. 289n, 295n, 296, 305n  
 Legros, P., *see* Ashenfelter, O. 930  
 Legros, P., *see* Ginsburgh, V. 930  
 Lehman, E., *see* Allmendinger, J. 771n, 779  
 Leibenstein, H. 898n  
 Leichenko, R.M., *see* Coulsen, N.E. 1246  
 Lele, S., *see* Putler, D.S. 423n  
 Lenman, R. 782, 783  
 Lennon, K.H., *see* Dacey, J.S. 511  
 Lerner, A.V., *see* Klein, B. 296, 708, 709n  
 Lerner, J. 293, 294  
 Lessig, L. 295n  
 Lessing, A. 258n  
 Letts, C. 1311  
 Levental, I. 459n  
 Levin, K. 221  
 Levine, D.K., *see* Boldrin, M. 290, 292n, 296, 300, 708  
 Levinsohn, J., *see* Berry, S. 958  
 Levmore, S. 237  
 Levy, A., *see* Corning, J. 417, 423, 425, 427, 432, 449, 453, 457  
 Levy, B. 1004n  
 Lévy-Garboua, L. 417, 418, 424, 425, 427, 427n, 430n, 442n, 443, 444, 444n, 445, 446, 447n, 448, 449, 451, 460, 465, 466, 466n, 467, 1073n, 1074n  
 Lewis, G. 440, 441, 461, 462, 462n, 463, 465, 1074n  
 Lewis, J. 18n

- Lichtman, D. 213  
 Lichtman, D., *see* Bakos, Y. 296  
 Liebowitz, S.J. 291, 296, 704n, 707n, 708, 708n, 709–713, 714n  
 Lilien, D., *see* Engle, R. 965  
 Linder, S. 493  
 Lipsey, R.G., *see* Eaton, B.C. 573n  
 List, J.A. 1028n  
 Listokin, B., *see* Listokin, D. 1240, 1245, 1246  
 Listokin, D. 1240, 1245, 1246  
 Litman, B. 620  
 Litvak, I.A. 1173n  
 Litwak, M. 647  
 Lobstein, D. 271  
 Locatelli Biey, M. 950, 973, 974  
 Longinus, 175n  
 Longley, R., *see* Scanlon, R. 1061  
 Longo, A., *see* Alberini, A. 1077  
 Loomes, G., *see* Bateman, I.J. 1077n  
 Lopes, P.D. 557  
 Lopéz Sintas, J. 423n  
 Lord, B. 1032n  
 Lorizzo, L. 102n, 103n, 116  
 Loughman, J. 13  
 Louviere, J., *see* Adamowicz, V. 1075n  
 Louviere, J., *see* Boxall, P. 1076n  
 Lovallo, D., *see* Camerer, C. 785, 793  
 Lowell, J., *see* McCarthy, K. 480, 491, 492, 496, 1239  
 Lubart, T.I. 511, 513  
 Lubart, T.I., *see* Sternberg, R.J. 513  
 Lubove, R. 1302n  
 Lucas, R.E. 364  
 Lucifora, C. 905  
 Luck, J. 1313  
 Lucking-Reiley, D., *see* List, J.A. 1028n  
 Luhmann, N. 193  
 Luksetich, W.A. 423, 425, 431, 433, 437n, 438, 439n, 453, 456, 463n, 483, 495, 1021, 1022, 1028  
 Luksetich, W.A., *see* Hughes, P.N. 1284  
 Luksetich, W.A., *see* Lange, M. 423, 425, 427, 430, 437n, 465, 489, 1025n  
 Lusht, K. 939, 940  
 Lyall, K., *see* Cwi, D. 1057, 1059  
 Lydecker, J.K. 76n, 77n  
 Lyk-Jensen, P., *see* Christoffersen, H. 1059  
 Lynch, J. 96n  
  
 MacDonald, G. 790, 815, 881, 898  
 MacDonald, M. 202  
  
 Mace, M.-A. 512  
 Macfie, A.L. 361  
 MacKenzie, I. 157  
 Mackerras, C. 482  
 MacKinnon, D.W. 514  
 Madden, C., *see* Rengers, M. 842  
 Maddison, D. 1019n, 1024, 1030, 1196, 1199  
 Madow, M. 900  
 Magurn, R.S. 1151  
 Mahajan, V., *see* Givon, M. 292n  
 Mairesse, F. 1026n  
 Mairesse, F., *see* Ginsburgh, V. 1019n  
 Maison, K.E. 258n  
 Malthus, T.R. 149  
 Mandelbrot, B. 621, 639  
 Mandeville, B. 30–32, 175, 960  
 Mandler, G. 160n  
 Mango, C. 313  
 Mann, B., *see* Tomlins, C. 1301  
 Marburger, D.R. 429n, 1034n  
 Marco-Serrano, F., *see* Köster, P.R. 427n  
 Mark, J.H. 964, 967  
 Marquis, A. 1232, 1309  
 Marshall, A. 53, 54, 184, 695, 777, 1104, 1105  
 Martell, C. 1256, 1258  
 Martens, D. 259–261, 271  
 Martens, M.P.J. 82n, 85, 87n, 98n, 115  
 Martin, F. 1019n, 1023n, 1077n  
 Martin, S.A., *see* Hood, R.S. 1280  
 Martinsson, P., *see* Carlsson, F. 1076n  
 Marvel, H.P. 747  
 Marvel, H.P., *see* Deneckere, R. 747  
 Maskin, E. 1210  
 Maskin, E., *see* Bessen, J. 293  
 Mason, C. 1162n  
 Mason, W., *see* Griliches, Z. 842n  
 Masson, R. 589n  
 Mathews, G. 1167  
 Mathewson, F. 747  
 Mathijs, E., *see* Mosselmans, B. 55  
 Matthew, L.C. 75n, 99n  
 Matthew, L.C., *see* Fantoni, M. 74n  
 Matthews-Grieco, S.F., *see* Fantoni, M. 74n  
 Maule, C.J., *see* Acheson, K. 988n, 1154n, 1155n, 1157n, 1158n, 1173n–1175n  
 Maule, C.J., *see* Litvak, I.A. 1173n  
 Mayer, C., *see* Genesove, D. 933  
 Mazza, I. 1002n, 1005, 1007, 1011  
 McAfee, R.P. 937–940  
 McAndrew, C. 923, 932, 951, 1258  
 McAndrew, C., *see* O'Hagan, J. 1258

- McCafferty, S., *see* Marvel, H.P. 747
- McCain, R.A. 151n, 153n, 160, 160n, 162, 162n, 163, 164, 164n, 447n, 448n, 514
- McCarthy, K. 480, 491, 492, 496, 998n, 1239
- McCartney, B.T. 227
- McCaughey, C. 439
- McCleary, R.M., *see* Barro, R. 393n
- McClelland, D.C. 392
- McClelland, R., *see* Kingma, B.R. 1284
- McConnell, K., *see* Cropper, M. 959
- McCulloch, J.H. 640
- McCulloch, J.R. 41
- McCulloch, R., *see* Wildman, S.S. 575n
- McDaniel, P.R. 1275n
- McDougal, D. 547
- McFadyen, S., *see* Hoskins, C. 1156
- McGuirk, A., *see* Barrett, K.S. 1273n
- McKee, M., *see* West, E.G. 447n
- McKensie, J. 621, 638
- McKinley, T. 994
- McMillan, H., *see* De Vany, A. 620, 636
- McNertney, E. 829n, 873n
- McNertney, E., *see* Waits, C.R. 832
- McPherson, G., *see* Bailey, S.J. 1021
- McVeigh, S. 129n, 139
- Meade, N.F., *see* Carson, R.T. 1078
- Meese, R. 967, 968
- Mei, J. 916, 918, 918n, 919, 920, 922, 922n, 924, 925, 927, 928, 932, 934, 949, 950, 951n, 966, 974
- Mei, J., *see* Ginsburgh, V. 918
- Meier, S. 1019n, 1029n
- Meiland, J.W. 271
- Melot, M. 257n, 263, 276n, 277, 278, 278n
- Melser, D. 959
- Menezes, F., *see* Jones, C. 939, 940
- Menger, C. 53, 57, 184
- Menger, P.-M. 773–775, 778, 779, 780n, 784, 785n, 791, 795n, 796, 798–800, 836n, 902
- Menger, P.-M., *see* Debeauvais, R. 778
- Menger, P.-M., *see* Ginsburgh, V. 1013, 1014, 1019n
- Menger, P.-M., *see* Gurgand, M. 880
- Mercer, C. 1069n, 1093n
- Merryman, J.H. 273, 273n, 276, 276n, 278
- Mertens, S., *see* Ginsburgh, V. 962, 974
- Merton, R.K., *see* Kingston, P. 829, 829n
- Messerlin, P. 1169
- Meyer, J., *see* Kuh, E. 439n
- Meyer, K. 1226n, 1260n
- Meyer, L.B. 279
- Mezias, J.M. 557
- Mezias, S.J., *see* Mezias, J.M. 557
- Michael, R.T. 440
- Michel, N.J. 711, 714
- Michel, P., *see* Ginsburgh, V. 365n
- Micklewright, J., *see* Atkinson, A.B. 798n
- Mikesell, J.L. 1268n
- Milgrom, P. 105n, 585, 585n, 932, 934, 937–940
- Milgrom, P., *see* Grief, A. 71n
- Mill, J.S. 46–48
- Millen, R.F. 1162
- Miller, D. 551
- Miller, L.J. 555n
- Miller, R.A. 790
- Miller, T., *see* Lewis, J. 18n
- Mills, D., *see* Throsby, D. 817, 824
- Mincer, J. 868, 872
- Ministère de la Culture et de la Communication, 1290
- Ministry of Finance. 1289n
- Mireur, H. 265, 281
- Mirowski, P. 183, 188
- Mirus, R., *see* Hoskins, C. 1156
- Missner, M. 901
- Mitchell, C. 1059
- Mitchell, R.C. 797, 1077n, 1078n
- Mitchell, R.C., *see* Carson, R.T. 1199
- Mobilia, P., *see* Carson, A. 425, 432n, 442, 451
- Moggridge, D.E. 61, 65
- Monballieu, A. 89n, 116
- Moncrieff, E. 154
- Montecuccoli degli Erri, F. 101n, 116
- Montgomery, S. 452, 466n, 816, 819, 841, 842
- Montias, J.M. 75n, 77n, 80n, 85, 88n, 92n–95n, 99n, 104n, 116, 260n, 261, 771n, 783n, 1019n, 1030
- Montias, J.M., *see* Loughman, J. 13
- Montmarquette, C., *see* Lévy-Garboua, L. 417, 418, 424, 425, 427, 427n, 430n, 442n, 443, 444, 444n, 445, 446, 447n, 448, 449, 451, 460, 465, 466, 466n, 467, 1073n, 1074n
- Moore, G. 51, 191
- Moore, J., *see* Hart, O. 541
- Moore, J.V. 139
- Moore, T.G. 421n, 423, 425, 426, 432n, 433, 435, 436
- Moorhouse, J.C. 996
- Morel, C. 1290
- Morrison, W.G. 441, 460, 1077n
- Mortier, R. 261n



- Moses, M., *see* Ginsburgh, V. 918  
 Moses, M., *see* Mei, J. 916, 918, 918n, 919,  
 920, 922, 922n, 924, 925, 927, 928, 932, 934,  
 949, 950, 951n, 966, 974  
 Mosselmans, B. 55  
 Mossetto, G. 155, 160, 1166  
 Moulin, R. 269n, 775, 784n, 787n, 796, 796n,  
 961, 1020n  
 Moulton, B. 682  
 Moulton, B., *see* Abraham, K. 952n  
 Mourato, S., *see* Bateman, I.J. 1077n  
 Mourato, S., *see* Maddison, D. 1199  
 Mueller, D. 1002n  
 Muirhead, S. 1307  
 Mukhtar, T., *see* DiMaggio, P. 423n  
 Muller, E., *see* Givon, M. 292n  
 Muller, J.M. 279  
 Mund, H. 275  
 Munnely, A. 1256n  
 Murphy, K.M., *see* Becker, G.S. 216, 240n,  
 447n, 448n, 586  
 Murphy, K.M., *see* Klein, B. 296, 708, 709n  
 Murray, J.M. 77n, 79n  
 Musgrave, R.A. 481n, 800, 1126  
 Musnick, I. 600  
 Mussa, M. 582  
 Mussoni, M. 1011, 1012  
 Muth, R.F., *see* Bailey, M. 918n, 954  
 Myerscough, J. 1059n, 1061, 1062, 1062n,  
 1067n  
 Myrdal, G. 401  
  
 Nantel, J., *see* Colbert, F. 419  
 Nash, D. 796  
 National Center for Charitable Statistics, 1309  
 National Committee for Responsive  
 Philanthropy, 1310  
 National Endowment for the Arts, 419, 420,  
 422, 494, 1029, 1228, 1308  
 National Music Publishers' Association, Inc.  
 and The Harry Fox Agency, Inc., 701, 703  
 National Research Center of the Arts, Inc.,  
 419n, 421, 464  
 Natzkoff, P. 939  
 Navarro, P. 1278  
 Navrud, S. 997n, 1077, 1077n, 1247  
 Neary, P. 1087n  
 Needham, G. 319  
 Needle, J.B. 322  
 Negus, K. 516  
 Nelson, K. 497  
  
