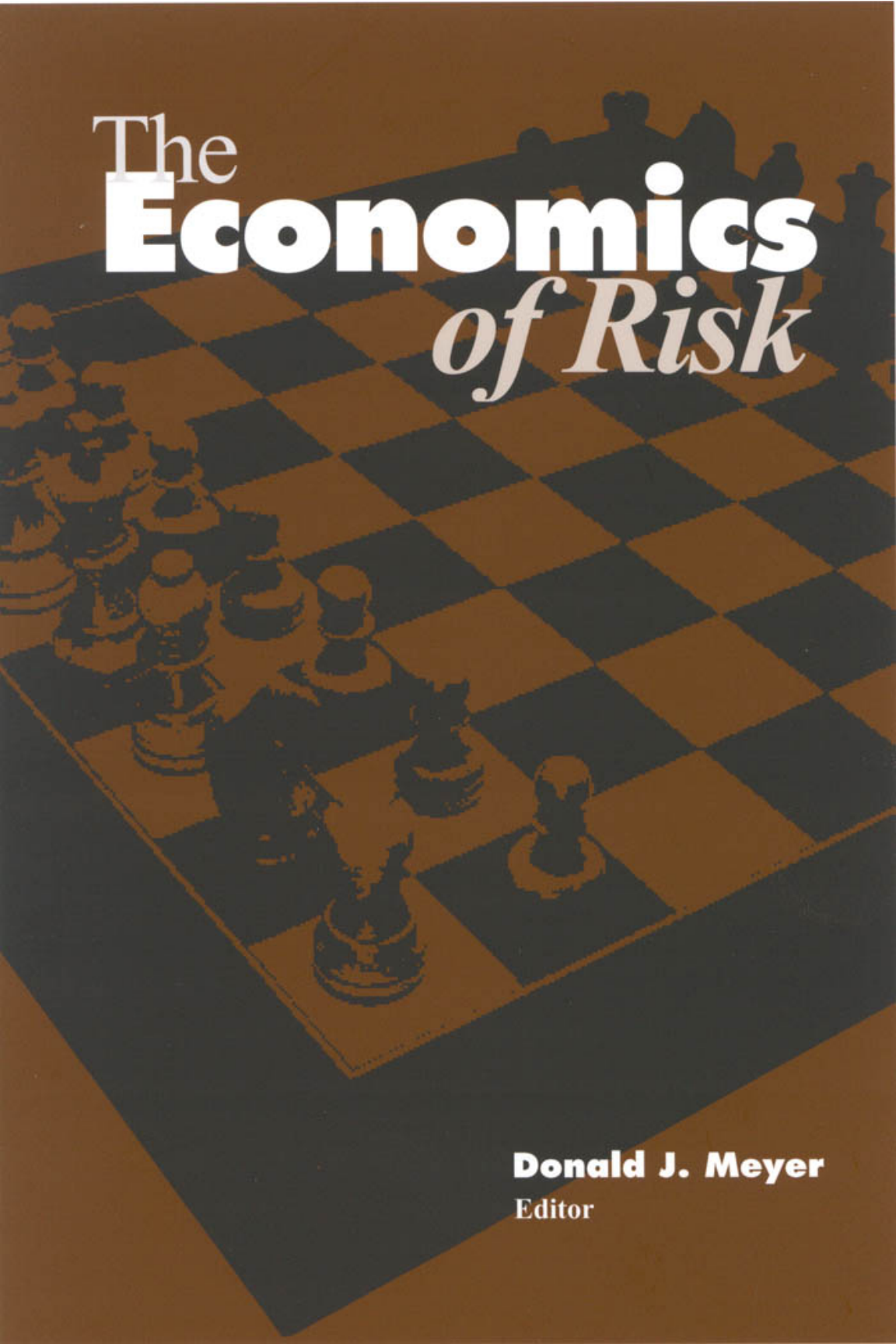
The Economics of Risk



Donald J. Meyer
Editor

The Economics of Risk

The Economics of Risk

Donald J. Meyer
Editor

2003

W.E.Upjohn Institute for Employment Research
Kalamazoo, Michigan

Library of Congress Cataloging-in-Publication Data

The economics of risk / Donald J. Meyer, editor. — 1st ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-88099-268-9 (hardcover : alk. paper) — ISBN 0-88099-267-0 (pbk. : alk. paper)
1. Risk. 2. Economics. 3. Risk management. I. Meyer, Donald J., 1957-
HB615.E3387 2003
368—dc22

2003020210

W.E. Upjohn Institute for Employment Research
300 S. Westnedge Avenue
Kalamazoo, Michigan 49007–4686

Cover design by Alcorn Publication Design.
Index prepared by Diane Worden.
Printed in the United States of America.
Printed on recycled paper.

Contents

Introduction

Donald J. Meyer
Western Michigan University

–Warning–

Under the Michigan Equine Activity Liability Act, an Equine Professional is not liable for an injury to or even the death of a participant in an equine activity resulting from an inherent risk of the equine activity.

Michigan Farm Bureau
Farm Bureau Insurance

This sign is posted at the horse stables where my two younger daughters ride horses on Saturday mornings. The sign communicates at least two distinct messages to its reader. First, riding horses is a risky activity. Even though it is a small percentage of riders overall, a number do get injured in horse-riding incidents with the possibility of even sustaining a serious or life-threatening injury. Second, when you decide to ride a horse at these stables, you take on and assume this risk; that is, the stable owner is not in any way liable for an injury that you may incur while participating in equine activities.

Riding horses is just one of almost countless situations in life in which you encounter risk—where you encounter the chance of injury, damage, loss, or of making a dangerous choice. Other examples include the risk of losing your job, the risk of contracting some debilitating disease, the chance of getting in an automobile accident on the way to the shopping mall, and the risk of being struck by lightning in a summer thunderstorm. Risk is all around us and affects us all. No one can escape from its clutches or attain perfect immunity to it. You may decide to not acquire the additional risk of riding a horse by not engaging in that activity, but one cannot so easily avoid the risk of getting cancer or the risk of being hit by a car while crossing the street on the way to school.

Risk is something that most of us dislike and try to avoid. Economists and other researchers have studied risk and have obtained considerable amounts of evidence that indicate that people are generally "risk averse" in their attitude towards risk. This means that, everything else held constant, people choose the less risky alternative and they will take measures to reduce or shift the risk to others when feasible. Purchasing health, life, or automobile insurance is one way to do this. Diversifying your investment portfolio by investing in different kinds of stocks, bonds, and money market securities is another. These measures help reduce the amount of risk that we must face to a more comfortable and manageable level. But we cannot completely insulate ourselves from the many risks that one can come across.

On occasion, people willingly seek to engage in a risky activity or situation instead of selecting the usual risk avoidance strategy. A slot machine or some other game of chance in a casino entices some people to gamble and take on risk. Other people invest a portion of their incomes in the risky stock market. This is not necessarily evidence contrary to risk aversion, because gambling can provide entertainment value or give an "adrenaline rush" to the gambler. Also, historically, the average return of the stock market has been higher than the return of safer money market assets, and this higher expected return can compensate for the higher risk of the stock market assets.

This book contains chapters that address various aspects of risk. Two chapters deal with risk directly by looking at risk management and how it is applied to decision making, or by assessing what researchers have learned over the last few decades in their theoretical investigations of risk. The other chapters look at risk indirectly by examining markets in which risk has a significant presence. Casino gambling enterprises, agriculture markets, auctions, and health insurance markets are places where risk makes a considerable impact. A number of problems that result from risk in these markets and in the economy will also be addressed. Auction participants may feel the sting of the "winner's curse" when the object they are bidding on has uncertain value. Significant health issues and potential problems face those who are without health insurance. Risk incentive problems have plagued the Farm Bill, government's response to farmers to decrease the risk in agriculture. Problem and pathological gamblers make up a percentage of those who enter a casino establishment.

The first chapter in this book is "Risk and Risk Management: Basic Concepts," by Keith J. Crocker. The main focus of this chapter is on risk management for the business firm and for the general consumer. Crocker discusses how one identifies risk and then how to deal with it. This is a fitting first chapter because it is a natural starting place in the investigation of risk, and it is very basic and applicable to every reader, regardless of how and where one fits into the economy.