 Nerlove, M. 958  
 Netzer, D. 488, 800, 1010, 1231, 1234, 1241,  
 1241n, 1242n, 1245, 1259, 1260n, 1265,  
 1281, 1282, 1308  
 Neumayer, E. 989  
 Neven, D.J., *see* Anderson, S.P. 581n  
 Newman, A.F., *see* Legros, P. 289n, 295n  
 Newman, K.S. 367  
 Nicholas, D. 77n, 79n, 86n, 115  
 Nicks, J., *see* Lord, B. 1032n  
 Nietzsche, F. 152, 186, 187  
 Nijkamp, P. 998  
 Nilssen, T. 573n, 607n  
 Nilssen, T., *see* Barros, P.P. 581n  
 Nilsson, G., *see* Khakee, A. 438  
 Nisbett, R. 800  
 Noble, J.V. 1019n  
 Nöldeke, G. 542  
 Noonan, D. 196, 997n, 1023n, 1077n, 1199  
 Nordhaus, W.D. 292, 708  
 Norman, G. 217n, 218n, 219, 220n  
 Norman, V., *see* Dixit, A. 585  
 North, D.C. 71n  
 Nourse, H.O., *see* Bailey, M. 918n, 954  
 Novos, I.E. 296  
 Nozick, R. 149n, 151, 152, 1150  
 Nuttall, P. 75, 80n, 82n  
 Nygren, T.I., *see* Bowen, W.G. 562  
  
 Oates, W.E., *see* Baumol, W.J. 491  
 Oberholzer-Gee, F. 688n, 707, 713, 1034n  
 Oberholzer-Gee, F., *see* Frey, B.S. 997n, 1247  
 O'Brien, G.E. 365n  
 O'Brien, J. 769, 770  
 Observatoire de l'emploi culturel, 769  
 Ochse, A. 511  
 OCW, 1183n  
 Odendahl, T. 561n  
 Oehser, P. 1308  
 Office of Communications, 1132n  
 O'Hagan, J. 419, 481, 495, 737, 742, 1012n,  
 1019n, 1021n, 1032, 1033, 1033n, 1039n,  
 1060, 1072n, 1083n, 1166, 1207n, 1258,  
 1265, 1274  
 O'Hagan, J., *see* McAndrew, C. 1258  
 O'Halloran, S., *see* Epstein, D. 1211  
 O'Hare, M. 1033, 1231n, 1239n  
 O'Hare, M., *see* Borgonovi, F. 1244, 1290n  
 O'Hare, M., *see* Feld, A. 1241n, 1258, 1260,  
 1260n, 1261, 1263, 1264n, 1268, 1268n,  
 1269, 1270, 1273, 1274, 1281, 1286n, 1293

- Okun, A.M. 768  
Olson, M. 361  
Olson, M., *see* Keser, C. 939, 940  
O'Malley, M. 72n, 75  
Ondaatje, E.H., *see* McCarthy, K. 998n  
Orban de Xivry, B., *see* Ginsburgh, V. 973  
Ordovery, J. 296, 296n, 559n  
Ordovery, J., *see* Baumol, W.J. 352n  
Orend, R.J. 439  
Ornstein, S.I. 743n  
Osbert, L.S., *see* Hood, R.S. 1280  
Osborne, H. 204  
Osbourne, S. 675  
Oster, S. 1022, 1031n  
Ostrower, F. 561n, 562  
Ostrower, F., *see* DiMaggio, P. 461, 461n, 462, 463  
O'Sullivan, A. 1089  
Oteri, M. 425, 442  
Ottaviano, G. 736, 1068  
Owen, B., *see* Spence, M. 583, 1208  
Owen, D. 1301n  
Owen, J.D. 434  
Özdemiroglu, E., *see* Bateman, I.J. 1077n
- Paarsch, H., *see* Donald, S. 939, 940  
Pacque, K.-H. 1280  
Pakes, A. 958, 958n  
Pakes, A., *see* Berry, S. 958  
Palma, A., *see* Anderson, S.P. 609n  
Palmquist, R.B. 958, 958n, 963  
Papadopoulo, A. 316  
Paradeise, C. 795n  
Páramo, J.P. 96n  
Pareles, J. 688  
Pareles, J., *see* George-Warren, H. 677n  
Pareto, V. 187  
Parker, G., *see* Wilson, C. 115, 116  
Parkinson, M., *see* Bianchini, F. 1093n  
Partridge, M.D., *see* Luksetich, W.A. 1021, 1022, 1028  
Pasquali, F. 1010  
Passman, D.S. 543n, 697, 698, 699n, 1159n  
Pastoureau, M. 271  
Pater, W. 201  
Paton, D. 431n  
Pattanaik, P.K. 994  
Pattullo, P. 407  
Paul, A. 797  
Paul, J.C.N. 318  
Payne, A.A., *see* Andeoni, J. 1244
- Peacock, A. 19, 339n, 484, 493, 784n, 797, 985n, 988, 1001, 1003, 1004, 1014, 1015, 1019n, 1024n, 1064n, 1125–1127, 1131n, 1132n, 1134n, 1138n, 1186, 1192  
Peacock, T.L. 48  
Pearce, D.W. 989, 990, 990n, 994  
Pearce, M. 157–159  
Pearce, P.W., *see* Bateman, I.J. 1077n  
Peck, J., *see* Deneckere, R. 747  
Peeters, N., *see* Martens, M.P.J. 85, 87n, 115  
Peitz, M. 290n, 581n, 593n, 605, 709, 711, 712, 714, 715  
Penders, F., *see* Ginsburgh, V. 974  
Peng, L., *see* Goetzmann, W.N. 950, 955n, 963  
Penhallurick, L. 1287  
Peri, G., *see* Ottaviano, G. 1068  
Perino, G. 1082, 1082n  
Peronnet, B. 265, 281  
Perotti, R., *see* Alesina, A. 1192  
Pesando, J. 916, 918, 918n, 920, 921, 923–928, 938–940, 950, 966, 973, 974, 974n  
Pesando, J., *see* Hosios, A. 918n  
Pestelli, G., *see* Bianconi, L. 130n  
Peterson, R.A. 419, 420, 422, 423n, 441, 466n, 557  
Petry, G., *see* Cardell, N. 973  
Petry, G., *see* Thiel, S. 939, 940  
Pezanis-Christou, P. 939, 940  
Pezzolo, L. 115  
Pflieger, S. 973  
Phelps, G. 323, 325  
Philbrick, N. 1226n  
Pick, J., *see* Protherough, R. 372  
Pickering, M., *see* Negus, K. 516  
Pierce, R.K., *see* Agnello, R.J. 973  
Pigliaru, F., *see* Brau, R. 1005  
Pignataro, G. 1001  
Piketty, T. 684  
Pindyck, R.S., *see* Dixit, A. 1146n  
Pine, B.J. 1075n, 1086, 1200  
Piore, M. 1105  
Pitman, J. 326  
Plant, A. 739  
Plato, 171, 311  
Plaza, B. 1037n, 1063  
Plug, E., *see* Rengers, M. 1197  
Png, I., *see* Hui, K. 290n, 709n, 710n  
Poggi, J. 547  
Pohlmann, H. 137  
Pokorny, M., *see* Sedgwick, J. 1154n, 1160n  
Polanyi, M. 1107

- Pollak, R.A. 442, 442n  
 Pollakowski, H., *see* Case, K.E. 968n  
 Pollakowski, H., *see* Halvorsen, R. 953n, 960n  
 Pommer, E., *see* Goudriaan, R. 487  
 Pommerehne, W. 418n, 423–426, 431, 432n, 449, 924, 974, 1032n, 1073, 1078  
 Pommerehne, W., *see* Frey, B.S. 339n, 770, 787, 918, 920, 950, 957, 974, 1019n, 1023, 1029n, 1035n, 1036n, 1078, 1079n, 1198n, 1242, 1270  
 Pommerehne, W., *see* Krebs, S. 424, 425, 442, 442n, 453, 456, 1073n, 1192  
 Pommerehne, W., *see* Schneider, F. 974  
 Pontier, J.M., *see* Bourdon, J. 277n  
 Poor, J. 1076n  
 Poor, P.J. 1248  
 Port Authority of New York and New Jersey, 1238  
 Porter, M. 1307, 1311  
 Portney, P.R. 1078n  
 Posner, R.A., *see* Chiang, T.J. 213n  
 Posner, R.A., *see* Landes, W.M. 213n, 214n, 224n, 234n, 237n, 292n, 739, 890, 1111  
 Posnett, J., *see* Jones, A. 1280  
 Powell, W., *see* DiMaggio, P. 1310  
 Powell, W.W. 548, 772, 802  
 Powell, W.W., *see* Coser, L.A. 739, 784n  
 Powers, K., *see* Kafka, P. 673n  
 Pradhan, M., *see* van den Berg, G. 939, 940  
 Preece, S.B., *see* Fisher, C.G. 423n  
 Previer, W. 115  
 Prieto Rodríguez, J. 452, 1029n, 1198  
 Prieto-Rodríguez, J., *see* Fernández-Blanco, V. 452  
 Prosch, H., *see* Polanyi, M. 1107  
 Protherough, R. 372  
 Pryor, F.L. 367  
 Purdy, M., *see* O'Hagan, J. 481, 495  
 Putler, D.S. 423n  
 Putnam, R. 368, 885  
 Puttfarcken, T. 960  
 Pyke, F. 1105  
  
 Quah, D. 290, 290n, 296n, 297  
 Quan, D., *see* McAfee, R.P. 937  
 Queau, P. 280  
 Queen, E., *see* Ilchman, W. 1315  
 Quigley, J. 784, 964  
 Quigley, J., *see* Case, K.E. 918, 963, 964  
  
 Radford, M. 156  
 Radich, A.J. 1057n  
  
 Rae, J. 47  
 Rafool, M. 1291  
 Raiff, M.E., *see* De Marchi, N. 90n  
 Rajadhyaksha, A. 1164  
 Ram, M., *see* Masson, R. 589n  
 Ramsdale, P. 1177  
 Randhawa, M.S., *see* Galbraith, J.K. 60  
 Ranfft, X., *see* Hughes, A. 271  
 Rannou, J. 775  
 Rannou, J., *see* Debeauvais, R. 778  
 Raux, S. 93n, 104n  
 Ravid, S.A. 653, 658  
 Raymond, S. 1307  
 Raynor, H. 128  
 Ready, R.C., *see* Navrud, S. 997n, 1077n, 1247  
 Reimer, E. 599n  
 Reisman, D. 364  
 Reitlinger, G. 950, 967  
 Relish, M. 439  
 Rengers, M. 816, 817, 825, 842, 883n, 1197  
 Renz, L. 1306, 1312  
 Ressler, R., *see* Ekelund, R. 950, 974  
 Reynolds, R., *see* Masson, R. 589n  
 Ribar, D.C. 1028n  
 Ricardo, D. 46, 47, 180  
 Ricci, J.C., *see* Bourdon, J. 277n  
 Richards, R.D. 322  
 Richardson, J. 815, 818, 819  
 Ricketts, M. 1127n  
 Riganti, P., *see* Alberini, A. 1077  
 Riley, G.B. 312  
 Riley II, C., *see* Schuster, J.M. 985n, 1265  
 Rimel, R. 1311  
 Rinaldi, S. 515  
 Ringstad, V. 743n  
 Ritenour, S., *see* Ekelund Jr., R.B. 425, 436, 437, 450  
 Rizzo, I. 1001, 1003, 1005, 1008, 1010  
 Rizzo, I., *see* Giardina, E. 1003  
 Rizzo, I., *see* Hutter, M. 985n, 1014, 1015  
 Rizzo, I., *see* Mazza, I. 1005, 1007  
 Rizzo, I., *see* Peacock, A. 985n, 1014, 1015, 1019n  
 Rizzo, I., *see* Pignataro, G. 1001  
 Robbins, L. 58, 150, 1019n, 1186  
 Robert, J., *see* Donald, S. 939, 940  
 Roberts, B. 391n  
 Roberts, C., *see* Wall, G. 1024n  
 Roberts, J., *see* Milgrom, P. 585, 585n  
 Robertson, R. 1150n