Crocker goes through a number of steps involved with risk management, starting with risk identification and then quantifying the magnitude of the existing risk. Risk mitigation and control follows in the process. Crocker examines both loss prevention and loss reduction measures and looks at the decision of which risks to retain and which to transfer.

Crocker highlights this process with a number of interesting applications and examples. His central backdrop is the February 1999 natural gas explosion occurrence at the Ford Motor Company River Rouge power plant in Detroit, Michigan. Detailed information associated with events leading up to and immediately following this catastrophe are used to exemplify the presence of risk and of risk control and management. Asbestos exposure and its subsequent cleanup and the tainting of Tylenol capsules with cyanide also serve as illustrations in Crocker's discussion.

The second chapter, by Mark J. Machina, is entitled "States of the World and the State of Decision Theory." As the title suggests, Machina assesses the state of the profession regarding its theoretical investigation of risk and uncertainty. He reviews and discusses where we are in terms of the modeling and development of risk analysis. The major theoretical risk research of the last several decades is divided into two major approaches.

Choice under objective uncertainty was the first theory developed about decision making under risk and comprises Machina's initial discussion. Its roots go back several centuries to early work by Pascal and Fermat. Objects of choice can be described as objective lotteries in which all outcomes and the objective probabilities of these outcomes are known. An individual's preference function over these lotteries is generally assumed to follow the objective expected utility form for some von Neumann-Morganstern utility function. Violations of the expected utility hypothesis, including the famous Allais Paradox, are

noted, and Machina discusses some nonexpected utility models that have developed as a response to these violations and paradoxes.

Choice under subjective uncertainty is the other main branch of theoretical risk research. Machina traces this theory back to work by Savage, who was instrumental in the formulation and early development of the subjective uncertainty model. The main tenets of this approach include states of nature, events, and outcomes or consequences. Individuals are allowed to have probabilistic beliefs, and these beliefs can differ across individuals. Thus, under this theory you and your friend can differ as to the assessed chance that the stock market will rise in the next few weeks. Machina then addresses violations of this theory including the Ellsberg Paradox and subsequent modeling adjustments in response to the violations.

After careful discussion of each of these theories, Machina offers his personal insights into the similarities and differences between them and how the approaches actually are more related than one might initially assess. Included in his observation is an intuitive discussion of his recently completed work on this subject. For more details, the interested reader is invited to consult Machina's fascinating developments and findings.

The third chapter is "Gambling with the Future: Economic and Social Perspectives on Casinos in America," by William R. Eadington. Strong demand for gambling activities for entertainment or risk-seeking value has existed as long as mankind has. Gambling has gone the gambit from an activity that has been largely banned and viewed by many as immoral to a more widely accepted, legalized, and controlled setting in which special interest groups compete for the industry revenues. Casino gambling is a significant and fast-growing industry, one whose growth often occurs in economic downturns when tax revenue generation and job creation are highly desirable. In this chapter, Eadington focuses on the economics of casinos, establishments that house numerous varieties of games of chance.

Eadington starts by tracing through U.S. casino history of the last century, beginning with Nevada's casino legislation in 1931. He examines the current status of the industry as well as looking at projected trends into the future, including new gambling forms such as Internet gambling. Different casino markets are described, including destination resort casinos such as Las Vegas, Atlantic City, or Biloxi, urban or

suburban casinos located in major metropolitan areas, and rural casinos which include most tribal casinos. Gambling is an ever-changing industry, as evidenced by the creation of "racinos"—horse race tracks that have become equipped with slot machines or other gaming devices.

Eadington explores a social as well as economic perspective on casinos by looking at cost–benefit analysis of casinos. The benefit side is well founded in consumer surplus theory in economics. Consumers benefit by having access to a legalized casino establishment, and many would be willing to pay money to do so. The cost side is much less developed and is harder to quantify. Problem and pathological gamblers impose a cost on themselves and others. Other researchers have tried to link casinos or gambling with increased crime rates, gambler financial troubles and higher divorce and suicide rates, and a general decline in the nation's "moral fiber."