- Robins, K., *see* Aksoy, A. 773  
 Robinson, M., *see* Montgomery, S. 452, 466n, 816, 819, 841, 842  
 Robinson, W. 228n  
 Rochet, J.-C. 580  
 Rockefeller Foundation, 1312  
 Rockefeller Panel Report, 1259n  
 Rodríguez, P., *see* García, J. 428n, 429n  
 Rogers, E.M. 548  
 Rogerson, R. 575n  
 Rogoff, K. 1211  
 Roharik, I., *see* Rannou, J. 775  
 Romano, R.E., *see* Blair, R.D. 599  
 Romanowski, P., *see* George-Warren, H. 677n  
 Romein, E. 93n  
 Romer, P. 293n, 708, 801  
 Ronte, D., *see* Bonus, H. 899n, 961  
 Rose, A., *see* Schulze, G. 1071n, 1074n, 1083n, 1084, 1087, 1192  
 Rose-Ackerman, S. 476, 495, 496, 496n, 1026n  
 Rosen, H., *see* Brown, J. 958  
 Rosen, S. 10, 141n, 515, 558, 652, 675, 678, 687, 695, 736, 777, 780, 784, 790, 835, 838, 868, 869, 872, 873, 876, 877, 897, 897n, 899, 902, 957–959, 959n, 961, 975, 978, 1036, 1076n, 1204n  
 Rosen, S., *see* Lazear, E.P. 781  
 Rosen, S., *see* Mussa, M. 582  
 Rosenberg, B. 549  
 Rosenberg, N. 392  
 Rosenfield, A.M., *see* Rosen, S. 675, 678, 687  
 Rosett, R.N. 1025n, 1029n  
 Ross, L., *see* Nisbett, R. 800  
 Ross, M.H. 919, 920, 922, 973  
 Rosse, J.N. 600, 605  
 Rosselli, J. 130n  
 Rothenberg, A. 512n  
 Rothenberg, J. 572n  
 Rothfield, L., *see* Schuster, J.M. 1268n  
 Rothschild, M.L., *see* Brown, T.J. 599  
 Rouet, F. 1258  
 Rubinfeld, D.L. 1213, 1214n  
 Rubinstein, A.J. 485, 494  
 Rubinstein, R. 217  
 Rumelt, R., *see* Conner, K. 292n  
 Runco, M. 511  
 Runco, M., *see* Lubart, T.I. 513  
 Rürup, B. 743n  
 Rushton, M. 1146, 1233  
 Ruskin, J. 51  
 Russell, F. 1285n  
 Ryan, M. 1076n  
 Ryan, W., *see* Letts, C. 1311  
 Ryans, A.B. 429n, 460  
 Rysman, M. 605  
 Sabel, C., *see* Bagnasco, A. 1105  
 Sabel, C., *see* Piore, M. 1105  
 Sacerdote, B., *see* Glaeser, E. 886  
 Sadie, S. 130  
 Saez, E., *see* Piketty, T. 684  
 Sagot-Duvaouroux, D. 960  
 Sagot-Duvaouroux, D., *see* Pflieger, S. 973  
 Sahuguet, N., *see* Ashenfelter, O. 930  
 Sahuguet, N., *see* Ginsburgh, V. 930  
 Sainath, P. 405  
 Salais, R. 1106  
 Salamon, L.M. 475–478, 484, 493, 1314, 1317  
 Salem, M. 483n  
 Saloner, G., *see* Ordover, J.A. 559n  
 Salop, S. 586, 608, 746n  
 Salop, S., *see* Besen, S.M. 699n, 700  
 Salop, S., *see* Ordover, J.A. 559n  
 Samorodnitsky, G. 662  
 Samuelson, P. 569n, 1129  
 Sanchez-Capdequi, C. 510  
 Sanders, G.L., *see* Gopal, R.D. 710n, 714  
 Sandler, T., *see* Cornes, R. 481  
 Sanjek, D., *see* Sanjek, R. 547  
 Sanjek, R. 547  
 Santagata, W. 1037n, 1077, 1103n, 1106  
 Santagata, W., *see* Cellini, R. 1113, 1114n  
 Santagata, W., *see* Cuccia, T. 1110, 1113, 1114n  
 Santagata, W., *see* Jalil Moreno, Y. 1106  
 Santilli, M., *see* Hansmann, H. 227  
 Santos, F.P. 816, 832, 835, 838  
 Sartre, J.P. 791  
 Savile, A. 203  
 Scanlon, R. 1061  
 Schaer, R. 274  
 Schechter, R., *see* Waterman, D. 452  
 Scheff, J. 429n  
 Scherer, F.M. 130, 138n, 139, 357n, 619n  
 Schervish, P. 1314  
 Schiff, J. 489, 1029n  
 Schiller, F. 198  
 Schiller, H. 1150n  
 Schimmel, D., *see* Cahn, S. 1170, 1173  
 Schimmelpfennig, J. 423, 425, 426, 431, 464, 465  
 Schleifer, A., *see* Hart, O. 1210

- Schlicht, E. 392
- Schmidt, K.M., *see* Nöldeke, G. 542
- Schmitz, M.D., *see* Abrams, B.A. 1243
- Schnapper, A. 85, 91n, 116
- Schnapper, B., *see* Chamoux, F. 256n, 260, 262, 262n, 273
- Schneider, F. 974
- Schneider, F., *see* Pommerehne, W. 1078
- Schoeffer, S. 391n
- Schoenfeld, S. 924, 925
- Schoenhaus, T., *see* Stern, S.L. 556
- Schopenhauer, A. 186, 287
- Schultz, T. 868
- Schulz, B. 1154n
- Schulze, G. 903, 904, 1023, 1071n, 1074, 1074n, 1078, 1078n, 1079n, 1083, 1083n, 1084, 1087, 1087n, 1126n, 1136n, 1146n, 1178, 1192, 1212
- Schulze, G., *see* Perino, G. 1082, 1082n
- Schum, P., *see* Pesando, J. 950, 973, 974, 974n
- Schumpeter, J.A. 177, 513, 878
- Schuster, J.M. 498, 985n, 1001, 1004, 1019n, 1029n, 1031n, 1065n, 1084n, 1234, 1257, 1257n, 1258, 1263–1265, 1268n, 1269, 1270n, 1275n, 1279n, 1280, 1280n, 1286, 1286n
- Schuster, J.M., *see* Feld, A. 1241n, 1258, 1260, 1260n, 1261, 1263, 1264n, 1268, 1268n, 1269, 1270, 1273, 1274, 1281, 1286n, 1293
- Schwartz, G. 240
- Schwartz, H. 260, 260n, 272, 272n
- Schwartz, M.L., *see* Paul, J.C.N. 318
- Schwarz, S. 485
- Schwed, N., *see* Ginsburgh, V. 973
- Schwer, R.K., *see* Gazel, R. 1060
- Scitovsky, T. 163, 194, 375
- Scoones, D., *see* Bernhardt, D. 939, 940
- Scorcu, A.E., *see* Candela, G. 950, 965, 973, 974
- Scotchmer, S. 292n, 293n, 1213
- Scotchmer, S., *see* Green, J.R. 293n
- Scott, A. 1164
- Scott, A., *see* Peacock, A.T. 1192
- Scott, A.J. 784, 1089, 1103, 1109, 1164
- Scott, J.A. 1088n
- SCP, 1194, 1198
- Scribner Jr., D. 1246
- Scully, G.W. 393
- Seagrave, K. 620
- Seaman, B.A. 418, 420n, 422n, 423n, 426n, 429n, 430, 440, 440n, 444n, 446, 451n, 452, 452n, 460n, 461n, 494, 495, 780n, 876n, 878n, 884, 892, 904, 1024n, 1064, 1064n, 1091, 1225, 1238
- Seaman, B.A., *see* Lewis, G. 440, 441, 461, 462, 462n, 463, 465, 1074n
- Sedgwick, J. 621, 621n, 1154n, 1160n
- Seeber, R.L., *see* Gray, L.S. 797n
- Segal, L.M. 1035
- Seguy Duclot, A. 271
- Seitz, J.A. 516
- Semenik, R.J., *see* Bamossy, G. 460
- Sen, A. 402, 776
- Sénéchal, P. 279n
- Sengenber, W., *see* Pyke, F. 1105
- Shachar, R., *see* Goettler, R. 606
- Shachmut, K.M., *see* Weinberg, C.B. 423n
- Shaffer, G. 555
- Shaftesbury Earl of, 175
- Shaked, A. 582
- Shamsie, J., *see* Miller, D. 551
- Shanahan, J. 161, 486
- Shanet, H. 562n
- Shapiro, C. 553n
- Shapiro, C., *see* Grossman, G. 255n, 584n, 585, 589, 590n
- Shavell, S. 293n
- Shaw, J.E. 101n
- Shea, W.R. 509
- Shemel, S., *see* Krasilovsky, M.W. 697, 698, 699n, 700, 706
- Sherman, A.L. 362
- Shields, M. 577, 577n, 584n
- Shiller, R.J. 955n, 963, 963n
- Shiller, R.J., *see* Case, K.E. 918, 918n, 963, 967, 973
- Shleifer, A., *see* Glaeser, E.L. 1027n
- Shockley, G.E. 986
- Shook, J.R. 151n
- Showell, E., *see* Facchina, B. 1257n
- Shum, P., *see* Pesando, J. 918, 920, 927, 928, 938–940, 950, 973
- Shusterman, R. 171, 176, 198, 201, 204
- Shy, O. 292n
- Shy, O., *see* Gayer, A. 291, 291n, 714
- Siegel, P., *see* Kominski, R. 832n
- Signorello, G., *see* Santagata, W. 1077
- Silver, M. 959
- Simmonds, J. 1257n
- Simmons, R., *see* Lucifora, C. 905
- Simon, H. 156, 621
- Simon, J.G. 1257n, 1267–1269

- Simonton, D.K. 156, 157n, 512, 513, 516  
 Simpson, C. 797  
 Sintas, J.L. 1198  
 Sirmans, C.F., *see* Carter Hill, R. 918  
 Siwek, S.E., *see* Messerlin, P. 1169  
 Siwek, S.E., *see* Wildman, S.S. 558, 1083n  
 Skeath, S., *see* Dixit, A. 517n  
 Skidelsky, R. 63, 188n  
 Slater, R. 547, 551  
 Slichter, J. 674  
 Slivinski, A., *see* Bilodeau, M. 481, 485  
 Slivinski, A., *see* Brown, E. 485  
 Sluijter, E.J. 92n, 94n, 116  
 Smith, A. 37, 38, 40, 45, 178, 179, 255, 867n, 869  
 Smith, H., *see* Latcovich, S. 732, 734  
 Smith, H.D. 478  
 Smith, J., *see* Poor, J. 1076n, 1248  
 Smith, J., *see* Renz, L. 1312  
 Smith, M.S., *see* Moorhouse, J.C. 996  
 Smith, S.R., *see* Schuster, J.M. 1268n  
 Smith, T. 448, 816, 820  
 Smith, V. 767, 801n  
 Smolensky, E. 1027, 1227  
 Snell, M., *see* Collins, A. 621  
 Snooks, G. 841  
 Sobel, J. 885  
 Sokolowski, W., *see* Salamon, L. 1317  
 Solkin, P.A. 176n, 178  
 Solomon, M. 52, 1151  
 Solow, R. 885, 989, 989n  
 Sonnac, N. 600  
 Sonnac, N., *see* Ferrando, J. 580, 602n  
 Sonnac, N., *see* Gabszewicz, J.J. 573n, 583, 587, 596, 597, 597n, 601n  
 Sonnino, E. 116  
 Sontag, S. 1167  
 Soon-Yong, C. 740n  
 Sørgaard, L., *see* Barros, P.P. 581n  
 Sørgaard, L., *see* Nilssen, T. 573n, 607n  
 Sornette, D. 619n, 638  
 Spanier, S. 154  
 Spear, R.E. 103n  
 Spence, A.M. 584n  
 Spence, M. 583, 1208  
 Spiegel, M., *see* Goetzmann, W.N. 915, 915n, 916, 919, 920, 950, 974  
 Spilerman, S. 779  
 Spiller, P.T., *see* Levy, B. 1004n  
 Spinnewyn, F. 447n, 448n  
 Spitznagle, M., *see* De Vany, A. 641  
 Spulber, D.F. 545n  
 Stabel, P. 77n, 79n  
 Stabler, M. 1064n  
 Stacey, S., *see* Katz, S. 1313  
 Stahl, D.O., *see* Soon-Yong, C. 740n  
 Stahl, K., *see* Legros, P. 305n  
 Staines, J. 1257n  
 Stalker, G.M., *see* Burns, T. 803  
 Stanley, D. 1059  
 Stanton, L.L. 275, 275n  
 Stark, C., *see* Cooper, S. 319, 321  
 Stark, R. 375n  
 Starkey, K. 551n  
 Stavins, J. 426n  
 Stebbing, P.D. 157  
 Stegeman, M. 585  
 Stein, J.P. 14, 918, 920, 949, 974  
 Steinberg, R. 483, 499, 501, 1253n, 1273, 1273n, 1277, 1280, 1280n, 1284  
 Steiner, F. 1034  
 Steiner, P.O. 572, 572n, 581, 606  
 Steiner, R. 322  
 Stephan, P. 788n  
 Stern, N.H., *see* Atkinson, A.B. 1201  
 Stern, S.L. 556  
 Sternberg, R.J. 513  
 Stevens, D. 163  
 Stevens, M., *see* Ashenfelter, O. 933–935, 950, 973  
 Stevenson, C.L. 202  
 Stevenson, D., *see* Gibson, L. 1070  
 Stewart, J.B. 929  
 Stigler, G. 162, 188, 447n, 448n, 545, 586, 769, 898, 1074n, 1083n  
 Stiglitz, J.E., *see* Dixit, A. 736  
 Stiglitz, J.E., *see* Salop, S. 746n  
 Stinchcombe, A.L. 772, 779  
 Stock, J., *see* Jackson, C. 791  
 Stohs, J. 816, 820  
 Stoker, T.M., *see* Blundell, R. 426n  
 Stone, J.E., *see* Facchina, B. 1257n  
 Storgaard, S., *see* Bille, T. 1073n  
 Storper, M. 551, 773, 784, 801n, 1107, 1109  
 Storper, M., *see* Christopherson, S. 551, 773, 778, 1109  
 Storper, M., *see* Salais, R. 1106  
 Stowe, D.W. 547  
 Strand, J., *see* Navrud, S. 1077  
 Stratton, P., *see* Hayes, N. 510  
 Streeten, P. 404, 406  
 Strumpf, K., *see* Oberholzer, F. 688n, 707, 713

- Stumpp, H., *see* Rürup, B. 743n  
 Sugden, R., *see* Bateman, I.J. 1077n  
 Sullivan, B., *see* Katz, S.N. 1303  
 Sullivan, M. 317  
 Sumka, H., *see* Wittke, A. 958  
 Sunstein, C.R. 1023n  
 Sunstein, C.R., *see* Holmes, S. 375  
 Suppes, P. 521n  
 Surowiecki, J. 706, 706n  
 Surrey, S. 1243, 1260  
 Sussmann, L. 773, 774  
 Sutton, J. 297, 746n  
 Sutton, J., *see* Shaked, A. 582  
 Swait, J., *see* Boxall, P. 1076n  
 Swansson, J., *see* Bateman, I.J. 1077n  
 Sweeting, A. 605  
 Syson, L. 98n  
 Szanto, M. 91n  
 Szenberg, M. 732n  
  
 Taalas, M. 489  
 Tabarrok, A., *see* Cowen, T. 875, 890  
 Tabassum, A., *see* Jalil Moreno, Y. 1106  
 Tabellini, G., *see* Alesina, A. 1002, 1210, 1211  
 Takeyama, L.N. 292n  
 Taleb, N. 638  
 Taleb, N., *see* De Vany, A. 641  
 Taqq, M.S., *see* Samorodnitsky, G. 662  
 Taylor, B. 777  
 Taylor, C. 177  
 Taylor, D. 1268n  
 Taylor, L.D., *see* Houthakker, H.S. 424, 425,  
 430, 442, 442n, 443, 464  
 Taylor, R. 1156  
 Temin, P. 1032  
 Tempest, S., *see* Starkey, K. 551n  
 Thaler, R. 1040n  
 Thaler, R., *see* Kahneman, D. 689  
 The Art Newspaper, 233, 233n, 234  
 The Economist, 1318  
 Theobald, M.M. 1035n  
 Thiel, S. 939, 940  
 Thisse, J., *see* Anderson, S.P. 609n  
 Thisse, J., *see* d'Aspremont, C. 596  
 Thisse, J., *see* Fujita, M. 1087n  
 Thisse, J., *see* Gabszewicz, J.J. 582, 601  
 Thisse, J., *see* Ottaviano, G.I.P. 736  
 Thisse, J., *see* Shy, O. 292n  
 Thomas, A. 75n, 98n  
 Thomas, B., *see* Johnson, P. 1019n, 1059,  
 1135n  
 Thomas, R.B., *see* Darnell, A.C. 1021  
 Thompson, B., *see* Throsby, D. 817, 824  
 Thompson, N.L. 1032  
 Thompson, R., *see* McAndrew, C. 923, 932,  
 951  
 Thompson, S.E. 317  
 Thomson, K. 1135n  
 Thornton, D., *see* Syson, L. 98n  
 Throsby, D. 5, 153, 153n, 155, 196, 339n, 408,  
 419, 419n, 421n, 423–425, 427n, 428, 429,  
 433n, 434n, 436n, 453–455, 455n, 456–458,  
 465, 483–486, 488, 491, 492, 501, 515, 522,  
 524, 525n, 736n, 737, 769, 775, 776, 791,  
 794, 795n, 796, 816, 817, 824, 825, 829, 834,  
 841, 842, 873–875, 878, 879, 882, 885, 885n,  
 890, 902, 986, 988n, 989n, 998–1000, 1019n,  
 1026n, 1027, 1055, 1056, 1065, 1065n, 1070,  
 1077n, 1107, 1139n, 1159, 1190, 1197–1199,  
 1207, 1238, 1242n, 1243  
 Throsby, D., *see* Baumol, W.J. 775, 779  
 Throsby, D., *see* Bryant, T. 161n  
 Throsby, D., *see* Hutter, M. 8n, 998n  
 Throsby, D., *see* Jeffri, J. 883  
 Thuillier, J. 960  
 Thurston, N. 939, 940  
 Tiebout, C. 1005n, 1214  
 Tiehen, L. 1273  
 Tietzel, M. 743n  
 Tighe, E., *see* Amabile, T. 156, 158, 158n  
 Tirole, J. 687  
 Tirole, J., *see* Dewatripont, M. 289n, 1210  
 Tirole, J., *see* Lerner, J. 294  
 Tirole, J., *see* Maskin, E. 1210  
 Tirole, J., *see* Rochet, J.-C. 580  
 Tobias, J., *see* DiNardo, J. 423n  
 Tobias, S. 438  
 Toepler, S., *see* Anheier, H.K. 1035, 1035n  
 Toepler, S., *see* Zimmer, A. 498, 1249  
 Toffler, A. 1259n  
 Tollison, R.D., *see* Ekelund Jr., R.B. 374  
 Tolstoy, L. 189  
 Tomlins, C. 1301  
 Tomlinson, J. 1150n  
 Topel, R. 798  
 Touchstone, S.K. 423, 425, 450n, 464n, 495  
 Tovey, D.C. 344n  
 Towse, R. 5, 339n, 539, 771, 834n, 842, 871,  
 873, 874n, 876, 876n, 877n, 878–880, 880n,  
 881, 881n, 882, 888–890, 890n, 893, 894,  
 1013–1015, 1019n, 1024n, 1080, 1081n,  
 1086, 1131n, 1139n

- Towse, R., *see* Rizzo, I. 1008  
 Trattner, W. 1301n  
 Trautman, W.B., *see* Dertouzos, J.N. 599n  
 Trienekens, S. 461  
 Trilling, 202  
 Trimarchi, M., *see* Oteri, M. 425, 442  
 Triplett, J. 952, 952n, 953n, 958n, 959, 960n, 963, 964n  
 Trites, R.S. 1160  
 Trompenaars, F., *see* Hampden-Turner, C. 377n  
 Trupiano, G. 1288n  
 Tullock, G. 746, 1205, 1209  
 Tung, L. 326  
 Tunstall, J. 1150  
 Turgot, A.R.J. 34, 35  
 Turner, R.K., *see* Pearce, D.W. 989  
 Turner, S.E., *see* Bowen, W.G. 562  
 Tuth, R.C., *see* Kohut, A. 393  
 Tversky, A., *see* Kahneman, D. 933n
- Uitermark, P.J. 743n  
 Ulibarri, C.A. 427n, 443, 986  
 UNESCO, 403, 987, 1202  
 Unger, F., *see* Glück, J. 510  
 United Nations Development Programme, 401  
 United Nations General Assembly, 401  
 United States Department of State, Washington Conference on Holocaust-Era Assets, 239  
 Upchurch, A. 52  
 Upright, C.B. 460  
 Urrutiaguer, D. 442, 453, 458, 459, 459n, 1207  
 Ursprung, H., *see* Schulze, G. 1074, 1078, 1078n, 1087n  
 Ursprung, H.W., *see* Schulze, G.G. 1023, 1126n, 1136n, 1212  
 US Bureau of the Census, 125, 138n, 141, 345, 828, 832  
 US Copyright Office, 226n, 228n  
 US Department of Commerce, Bureau of the Census, 476n, 477, 479, 494  
 Useem, M., *see* DiMaggio, P. 419n, 421, 422, 1039n
- Vajda, Á 1270, 1280  
 Vallely, P. 322n  
 Valletti, T., *see* Peitz, M. 581n, 593n, 605  
 Valsan, C. 974  
 Van Autenboer, E. 89n  
 van den Berg, G. 939, 940  
 van den Broek, A. 1217n  
 van den Broek, A., *see* Appelman, M.D. 736n, 737, 737n, 741, 743n, 751n  
 van der Ploeg, F. 744, 1174n, 1175n, 1183n, 1193, 1205n  
 van der Wielen, N., *see* Boorsma, P.B. 1265  
 van der Wielen, N., *see* van Hemel, A. 1029n  
 van der Woude, A. 93n  
 van der Woude, A., *see* De Vries, J. 75n  
 Van Eijck, K. 423n, 460  
 Van Haverbeke, P. 973  
 van Hemel, A. 1029n  
 van Hemel, A., *see* Boorsma, P.B. 1265  
 van Miegroet, H.J. 91n  
 van Miegroet, H.J., *see* De Marchi, N. 15, 71n, 75n, 82n, 90n, 91n, 96n, 98n, 99n, 104n, 108n, 179n, 261, 265, 279, 960, 960n  
 van Ours, J. 737n  
 van Ours, J., *see* Ginsburgh, V. 785n, 904, 939, 940  
 van Ours, J., *see* van den Berg, G. 939, 940  
 Van Puffelen, F. 1059n, 1061, 1064n, 1065  
 Van Tilborgh, L. 278n  
 van Ypersele, T., *see* Francois, P. 1149  
 van Ypersele, T., *see* Shavell, S. 293n  
 Van Zanden, J.L. 99n  
 Vanden Eeckaut, P., *see* Mairesse, F. 1026n  
 Vanden Houte, C. 78  
 Vanderporten, B. 939, 940  
 Van't Eind, G.J., *see* Goudriaan, R. 1021  
 Vari, S., *see* Debeauvais, R. 778  
 Varian, H. 296, 296n  
 Varian, H., *see* Bergstrom, T. 480, 481n, 499  
 Varian, H., *see* Kahin, B. 290n  
 Vasoo, S., *see* Wong, C.M. 1280  
 Vaughan, D.R. 1059  
 Veblen, T. 59  
 Vella, F., *see* Jones, C. 939, 940  
 Velmans, T., *see* Chamoux, F. 256n, 260, 262, 262n, 273  
 Velthuis, O., *see* Rengers, M. 883n  
 Venables, A., *see* Fujita, M. 1087n  
 Venture Philanthropy Partners, Venture Philanthropy, 497  
 Verbert, N. 950, 973  
 Verhoeff, R. 452  
 Vermeylen, F. 77n, 85, 87n–90n, 93n, 115  
 Vickers, J., *see* Harris, C. 1197  
 Vickrey, W.S. 586, 608  
 Victor, P.A. 990  
 Vidarte, J.I. 1070n  
 Vieira, M.A. 324



- Villani, A. 447n  
 Vincent, D. 937, 938  
 Vincent, D., *see* McAfee, R.P. 937–940  
 Vincent, S., *see* Chanel, O. 932, 934, 939, 940, 973  
 Viner, J. 364n  
 Vishny, R., *see* Hart, O. 1210  
 Vladeck, B.C. 1261n  
 Voith, R., *see* Crone, T. 968n  
 Volle, N. 279n  
 von der Fehr, N. 939, 940  
 Voss, G.B. 561n  
 Voss, Z.G., *see* Voss, G.B. 561n
- Wachter, S., *see* Case, K.E. 968n  
 Wackernagel, M. 78n, 81n, 115  
 Waelbroeck, P., *see* Peitz, M. 290n, 709, 711, 712, 714, 715  
 Wahlström, B. 1086  
 Waits, C.R. 832  
 Waits, C.R., *see* McNertney, E. 829n, 873n  
 Wakin, D. 815  
 Waldfoegel, J., *see* Berry, S. 557, 605  
 Waldman, M., *see* Novos, I.E. 296  
 Walker, L. 547  
 Walker, R., *see* Storper, M. 784  
 Wall, G. 1024n  
 Wall, G., *see* Mitchell, C. 1059  
 Wall Street Journal, 233  
 Wallace, J.A. 1267  
 Wallace, N. 960  
 Wallace, N., *see* Meese, R. 967, 968  
 Walls, W.D. 621n, 638, 640  
 Walls, W.D., *see* De Vany, A. 621, 627, 638–640, 642–644, 651, 652n, 653–658, 684, 685, 1156n  
 Wallsten, S., *see* Hahn, R.W. 293  
 Walras, L. 185  
 Wang, F. 967  
 Wanhill, S., *see* Cooper, C. 1058  
 Ward, T., *see* Mace, M.-A. 512  
 Warnke, M. 771n  
 Warr, P.G. 480  
 Wassall, G.H. 770, 794, 816, 829, 829n, 835n, 841, 842, 842n, 873, 902  
 Wassall, G.H., *see* Alper, N. 164n, 769, 771n, 774, 776, 794, 795, 816, 817, 821, 827n, 829, 847, 848  
 Waterman, D. 452  
 Watson, J., *see* Ekelund, R. 950, 974  
 Watson, M., *see* Engle, R. 965
- Watson, P. 914  
 Watt, R. 889n  
 Watt, R., *see* Towse, R. 890n  
 Wauthy, X., *see* Gabszewicz, J.J. 580n  
 Weber, M. 391  
 Weber, R., *see* Milgrom, P. 105n, 932, 934, 937–940  
 Weick, K. 775  
 Weil, S.E. 1019n, 1032, 1264  
 Weinberg, B., *see* Galenson, D. 818, 926  
 Weinberg, C. 423n  
 Weinberg, C., *see* Krider, R. 620n  
 Weinberg, C., *see* Ryans, A.B. 429n, 460  
 Weingast, B.R., *see* Grief, A. 71n  
 Weinraub, B. 687  
 Weinstein, M. 553, 657  
 Weir, R., *see* Peacock, A. 784n, 797, 1131n  
 Weisberg, R.W. 512n  
 Weisbrod, B.A. 480, 484, 1026, 1029n, 1035, 1264n, 1280n  
 Weisbrod, B.A., *see* Schiff, J. 489, 1029n  
 Weisbrod, B.A., *see* Segal, L.M. 1035  
 Weiss, A. 870, 875n  
 Welch, E. 72n  
 Welch, E.S. 1151  
 Welch, F., *see* Topel, R. 798  
 Welch, I., *see* Bikhchandani, S. 1156n  
 Wenzlick, R. 954  
 Werness, H.B. 256  
 West, E. 419n, 421n, 447n, 481, 488  
 West, E., *see* Dobson, L. 441, 453, 456, 457, 461  
 West, E., *see* Morrison, W.G. 441, 460, 1077n  
 Westby, D. 771n, 772  
 Westermann, M., *see* Dewenter, R. 1073, 1083  
 Weyers, S., *see* Ginsburgh, V. 173n, 328n, 960n  
 Wheeler, S., *see* Hammack, D. 1302  
 Whinston, A.B., *see* Soon-Yong, C. 740n  
 Whinston, M.D., *see* Bolton, P. 559n  
 White, C.A., *see* White, H.C. 540n, 771n, 782  
 White, H.C. 540n, 771n, 782, 796  
 White, J.L. 1032n  
 White Jr., L. 341n  
 Whitehead, J. 1072n  
 Whitesides, T. 548, 556  
 Whyte, J.L. 743n  
 Wiesand, A. 420, 1073n, 1074n  
 Wiesand, A., *see* Fohrbeck, K. 777  
 Wiggins, G.A., *see* Pearce, M. 157–159  
 Wijnberg, N. 540n, 878, 884n, 887, 1202n