The fourth chapter is by John H. Kagel and is entitled "Common Value Auctions and the Winner's Curse: Lessons from the Economics Laboratory." In this chapter, Kagel discusses the risk of experiencing the winner's curse in a common value auction, where the auctioned item's value is the same to all bidders but is unknown (risky) at the time of the bid. Bidding on an offshore oil tract is a fitting example, as the precise value of the hydrocarbons beneath the ocean floor is uncertain at the time the bids are placed.

Generally bidders obtain distinct private indicators or signals of the object's value. Some of the signals will be higher than this value while others will be lower. The winning or high bidder likely has the highest or one of the highest signals. If the bidder doesn't recognize this and factor it accordingly, the submitted bid may be more than the uncertain value even though it is less than the signal it is based on. If this happens, the winner is said to be "cursed," and below-normal profits, even losses and bankruptcies, can result.

Kagel traces through the history and early reporting of the winner's curse in auctions and in other markets. It is important to recognize that the winner's curse is not a theoretical, equilibrium concept or result but rather is a hypothetical empirical phenomenon, indicating that bidders do not properly account for receiving private signals in the bidding process. Kagel discusses the considerable evidence of the curse, focusing on that of sealed bid auctions. He also examines alternative hypoth-

eses for explaining the overly aggressive bidding behavior such as limited liability for losses. Auction experience is found to be an important learning variable, as inexperienced bidders are generally most susceptible to the curse.

Kagel then looks for the winner's curse in the context of English auctions and first price auctions with insider information. These market types were selected because they are environments that may eliminate or at least sharply reduce the winner's curse effect. Kagel compares the data results in these markets with earlier findings of sealed bid auctions. He concludes that evidence for the winner's curse is present in these markets, although the magnitude seems to be less severe.

Chapter 5 is "Sharing High Risks: How Government can make Health Insurance Markets more Efficient and more Accessible," by Katherine Swartz. In this chapter Swartz looks at the characteristics of those in the economy who are without health insurance and then suggests how to make these health insurance markets more accessible and more efficient. Recall from earlier discussion that the typical decision maker in our economy is risk averse and desires insurance to help reduce the risk of undesirable events. To be without health insurance is not only bad for the uninsured individual but for the general economy as well.

Swartz begins by describing the traits of the uninsured, including that of age, income, labor force status, and health status. The typical uninsured person is young, has low income, is working, and is in good health. She goes on to examine the insurance company or seller side of the market. Health insurance markets can be divided into three different types: large employer group, small group, and individual. The small group and individual markets significantly differ from the large employer group, and this difference is critical for the occurrence of the uninsured.

The small and individual markets suffer from what is known in the risk literature as the *adverse selection problem*. The pool of people seeking health insurance is comprised of those who are at higher risk for illness, disease, and sizable medical bills, and those that are of lower risk. The insurance applicant generally knows much more about the size of this risk than does the insurance company. Insurance companies fear that those applying for coverage are *dis*proportionately composed of the high risk or high cost group. This can lead to significant

losses for the insurance companies. This forces companies, Swartz suggests, to compete for the lower risk or relatively healthy individuals, and the high risk or sickly people subsequently get left out and tend to be uninsured.

Swartz proposes that the government should step in and cover the 2–3 percent of insured cases with the highest medical costs. This extreme cost segment makes up a large fraction of the costs paid out by insurance companies. By taking these cases over, the government frees up resources expended by insurance companies to try to screen out the cases expected to be high cost. This, Swartz argues, will enhance overall insurance market efficiency and also accessibility for those seeking health insurance coverage. She models her plan after a recently developed and implemented plan in New York state.

The sixth and final chapter is by Rulon D. Pope and is entitled "Risk and Agriculture: Some Issues and Evidence." Adverse weather, disease, and damaging insect pests are just some of the risks farmers face during growing season that could result in uncertain crop yields. Product price is also risky, as many agricultural markets are atomistic in nature and crop prices are subject to changing market supply and demand conditions. In his chapter, Pope highlights some central concepts pertaining to risk in agriculture.