- Wilbur, K.C. 575n, 604, 605  
 Willman, S.S. 558, 575n, 1083n  
 Wiles, P. 572n  
 Wilhelm, M.O., *see* Ribar, D.C. 1028n  
 Willemen, P., *see* Rajadhyaksha, A. 1164  
 Williams, M., *see* Adamowicz, V. 1075n  
 Williams, M., *see* Boxall, P. 1076n  
 Williams, R. 28  
 Williamson, O.E. 535  
 Willig, R.D., *see* Ordovery, J.A. 296, 296n  
 Willis, R. 869, 872  
 Wilson, C. 115, 116  
 Wilson, E. 1127n  
 Wilson, J. 88n  
 Winckelmann, J.J. 262  
 Winn, S. 1309  
 Winter, D.G., *see* McClelland, D.C. 392  
 Winter, R., *see* Mathewson, F. 747  
 WIPO, 1111  
 Wisdom, J. 204  
 Withers, G. 421n, 423, 425, 433, 435–437, 449, 450, 841, 842, 873, 1021n, 1022n  
 Withers, G., *see* Throsby, D. 419, 419n, 421n, 423–425, 433n, 434n, 436n, 485, 488, 1077n, 1198  
 Wittgenstein, L. 191, 203  
 Wittke, A. 958  
 Wolf, R.E., *see* Millen, R.F. 1162  
 Wolff, E.N., *see* Baumol, W.J. 347n  
 Wong, C.M. 1280  
 Wong, C.M., *see* Chua, V.C.H. 1280  
 Wooders, M., *see* Eaton, B.C. 593n, 608n  
 Woods, R., *see* Forrest, D. 423, 425, 427, 452, 465  
 World Commission on Culture and Development, 403, 988n  
 World Commission on Environment and Development, 988n  
 Wright, A. 75n, 78n  
 Wright, C.E. 1301  
 Wright, J. 580, 591n  
 Wright, J., *see* Kaiser, U. 605  
 Wulff, E., *see* Bille, T. 1073n  
 Wyngarden, H. 954  
 Wyzomirski, M.J. 1270n  
 Yarmolinsky, A. 1269n  
 Yetkiner, I.H. 740n  
 Young, J.R. 715n  
 Youngkoo Lee, E., *see* Szenberg, M. 732n  
 Zablottney, S. 99n  
 Zajdenweber, D., *see* Sornette, D. 619n, 638  
 Zakaras, L., *see* McCarthy, K. 480, 491, 492, 496, 998n, 1239  
 Zaleski, D., *see* Cooper, S. 319, 321  
 Zan, L. 1007, 1035n  
 Zang, I., *see* Ginsburgh, V. 1034n  
 Zanola, R., *see* Biey, M.L. 918  
 Zanola, R., *see* Locatelli Biey, M. 950, 973, 974  
 Zeigler, J. 1309  
 Zentner, A. 707, 712, 712n  
 Zhang, M.X. 714  
 Zimmer, A. 498, 1249  
 Zofío, J.L., *see* García, M.I. 1193, 1194  
 Zolberg, V. 1000n  
 Zollmann, C. 1301  
 Zonderman, S., *see* Ross, M.H. 973  
 Zondervan, S., *see* Ross, M.H. 919, 920, 922  
 Zorn, P., *see* Wang, F. 967  
 Zuidhof, P., *see* Klamer, A. 997n  
 Zuzanek, J. 452

## SUBJECT INDEX

- A-list/B-list 723, 734  
ability bias 870  
acquired skills 868  
acquired taste 429  
actors 619, 637–641, 649–663  
ad clutter 577  
ad volume 577  
administrative burdens 1029  
admission price 624–636, 1021  
advertising 567–612, 767  
advertising avoiders 600  
advertising expenditures by medium 578  
advertising lovers 589, 600  
advertising prime-time 579  
Aesop 31  
aesthetic value 169–205, 998  
aesthetics 7, 288  
affiliates 1158  
age 420  
age-earnings profiles 823, 872  
agent 543, 545, 546  
agglomeration economies 1087–1091  
aggregation bias 659  
aggregative data 424  
agriculture 409  
Akerlof's "lemons" model 92  
almost ideal demand system 423  
alternative leisure activities 1021  
ambiguous probabilities and sure things 660  
American Association of Museums 1284  
American Institutionalists 59  
American model of cultural support 1258  
American Society of Composers, Authors and  
    Publishers (ASCAP) 699  
Americans and the Arts 1281  
Amsterdam guild 92  
anchoring 933  
ancillary activities 1034  
ancillary goods 1034  
Andersen, Hans Christian 1060  
angel's nightmare 622, 648, 652  
Anglo Saxon tradition 888  
answering bias 1078  
anthropology 35, 150, 402  
antitrust issues 1160, 1162  
Antwerp 87  
AOL Time Warner 1161  
Apollonian 152  
appellation of origin 1111  
appreciation scores 446  
apprenticeships 883, 887  
appropriability 290  
appropriation art 217–220  
arbitrage 112  
arbitrating dealers 95  
architectural design 1037  
Aristotle 8, 40  
arm's length principle 1208  
Arnold, Matthew 49  
art appreciation 448  
art assets 914, 918  
art critics 44  
art dissemination 343  
art experts 912  
art for art's sake 536, 539, 561, 776  
art galleries 563, 1162  
art historians 255  
art house circuits 1157  
art installations 1162  
art investment returns 921  
art law 211–250, 273–279, 330  
Art Loss Register 237, 238, 1152  
art market 5, 63, 69–116, 949, 950, 958, 964,  
    975, 976, 1131  
art museum 992  
    *see also* museums  
Art Net 965  
art price indices 914, 922, 947–974  
art schools 880, 1090  
artist-in-residence placements 887  
artistic conventions 1040  
artistic goods 7  
artistic training 879, 881  
artistic value 155, 328–331, 333  
artists 255, 294, 411, 509, 767, 874, 1256  
artists' earnings 9, 825, 867–874, 879, 887  
artists' employment 767

- artists' gross return 538  
 artists' labour markets 5, 514, 765–905, 1089  
 artists' mobility 1151  
 artists' multiple job-holding 9, 770, 827  
 artists' productivity 652  
 artists' reputations 961  
 artists' resale right 113  
 artists' supply decisions 9, 768, 890  
 artists' training 879, 881, 1237  
 artists' unemployment 834  
 arts and culture 28, 29, 38  
 Arts Councils 1208  
 Arts Endowment 1231  
 arts in society 63  
 arts policy 26  
     *see also* cultural policy  
 arts financing 346  
 arts tourists 1062  
 artworks 985  
 Asian values 409  
 asset 296  
 asset class 949  
 associationism 362  
 asymmetrical information 999  
 atelier 257  
 attendances per capita 434  
 attribution rights 224, 226  
 auction data 960, 965  
 auction houses 113, 912, 929, 1152  
 auction mechanism 928  
 auction theory 1145  
 auctions 13, 104–107, 269, 295, 909–942, 965  
 audience segmentation 418  
 audience surveys 418  
 audiovisual services 767, 797, 1149, 1169, 1175  
 Australia 1287, 1289  
 Austria 14  
 authenticity 239, 242, 276  
 authenticity disputes 240, 243  
 autocorrelation 432, 438  
 automatic tax credit 1287  
 average price 953  
 awards 884  
 axioms 467  
  
 balanced growth 1090  
 Banfield, T.E. 58  
 Barnes Foundation 248–250  
 baseball franchises 429  
 basins of attraction 621  
 Basle 1078  
  
 Baumol's cost disease 4, 346, 489, 685, 767, 891,  
     1024, 1080, 1085, 1197, 1204  
 Bayes estimators 963  
 Bayesian actor 793  
 beauty 171, 174–179, 181–186, 189–191  
 beliefs 360, 363, 391  
 Bell, Vanessa 63  
 benefit–cost ratio 995  
 Bentham, Jeremy 4  
 Benthamite utilitarianism 59  
 bequest value 482, 996, 1023, 1072  
 Berne Convention 1171  
 biases 659, 950, 963, 1078  
 bidding wars and the producer's curse 658  
 bilateral investment treaties (BITs) 1173  
 Bilbao 1063, 1070  
 biodiversity 404, 988  
 blind bidding 1160  
 block booking 631, 1160  
 blockbuster 1024, 1031  
 Bloomsbury Group 4, 26, 61, 62, 65–66  
 BMG Music 1161  
 BMI (Broadcast Music Inc.) 699  
 bohemianism 782  
 book clubs 734  
 book industry 732, 1157  
 book market 721–759  
 book publishing 12, 536, 726, 784  
 book reading 724, 726  
 book reviews 740  
 booking 619, 624, 629–637  
 bookkeeping 1032  
 bookstores 742, 752, 1160  
 bootstrapping 967  
 bounded rationality 546, 882  
 Bowie theory 295, 685  
 box office 1193  
 brain drain 406  
 Brandts Klaedefabrik 1059  
 British Arts Council 65  
 British Museum 1024  
 broadcasting 570, 697, 767, 1132, 1229  
 Bruges 77  
 budgets 622, 650–652, 659  
 bulk marketing of images 86  
 bundling 1158, 1160  
 bureaucracy 1002, 1208  
 burning effect 14, 935  
 Burra Charter 987  
 business models 295, 714, 1160

- business stealing 584, 595  
 buyers' premiums 930
- Canada 1280, 1286  
 capacity 302, 453  
 capital 485, 1165  
   *see also* cultural capital; human capital;  
   natural capital; social capital  
 capital intensity 1261  
 capitalism 800  
 captive audience 455  
 career patterns in the arts 10, 294, 772–775,  
   813–861  
 Carnegie, Andrew 1302  
 Carnegie Corporation 1304  
 cartelization 685  
 catalogue effects 939  
 catalogues 256, 960  
 CD sales 709  
 CD-ripping and burning 1171  
 censorship 16, 311–332, 516  
 censure 315  
 census data 770, 828  
 Census Public Use Microdata Samples (PUMS)  
   828  
 certification 305, 881  
 certifiers 539, 544  
 chaos 642–647  
 charitable contributions 1258–1290  
 charity 1301  
 childhood participation 441  
 Chile 1287  
 China 316, 326  
 choreographers 774  
 Christianity 312, 313, 316, 327, 328, 404  
 Christie's 113, 912, 929  
 circle model 606  
 city government aid 1227  
 clash of civilizations 404  
 Classicists 64  
 Clear Channel Communications 686  
 clearance 626–634  
 climate change 990  
 club goods 367, 481  
 clusters 1107  
 clutter 577  
 co-financing 1287  
 co-production 1170–1173  
 cognition 173, 174, 181, 186, 200  
 cognitive skills 511  
 collaboration and creative networks 1179  
 collective property rights 1111  
 collectivism 361, 374  
 collectors 255  
 collusion 113, 929  
 commercial broadcasting 1132  
 commercial placement 604  
 commercial presence 1178  
 commercialisation 1035  
 commission exchange 72  
 Commission on Private Philanthropy and Public  
   Needs 1259, 1281  
 commissions 919  
 communication 288, 405  
 communications technology 509  
 Communist Party 320  
 community 367  
 community foundations 1304, 1312  
 compensating differentials 39, 778  
 compensating variation 1023  
 competition 71, 290, 360, 372, 489, 501, 1151  
 competition policy 1160, 1161  
 competitive bottleneck 594  
 competitive effect 302  
 complementarities 1034  
 complementary costs 1021  
 complementary goods 449, 586, 675, 687  
 complementary software 292  
 completion guarantee 650, 652  
 composers 13, 130, 774  
 computer game industry 767  
 computer painting 343  
 Comstock 319, 320  
 concentration of sellers 555  
 concert industry trends 676  
 concert pricing 675  
 concert supply curve 692  
 conditional distribution fee 629  
 conditional maximum likelihood estimation 423  
 conflict resolution 403  
 Confucianism 402  
 congestion costs 785, 1196  
 conservation 985  
 conservative political movement 1314  
 conservatories 129, 131–133  
 conspicuous consumption 31, 36  
 constant return to scale 302  
 consumer 289  
 consumer sovereignty 570, 1130  
 consumption capital 443, 545, 897  
 consumption gap 422  
 consumption goods 994

- contemporary art auctions 933  
 Contemporary Art Society 65  
 contingent demand 1145  
 contingent pricing 627  
 contingent valuation 196, 997, 1023, 1076, 1077,  
     1199, 1242  
 contingent work 768  
 continuously compounded return 954  
 contracts 11, 535, 617–639, 648–650, 657–673,  
     1160  
 controls on foreign ownership 1179  
 Convention on Cultural Diversity 1168, 1170,  
     1174  
 Convention on Illicit Trade in Cultural Property  
     1172  
 Convention on the Protection of the Underwater  
     Cultural Heritage 1170  
 convergence 393  
 cooperation 1108  
 copies 221, 253–281, 1079  
 copyright 10, 137, 138, 213–215, 217–222, 255,  
     285–306, 351, 516, 1085, 1086, 1145,  
     1155, 1170, 1171, 1178  
 copyright infringements 290  
 copyright law 113, 868, 871, 888, 889  
 copyright royalty payments 890  
 Copyright Treaty 1171  
 corporate charitable contributions 1266,  
     1277–1279, 1290, 1307  
 cost disease, *see* Baumol's cost disease  
 cost functions 1025  
 cost–benefit analysis 998, 1271  
 court orchestras 127  
 crafts 287, 767, 1117  
 creative alliances 1069  
 Creative Artists Agency 547  
 creative city 1090  
 creative class 803  
 creative commons license 295  
 creative destruction 788  
 creative industries 11, 1193  
 creative writing 522  
 creativity 7, 156, 255, 509, 780, 867, 878, 1040,  
     1067, 1069, 1106  
 creativity in consumption 160  
 creativity spillovers 1090  
 credentialism 871  
 credit 1293  
 credit cards 579  
 cross-elasticities 417, 451, 1021  
 cross-fertilization of ideas 1154  
 cross-media successes 1159  
 cross-section analysis 433, 879  
 cross-subsidization 488, 489, 748, 1205  
 crowding 887, 1243  
 crowding in 1283, 1284  
 crowding out 498, 1153, 1283–1285  
 cultivation of taste 162  
 cultural accountancy 770  
 cultural arms race 128  
 cultural assets 1092  
 cultural budgets 1192  
 cultural capital 20, 153, 885, 887, 985, 1092,  
     1139, 1145  
 cultural change 1081  
 cultural convergence 462  
 cultural development 1053, 1065, 1071  
 cultural districts 21, 1101–1118  
 cultural diversity 18, 402, 1068, 1148, 1150,  
     1168, 1170, 1174  
 cultural economics 4, 28, 1144, 1145  
 cultural education 1195  
 cultural employment 1187  
 cultural exception/exemption 569, 1173  
 cultural expenditures 1190  
 cultural exports 345  
 cultural goods 5  
 cultural heritage 18, 983–1013, 1077, 1166,  
     1244, 1246, 1264  
 cultural indicators 994  
 cultural industries 11, 533–759, 1081, 1086,  
     1154  
 cultural objectives 743  
 cultural observatories 1179  
 cultural occupations 1188  
 cultural participation 1194  
 cultural patrimony 1146  
 cultural planning 1069  
 cultural policy 17, 770, 799, 1092, 1183–1294  
 cultural programming 1069  
 cultural resistance 462  
 cultural statistics 1177  
 cultural subsidies 1012, 1172  
 cultural sustainability 1055  
 cultural trade 1177  
 cultural value 152, 482, 518, 986  
 culture 5, 150, 360, 361, 401, 770, 1054  
 Culture Wars 1309  
 culture-based goods 1103  
 Cupid and Psyche 63  
 curators 255

- customs 368, 378, 985  
 Czechoslovakia 14
- Daguerreotype 287  
 dance 451, 514, 1231  
 Daphne and Apollo 63  
 data 464, 951, 965, 968, 1166  
 day-wage equivalents for painters 98  
 de Piles, Roger 960, 976, 978  
 de-accessioning 1032  
 dealers 90, 261, 1152  
 decency 1233  
 decentralization 1265  
 decision making 659, 1265, 1269  
 decision rights 541  
 Declaration of the Responsibilities of the Present  
     Generation Towards Future Generations  
     1170  
 Declaration on the Rights of Indigenous People  
     1170  
 declining price anomaly 938  
 dedicated sales venues 87  
 dedicated taxes 1256  
 deductibility of contributions 1258, 1267, 1293  
 deed of covenant 1263, 1280  
 defendants 632–636  
 delegation 1002, 1210  
 DeLong, J. Bradford 344  
 demand 37, 74, 302, 415–467, 1073, 1074, 1078  
 demand curve 82, 691  
 demand elasticities 423  
 demand for artists 876  
 demand for cultural goods 1073  
 demand for paintings 260  
 demand for the arts 8  
 demand function 1022  
 democracy 410  
 depreciation 885  
 derivative work 221, 222, 295  
 descending price auctions 104  
 désétatisation 1265  
 design 767, 1106  
 development 401, 1055, 1069, 1083, 1104, 1117,  
     1139  
 devolution 1004, 1265  
 diagonal weighting matrix 973  
 Dickens, Charles 26, 515  
 diffusion 393, 1151  
 digital files 1157  
 Digital Millennium Copyright Act 707  
 digital performance right in sound recordings 706  
 digital photography 287  
 digital rights management 715  
 digital satellite transmission 1157  
 digitization 258  
 Dionysian 152  
 directors 458, 618, 619, 649–658, 744  
 discovery rule 237  
 discrete choice 581, 1075  
 discrimination 128, 632  
 discriminatory prices 349  
 disputes 239, 1145  
 distance-weighted repeat-sale 915  
 distinction hypothesis 1198  
 distortionary taxes 1201  
 distribution 295, 434, 684  
 distribution and retailing of books 1157  
 distribution fee 629  
 distribution networks 1180  
 distribution of artists' earnings 876  
 distributive justice 804  
 divergent thinking 511  
 diversification of risk 795, 949  
 diversified portfolio of artworks 922  
 diversity 364  
     *see also* biodiversity; cultural diversity  
 dividends 302  
 division of labor 89, 367, 368  
 doctrine of the unintended consequences 886  
 dominance of extreme events 644, 653, 654  
 donations 495, 1031, 1264, 1266, 1276  
 donor-advised funds 1313  
 donors 437, 561, 1031, 1261  
 double-log function 423, 959  
 downsizing 406  
 drawing 259  
 dubbing 1167, 1168  
 duplication 259, 572  
 Durkheim, Emile 803  
 Dutch Republic 93  
 DVDs 1157  
 dynamic optimization 162
- e-books 734  
 e-Musik 420  
 e-philanthropy 497  
 earned revenues 494  
 earnings 9, 825, 867–871, 874, 879, 887  
 earnings differentials 875  
 earnings forgone 872  
 earnings functions 823, 841, 845, 868–876  
 earnings gap 822  
 earnings inequality 840

- earnings penalty 836, 847  
 economic development 399–412, 1053, 1055,  
     1069, 1074, 1081, 1118, 1139  
 economic growth 1053, 1055, 1056, 1071, 1138  
 economic impact studies 1024, 1057–1059, 1065  
 economic performance 359–395  
 economic rationality 887  
 economic value 150, 169–205, 518, 986  
 economies of agglomeration 1104  
 economies of scale 784, 1037, 1079  
 economies of scope 1087–1088  
 ecosystems 988  
 education 8, 35, 366, 419, 511, 1073, 1074  
 education elasticities 441  
 education gap 422  
 education value 482, 1023, 1068  
 educational production function 877  
 educational value of museums 1227  
 efficiency 488, 620, 899  
 efficient component-pricing rule 353  
 Egypt 312  
 Eiffel Tower 1072  
 Einstein, Albert 901  
 electronic rights management information 1171  
 elitism 419  
 emancipation of peasants 129  
 embedding effect 997  
 EMI 1161  
 emotion 361, 367, 370, 390  
 endogenous growth 801  
 England 317  
 engravings 257  
 Enlightenment 26, 129, 1151  
 Enlightenment thinkers 42  
 enterprise culture 378  
 entertainment conglomerates 558  
 entrance fees 1033  
 entry deterrence 593  
 environment 402  
 equalizing differences 777  
 equity 994, 1267  
 estate management 244–248  
 estimation errors 452  
 estimators 957  
 ethical principles 403  
 ethnic background 461  
 European Capitals of Culture 1060  
 European civil law tradition 888  
 European Commission 1161  
 European Group on Museum Statistics  
     (EGMUS) 1195  
 European Union 1172  
 Eurostat 1188  
 evaluation 173, 176, 181, 189–192, 201, 203  
 excess burden of taxation 1243  
 excess supply of artists 876  
 existence value 482, 996, 1023, 1072, 1079  
 exit option 1112  
 experience 511  
 experience good 5, 675  
 experts 931  
 EXPO 2000 1061  
 export license 112  
 externalities 7, 31, 34, 53, 193, 194, 999, 1022,  
     1242  
 extreme events 617, 622, 638, 644, 645, 654–658,  
     660, 663  
 extrinsic motivations 515  
 extrinsic reward 887  
  
 fair use 215–221, 290  
 fairness 1256  
 fakes 256  
 fame 256  
 families 367  
 family background 511  
 fashion 38  
 federalism 1317  
 female artists 849  
 festivals 1166  
 file sharing 291, 707, 1171, 1176  
 film industry 557, 615–664, 773, 1089  
 film libraries 1157  
 film writers 822  
 films 12, 630, 797, 1083  
 finite expansibility 297  
 firm concentration 480  
 First Amendment 311, 324, 330–333  
 first sale doctrine 215, 222  
 fiscal illusion 1002  
 Fisher Ideal index 680  
 Five College Dance Program 819  
 fixed book price agreement 743, 749, 1205  
 fixed costs 1024  
 fixed factor 300  
 Flanders 13  
 flat price 630  
 flexibility 378, 772  
 flexible specialization 552  
 Florence agreement 1171  
 followers and leaders 360, 391  
 Ford Foundation 1281, 1306  
 forecast 619, 642–647, 660, 663



- foreign investment 405  
 forgeries 255, 256  
 forgone taxes 1263, 1269  
 formalist theory of aesthetics 62  
 formation of identity 1068  
 formula deals 627, 634, 635  
 Forster, E.M. 63  
 foundations 1290, 1303  
 France 1009, 1290  
 franchise 629, 631, 635  
 fraud 305  
 Free Trade Agreement of the Americas (FTAA)  
     1173  
 free-lance activity 129–132, 778  
 free-rider problem 997  
 freedom 393, 395  
 Freud, Sigmund 62  
 Frisch demand function 445  
 fundamentalist Protestant 463  
 funding diversification 496  
 fundraising 483, 496, 497, 501  
 future generations 1133
- gallerists 113  
 gallery 258, 539  
 game theory 516  
 Ganz, Victor and Sally 923  
 Garnett, David 63  
 gatekeepers 289, 537, 543, 547, 548, 557  
 gay affinity for the arts 463  
 gender 461  
 gender gap 823  
 General Agreement on Tariffs and Trade (GATT)  
     1168, 1175  
 General Agreement on Trade in Services (GATS)  
     1168–1169, 1175–1178  
 generalized central limit theorem 638, 662  
 generative linguistics 157  
 genius 512  
 geometric mean estimator 957, 967  
 geometric repeat-sales estimator 957  
 Germany 1011, 1280  
 gift exchange 72  
 giving to cultural organizations 1276  
 Glasgow 1062, 1070  
 global ethics 403  
 globalization 17, 378, 404, 583, 905, 1081, 1082,  
     1138, 1150, 1186  
 GNU General Public License 295  
 Goethe, Johann Wolfgang von 50  
 Gossen's second law 1197
- government arts funding 5, 293, 498, 742, 797,  
     1030, 1072, 1083, 1132, 1317  
 Grant, Duncan 63  
 Grantmakers in the Arts 1306  
 grants 523, 794, 1030, 1244  
 Grateful Dead 1060  
 Gray, Thomas 344  
 Great Depression 1304  
 Great Society 1305  
 Greece 34, 311, 314  
 greenlighting 659  
 Guggenheim Museum Bilbao 1064  
 guilt 367, 371, 390
- H-creative 158  
 habit formation 442  
 hammer price 950, 951, 965, 968  
 Hanks, Nancy 1308  
 hazard 643, 659  
 Hebraism and Hellenism 50  
 Heckman, J. 966, 977  
 hedonic and repeat sales price indices 916  
 hedonic approach 1076  
 hedonic models 915, 952–968  
 hedonic pricing 996, 1076  
 hedonic regression 15, 927, 950–978  
 Hendon, William 4  
 Heraclitus 404  
 herd behavior 64, 781  
 heritage, *see* cultural heritage  
 heritage tourism 1166  
 Hermann, F.B.W. von 54  
 heterogeneity 458, 949, 950, 958  
 heteroskedasticity 432, 958, 968  
 hijacking 294  
 Hislop Art Sales Index 965  
 historical artefacts 1131  
 Hofstede's classification 375  
 Hogarth Press 64  
 hold ups 552, 648  
 Hollywood 773, 1089  
 Holy Roman Empire 128, 134, 135, 137  
 home radios 140  
 homosexuality 463  
 Hotelling, Harold 1199  
 household production model 296, 440  
 human capital 10, 133, 442, 513, 769, 825, 869,  
     875, 883, 885, 890, 1068, 1145  
 human development 401  
 human resource management 801  
 human rights 403

- humanities and the arts 48  
 Hume, David 4  
 hysteresis 1146
- Ibsen, Henrik 515  
 Iceland 1289  
 iconoclasm 312–314  
 idealism 180  
 ideas 288  
 illegal music downloads 1082  
 impact studies 1024, 1057–1059, 1065  
 implicit contract 370  
 implicit opportunity cost of time 429  
 implicit price 953–969  
 Impressionists 782, 916  
 inalienability of human capital 877, 890  
 incentives 290, 885, 887, 1288  
 income averaging 1268  
 income effect 430, 1085, 1276  
 income elasticity 139, 417, 1073, 1276–1278  
 income gap 422, 776  
 Income Tax Amendment 1315  
 income tax 1240, 1280  
 income variability 769  
 incomplete contracts 540  
 indemnity insurance 1235  
 independent exhibitors 634–636  
 independent record producers 1159  
 independent record promoters 705  
 indigenous cultures 1118  
 indirect aid 1258, 1260, 1263, 1268  
 indirect cost 880  
 indirect federal subsidies 498  
 individual charitable giving 1271  
 individualism 361, 362, 374, 394  
 induced gift 1274, 1275, 1281  
 industrial cultural district 1103  
 industrial organization 13  
 infinite variance 640–646, 656–663  
 infinite variety 723, 734  
 Infomercials 571  
 information 545, 547, 548, 626, 1156  
 information industry 618, 620, 622, 636  
 informational asymmetries 834, 870  
 informational cascade 1156  
 informative advertising 585  
 innate ability 867, 868, 873–875, 877, 881, 891  
 innovation 292, 341–342, 509, 1081  
 input markets 1089  
 Inquisition 314, 315  
 inside and outside odds 660  
 installation artists 1162  
 institutions 6, 360, 1027  
 Intangible Heritage Convention 987  
 integrated international auction circuit 112  
 integrated international traders in paintings 97  
 integrity right 225, 226  
 intellectual capital 34  
 intellectual property 7, 164, 708  
     *see also* copyright  
 intelligence 510  
 intergenerational equity 989  
 intergenerational transfer of wealth 1314  
 intergenerational value 1201  
 intermediation 294, 545  
 internal labor markets 801  
 internal rate of return 872, 995  
 International Intellectual Property Alliance 1178  
 International Standard Classification of  
     Occupations (ISCO) 1188  
 international trade 345, 405, 1082, 1152, 1177,  
     1178  
 international trade agreements 1148, 1168, 1175  
 Internet 12, 285–306, 1082, 1186  
 intertemporal dynamics 441  
 intertemporal resource allocation 994  
 intertemporal separability 444  
 intrinsic motivation 515, 887  
 intrinsic value 151, 190–193, 196, 198  
 investment appraisal 986  
 investment portfolio 949  
 Islam 316, 327  
 Italy 13, 1001, 1288
- Japan 317, 326  
 Jevons, William Stanley 54  
 job satisfaction 777  
 job-matching 790  
 joint consumption 780  
 joint-venture contracts 540  
 Journal of Cultural Economics 1078
- Keynes, John Maynard 4, 61  
 knowledge workers 802  
 knowledge economy 1149  
 Konstkompass rankings 924  
 Koons, Jeff 218, 220  
 kurtocracy 638, 641
- labor demand 768  
 labor markets 5, 514, 765–905, 1089  
 labor productivity 490, 493  
 labor rationing 791

- labor supply 9, 768, 890  
 language 21  
 Laspeyre's index 679, 964, 965  
 Last Judgment (Michelangelo) 315, 329  
 latent demand 83  
 Law of Anne 137  
 law of champions 643  
 law of one price 926  
 laws of production and cost 647  
 laws of profit 653  
 laws of the box office 641  
 leadership 362, 369, 380, 390  
 learning by consuming 161  
 learning by doing 365, 790, 872  
 leisure price effect 435  
 leisure theory 834  
 lending policy 1033  
 level playing field theorem 355  
 leveraging ratio 1275  
 Lévy stable distribution 621, 622  
 libraries 726, 752, 1288  
 licensing prices 1149  
 licensing rights 291  
 life cycle of an artist 925  
 life style factors 459  
 line item budgeting 1270  
 linguistic diversity 1170  
 lithographs 257  
 local government 1084, 1240  
 local public goods 1213  
 location choice 1066  
 location competition 599  
 location of arts and cultural industries 1086  
 logit models 432, 446, 609  
 London Artists' Association 64  
 long-run equilibrium 433, 593  
 Los Angeles 1109  
 low price strategy 438  
 loyalty 369, 371, 375  
 luck 652–654, 898  
 luxury 29, 33  
 luxury goods 31, 417, 1073  
  
 made to order and on spec production 108  
 Madonna 904  
 magazines 1157, 1173  
 magnet cities 135  
 Mapplethorpe, Robert 321  
 marginal costs 1024  
 marginal revolution 26, 53  
 marginal tax rates 1268, 1274, 1278, 1282  
  
 margins 75  
 mark-to-market 949  
 market access 1169  
 market capacity expansion effect 302  
 market development 1163  
 market exchange 72  
 market failure 19, 999  
     *see also* externalities; public goods  
 market power 301  
 market segmentation 457  
 market shares 80, 620, 633, 646, 656, 657  
 market structures 513  
 market value 636  
 marketing campaigns 1156  
 Marshall, Alfred 777  
 Marx, Karl 792  
 mass media 355  
 master agreements 635  
 master copy 295  
 masterpiece effect 14, 923  
 match value 581  
 matching grants 1270, 1275, 1276  
 material culture 988  
 Mathew effect 781  
 measurement error 917  
 mechanical license 698  
 media arts 452  
 media industries 12, 372, 567–612, 1160  
 Mei–Moses All Art Index 966, 968  
 Mellon, Andrew 1308  
 merchant dealers 96  
 merit bads 569  
 merit goods 481, 569, 800, 1126, 1203  
 methodology 875, 1020  
 Metropolitan Opera 125, 136  
 Midas 63  
 Mill, John Stuart 1185  
 Mincerian earnings function 872  
 minstrels 126  
 mobility of artists 1151  
 modernization 401, 1150  
 monetized transactions 1085  
 money burning 585  
 monitored tax credit 1287  
 monopolistic competition 735, 786  
 monopoly 32, 513, 581  
 Monte-Carlo 968  
 Montmartre 1090  
 moonlighting 794, 827  
 moral hazard 798, 1003  
 moral rights 16, 224, 227, 228, 230, 277

- Morris, William 26  
 most-favored-nation 1160, 1169  
 Motion Picture Association of America (MPAA)  
     714, 1148  
 movie industry 557, 773, 1089  
 movie price 450  
 movies 12, 615–664, 797, 1083  
 Mozart, Wolfgang Amadeus 515, 782  
 MP3 707  
 multi-homing 581, 585  
 multicollinearity 439, 958  
 Multilateral Agreement on Investment (MAI)  
     1173  
 multinationals 1153, 1157, 1160  
 multiple job-holding 9, 770, 827  
 museums 20, 56, 257, 992, 1017–1042, 1162  
 music 12, 41, 43, 123–142, 290, 311, 315, 514,  
     667–716, 771, 1158  
 music labels 1158  
 music majors 290  
 music piracy 709  
 music, popular 667–716  
 music publishing 125, 136, 701  
 music recording 125, 536  
 musical plays 454  
 musicians 670, 767, 1089  
 Muslims 404
- Napster 291, 708  
 narrowcasting 1208  
 Nation magazine 64  
 National Arts Council 1289  
 National Endowment for the Arts 321, 828, 829,  
     1229, 1259, 1281  
 National Foundation for the Arts and Humanities  
     1308  
 National Gallery of Art 1308  
 National Institutes of Health 1306  
 National Longitudinal Survey of Youth 815, 847  
 National Science Foundation 821, 1306  
 natural capital 20, 985  
 natural sale rate 937  
 neighboring rights 1171  
 networks 95, 439, 773  
 New Deal 1305  
 new economy 16, 341, 509  
 New Jersey Cultural Trust 1291  
 new media 767  
 new technologies 287  
 New York City 1228, 1230  
 New York Philharmonic Society 136  
 New York State Council on the Arts 1228  
 New York–New Jersey Metropolitan Area 1061  
 niche markets 1155, 1157  
 nobody knows principle 618–623, 626, 629, 630,  
     635, 647, 661, 734, 736, 784  
 non-arts work 795, 829, 850  
 non-distribution constraint 476, 1031  
 non-economic values 150  
 non-government organizations 475  
 non-market benefits 20, 994  
 non-market demand 1075  
 non-market values 1071  
 non-pecuniary rewards 882  
     *see also* psychic income  
 non-profit firms 11, 475, 487, 560, 799, 1031,  
     1127, 1226, 1256, 1266, 1287, 1293  
 non-profit sector 773, 1265, 1267  
 non-separability 448  
 non-use values 1075, 1076  
 Nordic Contact Group for Cultural Policy  
     Research 1191  
 North American Free Trade Agreement  
     (NAFTA) 1172  
 not-for-profit firms, *see* non-profit firms  
 novelty 90, 156  
 nudity 311, 312, 314, 315, 319, 326  
 nuisance cost 581
- O-rings 537, 553, 781  
 obscenity 318, 321, 324–326, 329, 330  
 occupational choice 419, 775, 777, 869, 879, 884  
 offensiveness 326, 327, 329, 330, 333  
 O’Keeffe, Georgia 237, 238, 244  
 Old Testament 63  
 Omega Workshops 62, 64  
 on-the-job experiences 791  
 on-the-job training 872  
 online auctions 113  
 open source 294  
 opera 129, 132, 136, 514, 1185  
 Opera America 464  
 opera houses 771  
 opera singers 819  
 operating costs 491  
 operationalism 149  
 opportunity cost 1021  
 optimal division of labor 789  
 optimal number of artists 902, 903  
 optimal reserve price 937  
 optimum advertising 594  
 option contracts 541, 552  
 option demand 996

- option value 482, 1022, 1072, 1079
- orchestras 437, 780, 1185, 1192
- Oregon Cultural Trust 1291
- originality 255
- originals and derivatives 74
- originals as capital assets 98
- output decisions 487
- overconfidence 785
- overcrowding 783
- oversupply of artists 769, 880, 882
- ownership constraints 1179
- ownership disputes 234–236
- ownership rights 237
  
- P-creative 158
- Paasche's index 679, 965
- pacifism 44
- painting 256
- paintings 85, 108, 255–280, 933, 1036
- Pan 63
- panel data 816
- Paramount 550, 620, 625–635
- parenting 366
- Pareto distribution 638, 685
- Paris Salon 318, 319
- parity price 354
- parody 220
- participation 418
- passions 30, 32
- passive inertia 466
- passive-use values 996
- patents 293
- paternalism 1198
- patronage 31, 35, 1317
- patrons 460, 1163
- payback method 995
- payola 702, 1160
- peer review 1270
- peer-to-peer networks 291
- Pensée Unique 606
- performance 360, 361, 390
- Performance and Phonograms Treaty 1171
- performance contracts 772
- performance fees 138
- performance firms 438
- performance indicators 1137
- performers 875
- performing arts 6, 415–467, 473–502, 798, 1075, 1079
- performing rights 697
- performing rights organizations 699
  
- persistence 821
- personality 511
- persuasive advertising 585
- Pew Charitable Trusts 1306
- philanthropy 18, 495, 1011, 1266, 1279, 1292, 1299–1319
- Philistines 64
- philosophy 5
- phonographs 140
- photography 16, 256
- physical capital 986
- Physiocrats 45, 46
- piano production 139
- piracy 290, 1171, 1178
- Piss Christ (Andres Serrano) 321, 327–330
- plagiarism 137
- platform competition 579
- Plato 7, 50, 311, 315, 326
- play-or-pay contracts 552
- pluralism 1225
- poetry 41–43
- policy instruments 743
- political externalities 34
- political 'markets' 1129
- politically conservative foundations 1310
- politicians 1208
- Pollstar 676
- Poor Relief Act 1301
- popular culture 345, 450, 569
- popular music 667–716
- popular referenda 1023
- post-artist employment 854, 855
- poverty 404
- pragmatism 151, 331, 333, 361, 362, 374, 394
- pre-sale estimate 951
- precautionary principle 989
- preferences 290
- preservation of art 346
- preservation of cultural heritage 1170
- President's Committee on the Arts and Humanities 1289
- press concentration 599
- prestige value 482, 1023, 1029, 1072, 1079
- price 82, 296, 677, 928, 1290
- price adjustment 618, 627
- price discovery 444
- price discrimination 138, 561, 620, 633, 634, 676, 1034, 1158, 1160, 1199
- price elasticity 417, 445, 1021, 1073, 1272–1276
- price fixing 911, 929
- price floor 598

- price incentive 1280  
 price indices 435, 679, 918, 950–964, 974–978  
 price pairs 954  
 price search model 443  
 price sensitivity curve 82  
 pricing 348, 591, 625, 627, 657, 1160  
 principal-agent model 1002, 1136  
 Principle of Duplication 572  
 Principle of Minimum Differentiation 573  
 print 257  
 private museums 1019, 1030  
 privatization 19, 1318  
 prizes and competitions 884  
 procedural fairness 1161  
 proceduralism 361, 374  
 product differentiation 786  
 product homogeneity 465  
 product selection 596  
 Production Code Administration 323  
 production functions 485, 501, 525  
 production of ideas 288  
 production process 648, 1079  
 productive versus unproductive labor 39, 45  
 productivity 93, 347, 622, 649, 652–658, 868,  
 870, 877–891, 1080  
 professional workers 831, 832, 836  
 professionalism 776  
 profit shares 552  
 programming diversity 1207  
 promotional campaigns 1160  
 property rights 255  
 property tax 1261, 1262  
 protectionism 1148  
 Protestantism 391, 402, 1316  
 provenance 154, 960, 961, 965, 966, 1152  
 proximity to market 1087  
 psychic income 10, 777, 882, 902, 903  
 public benefit 1287  
 public broadcasting 570, 1229  
 public domain 277  
 public expenditure 999  
 public goods 7, 348, 360–364, 390, 393, 480,  
 501, 569, 603, 996, 1072, 1145, 1196  
 public libraries 726, 752, 1288  
 public participation 1242  
 public support for the arts 5, 293, 498, 742, 797,  
 1030, 1072, 1083, 1132, 1317  
 public television 605  
 Public Use Microdata Sample (PUMS) 815, 827,  
 829, 836  
 publicity rights 900, 901  
 publishers 543, 736  
 publishing industry 548, 554  
 push-pin 43  
 putting-out system 99  
 PVRs (personal video recorders) 1157  
 quality 77, 276, 305, 453, 486, 501, 952, 1028,  
 1134  
 quality of life 361, 362  
 quasi-panels 816, 822  
 R&D 293, 527  
 Rachmaninov, Sergei 522  
 racial oppression 461  
 radical uncertainty 113  
 radio broadcasting 697  
 Ramsey price 352  
 rankings 619, 639, 643, 644, 646, 691  
 rate of return 869, 879, 880, 885  
 rate of time preference 448  
 rational addiction 162, 447  
 rational choice theory 1020  
 rationality 882  
 reading habits 727  
 record industry 543, 556, 773, 784, 1159  
 Recording Industry Association of America  
 (RIAA) 708  
 recording technologies 492  
 recursive process 432  
 reference group 1029  
 referenda 1023, 1078, 1136, 1212, 1270  
 Reformation 36, 314  
 regional agreements 1172  
 regional development 1054  
 regulation 255, 293, 999, 1041, 1149, 1160  
 regulatory capture 1149  
 relative shadow price 448  
 relativism 330, 331  
 religion 21, 126, 361  
 remixes 295  
 rent-seeking 1002  
 rental contract 627  
 rental escalation 628  
 rental rate 624–628, 636, 656, 657  
 rents 294, 544, 558  
 repeat-sales 915, 955, 957, 963–979  
 repeat-sales estimator 951–968, 977  
 repeat-sales regression (RSR) 950, 951, 963–977  
 repeated game 362  
 repertoire 453  
 representative consumer 581, 607  
 reproducibility 888, 889

- reproducible art 1079
- reproduction right 697
- reproductions 255
- reputation 294, 459, 540, 618, 634, 648–650, 780, 1031
- resale markets 72
- resale price maintenance 554
- resale rights 232, 233, 277, 304
- research 289, 293, 527
- research and development (R&D) 293, 527
- reserve prices 13, 912, 935
- residual demand 302
- retailers 554, 739
- retrospective surveys 817
- returns to art 921, 923
- returns to scale 292
- Revenue Act 1307
- revenue-maximizing 428
- revenue-neutrality property 591
- revenues of performing arts firms 493
- revision volatility 967
- Ricardian theory of rent 877
- right of preemption 112
- right of public display 215
- risk aversion 299
- risk-loving behavior 539, 543
- risk-taker 776
- rivals 451
- Robber Barons 1302
- Robbins, Lionel 58
- rock concerts 786
- Rockefeller, John D. 1302, 1306
- Rockefeller Foundation 1303, 1312
- role versatility 796
- Rolling Stone Encyclopedia of Rock & Roll 677
- Romania 1292
- Rome 312
- Rosenwald, Julius 1303
- royalties 278, 542, 697, 1159
- royalty contract 889
- runs 626–637, 641–643, 649
- runs and clearances 634
- Ruskin, John 4
- Russell Sage Foundation 1302
  
- sale rates 912
- sales 88, 105, 113, 935, 950–954, 958, 963–966, 968
- sales catalogues 960
- sales tax 1262
  
- sales without reserves 105
- sampling 466, 710, 919, 950, 966
- Samuelson rule 1201
- scale economies 784, 1037, 1079
- scalping 688
- scarcity 263
- Schiller, Johann Christoph Friedrich von 50
- schooling 867, 884
- Schumpeter, Joseph 341
- screen credits 618, 649
- Screen Writers Guild 552
- screening hypothesis 870, 882, 884
- search and information costs 781, 881
- seasonal attendance 458
- seasonality 457
- secondary labor markets 801
- secondary performance venues 355, 356
- secret reserve prices 937
- segmentation 427, 778
- self-actualization 789
- self-confidence 377, 460
- self-designing organizations 775
- self-employed artists 1130
- self-generating genius 775
- self-similarity 633
- sequential production 647
- Serrano, Andres 321
- sexual orientation 462
- shame 367, 390
- Sibelius, Jan 515
- signalling 585, 870
- simultaneity bias 436, 438
- Singapore 1280, 1289
- skewed distribution of earnings 876
- skilling, screening and socialisation 883
- slotting allowances 555
- small businesses 774
- Smith, Adam 4
- Smithsonian Institution 1226, 1308
- snobbists 64
- snowballing 901, 904
- social capital 404, 885, 886
- social cost 291
- social group 361, 367
- social rate of return 880
- socialization 459
- societies 367
- Society of European Stage Authors and Composers (SESAC) 699
- software 290
- Sony Music 1161

- sorting models 870, 875, 881, 884, 891  
 Sotheby's 113, 912, 929  
 sparse data 966  
 spatial concentration 785  
 Spears, Britney 904  
 special exhibitions 1038  
 specialist arts training 881  
 specification biases 966  
 spectrum constraints 593  
 spillover benefits 1134  
     *see also* externalities  
 split runs 1158, 1173  
 sponsorship 1011, 1041, 1185, 1193, 1266, 1279,  
     1292  
 sport 21, 452  
 spot price 889  
 Springsteen, Bruce 689  
 stable distributions 637–639, 655, 661–663  
 stable Paretian costs 650  
 stable profit laws 655  
 stage directors 774  
 star system 355, 658  
 state arts endowments 1291  
 state tax law 1292  
 state universities 1237  
 stated preference methods 997  
 states' rights 631  
 status competition 462  
 Statute of Charitable Uses 1301  
 stereotypes 366  
 stimulus goods 163  
 Strachey, Lytton 63  
 strange law of production 652  
 strategic release 298  
 strategic variable 607  
 strong sustainability 990  
 studio musicians 1089  
 subcontracting 778  
 subscription system 295, 427, 581, 590  
 subsidiarity 1213  
 subsidiary rights 544  
 subsidies 419, 481, 794, 1186, 1239  
 substitutability 989  
 substitute prices 436  
 substitutes 430, 449  
 substitution effect 466, 710  
 sunk costs 348, 536, 556  
 sunk investment 348  
 superdeductions 1286, 1289, 1293  
 superstar effect 558, 695  
 superstar museums 1036  
 superstar theory 780, 838  
 superstars 10, 141, 355, 835, 889, 895–905  
 supply of artistic creativity 886  
 supply of cultural goods 1079  
 supply-side interventions 1230  
 sure things 659  
 Survey of Public Participation in the Arts 1282  
 survivorship bias 919  
 sustainability 20, 153, 411, 990  
 sustainable development 988, 1055–1056  
 symphony orchestras 771  
 synchronization license 698  
  
 tacit knowledge 868, 883, 1107  
 talent 259, 521, 779, 867, 876, 878, 897–900,  
     903–905, 1067  
 talent agencies 773  
 talent differential 898, 900  
 target rate of return 96  
 targeted incentives 1264  
 taste 5, 44, 74, 174–181, 187, 188, 193–196, 444,  
     449  
 tax base 1267, 1269  
 tax concessions 1240  
 tax credit 1280, 1290–1292  
 tax deduction 1280, 1290  
 tax exemption 1271  
 tax expenditure 999, 1260, 1268–1275, 1281–  
     1288  
 tax incentives 1257–1264, 1269, 1274, 1285,  
     1293  
 tax policy 1256, 1266, 1267  
 teaching 796  
 teams 779  
 technical innovations 783  
 technical performance criteria 455  
 technological change 404, 513, 1080, 1081, 1165  
 technology 287, 1067  
 television 549, 1082  
 Television Without Frontiers (TWF) 1172  
 television writers 822  
 Thatcher, Margaret 1318  
 theater 514, 634  
 theme parks 1166  
 Thirty Years War 127  
 three-stage least squares estimator 973  
 ticket prices 494, 688  
 time allocation 435, 1238  
 time flies principle 734, 738  
 timeliness 951  
 timing of donations 1277  
 tolerance 1067



- Tolstoy, Leo 62  
 top-coding 830, 840  
 topiary 40  
 touring 452  
 touring exhibitions 1153  
 tourism 407, 992, 1005, 1036, 1039, 1062, 1063, 1166  
 tournaments 538, 779, 781  
 toys and games industry 556  
 tracking errors 970  
 trade 345, 405, 1082, 1152, 1165, 1177, 1178  
 trade liberalization 1149  
 trade policy 18, 1141–1180  
 trade statistics 1178  
 Trade-Related Aspects of Intellectual Property Rights Agreement (TRIPS) 1169, 1171  
 trademark 1111, 1114  
 trademark law 223  
 traditionalism 460  
 traditions 985  
 training 879, 881, 1237  
 transaction prices 426  
 transactions costs 360, 390, 773, 1007  
 transition 850  
 translations 1167  
 transparency 1161, 1169  
 travel cost method 996, 1076  
 trust 361, 362, 370, 883  
 Turgot, Anne Robert Jacques 4  
 two-sided markets 571, 579, 586, 594, 599, 603  
 two-stage Bayesian estimator 973  
  
 U-Musik 420  
 UIP (United International Pictures) 1161  
 uncertainty 273, 536, 541, 554, 559, 793, 1156, 1179  
 under-price tickets 428  
 underemployment 769  
 unemployment 769, 797, 798, 834  
 unions 797  
 United Kingdom 1009, 1280  
 United Nations Development Strategy 401  
 United Nations Educational, Scientific and Cultural Organization (UNESCO) 987, 1171, 1174–1177  
 United States 1289, 1291  
 United Way 1312  
 Universal Declaration of Human Rights 401, 1170  
 Universal Declaration on Cultural Diversity 403  
  
 universities 1237  
 urban and regional development 20, 1051–1093  
 urban planning 1069  
 urban regeneration 767, 1007  
 Uruguay Round 1169  
 US Copyright Act 697  
 US General Social Survey 465  
 United States of America 1010  
 use values 1072, 1076  
 utilitarians 151  
 utility function 488  
 utility maximizing framework 433  
  
 V-chip 605  
 valuation 99, 176–179, 189, 194–205, 248, 949, 1075, 1077, 1247  
 value 150–156, 245, 255  
     *see also* aesthetic value; bequest value; cultural value; economic value; existence value; option value; prestige value  
 value-added tax (VAT) 751, 1266  
 values 360, 363, 365, 391  
 van Gogh, Vincent 515  
 variance of artists' earnings 876  
 vaudeville 547  
 Vauxhall Gardens 139  
 Veblen, Thorstein 4, 64  
 Venice 1007  
 venture philanthropy 497, 1311  
 versioning content 1166  
 vertical and horizontal differentiation 545, 551, 553, 601, 787  
 vertical disintegration 773  
 vertical integration 550, 555, 559, 631, 632, 1160  
 video 259  
 viewer sovereignty 570, 588  
 visitor surveys 1073  
 Visual Artists Rights Act (VARA) 224, 225, 227–231  
 visual arts 213, 311, 317, 328, 771, 1079, 1233  
 volatility 378, 644, 1234  
 voluntarism 377, 1277  
 voluntary contracts 1032  
 voters 1136  
 vouchers 1200, 1270  
  
 wage dispersion 769  
 Wagner's law 1192  
 Wallace Funds 1306  
 war 404  
 Warhol 216, 219, 244  
 warranted trust 377

- waste 29
- waves of fashion 37
- weak and high centrality 458
- weak copyright 292
- weak sustainability 989
- wealth-elasticity of demand for art 79
- Weber, Max 793
- Weber thesis 391
- welfare economics 6, 594
- Welfare State 370, 1318
- Wexford Festival 1060
- willingness to accept 1077
- willingness to pay 997, 1076–1078
- winner takes all 641, 736, 786, 825, 841
- Wisconsin Artistic Endowment Foundation 1291
- women 1074
- woodblock print 317
- Woolf, Leonard 63
- Woolf, Virginia 63, 64
- work-preference model of artist behavior 875
- works-for-hire 889
- World Bank 401
- World Commission on Culture and Development 403
- World Culture Reports 1177
- World Expositions 1060
- World Heritage List 1000
- World Intellectual Property Organization (WIPO) 1171
- World Trade Organization (WTO) 1161, 1168, 1169, 1171, 1174, 1176, 1178
- World War II 1304
- Worpswede 1090
- writers 751, 822
- Writers Guild of America (WGA) 822
- X inefficiency in cultural services 1130
- youth bias 420
- zonal travel cost 452
- Zurich 1078