One of Pope's primary concepts of interest is that of diversification as a response to risk. Diversification can be accomplished in a number of ways; for example, crop diversification, area diversification, or the diversification of labor income away from the farm. Pope also looks at risk reduction and input usage. He discusses hedging and forward markets as a means of risk reduction. All of these actions are consistent with farmers who exhibit risk aversion, and Pope cites empirical evidence in support of this position. Finally, food safety is addressed as a concern to both farmers and consumers alike.

Part of Pope's analysis is devoted to government support programs and crop insurance. The government programs have not proven to be financially successful, because there is generally not enough money collected in insurance premiums to cover the crop loss payments. Two problems—adverse selection and moral hazard—can help explain why the government may have financial difficulty. Katherine Swartz discusses in Chapter 5 the adverse selection problem as it pertains to health insurance markets, which applies here as well. Moral hazard is

the idea that farmers may take fewer precautions to avoid risk when they hold insurance than when they do not. This can result in more crop losses than expected and more payouts by the government.

The six chapters in this book look at a broad array of research relating to risk. The authors convey much information about risk as it pertains to the various markets that they address. A comprehensive list of references accompanies each chapter to guide the interested reader who wants to pursue some particular facet in more detail. Learning more about what risk is and how it affects us reduces some of the uncertainty in life that we all face, and then helps us make more informed decisions. We will never be able to eliminate all the risk that we can potentially encounter, but we can strive to better understand the risks involved and then deal with them in the best light possible.

1
Risk and Risk Management

Basic Concepts

Keith J. Crocker
University of Michigan Business School

"This has got to be the worst day of my life," observed William Clay Ford Jr., Ford Motor Company Chairman, as he contemplated the February 1999 natural gas explosion in boiler number six that had just leveled part of the River Rouge powerhouse in Detroit, Michigan. The disaster killed 6 people and seriously injured 14, and cut off power to the 1,100 acre facility.

While Ford Jr.'s remarks were directed toward the human dimension of the tragedy, from a corporate standpoint the prognosis must have appeared equally sobering. The Rouge complex powerhouse—the centerpiece of Henry Ford's dream of building entire cars in a single location—had supplied electricity, compressed air, mill water, and steam to six assembly and parts plants employing 10,000 workers, and also to the independently owned Rouge Steel plant. Although an engineering marvel of its time, the concentration of production at River Rouge had precipitated a risk manager's worst nightmare, as the effects of the integrated plant's shutdown rippled through Ford's internal supply network.

First hit was Rouge's own Mustang assembly plant, which had been working overtime with two 10-hour shifts daily cranking out the popular sports compact. Next came Rouge's metal stamping plant, supplying metal parts (fenders and similar products) to 16 of Ford's 20 North American plants. Results were predictable. Shifts were cut from 8 hours to 4 hours at three Midwest assembly plants, and lost production at Rouge's frame plant resulted in the elimination of scheduled overtime at truck plants in Kansas City, Missouri; Norfolk, Virginia; and Oakville, Ontario.

Even at these reduced levels, production was supported only by the buffers of existing inventories and supplies in transit that, once exhausted, would necessitate plant shutdowns. And a previously planned $240 million replacement powerhouse would not be completed for at least a year (*Financial Times* 1999).

Risk is endemic to our personal, as well as professional, experiences. Every time we decide to cross the street or ascend the stairs in our homes, we are making personal decisions involving risks and their management. How we handle these situations has an important impact on the quality (and, in many cases, the length!) of our lives.

WHAT IS RISK?

Webster's dictionary defines risk as "the chance of injury, damage, or loss." Unlike, say, a portfolio of stocks, which has a potential for gain, risks present only a down side. A risk is a chance of something bad occurring and, hence, to be avoided. Of course, even bad things can provide a profit opportunity to somebody—the city taxes me to haul away my garbage, thereby providing employment, and the "Orkin Man" is happy to fumigate my house, for a fee. But I do not generally bring home extra garbage or encourage termites to infest my house. Nor do sensible people seek out risk. However, risk can be managed. This chapter lays out the key elements of risk management: identification and quantification, mitigation and control, financing, and catastrophe planning.

IDENTIFICATION AND QUANTIFICATION

Given that risks are endemic in our uncertain world, adopting appropriate strategies to deal with risk exposures and their consequences is an everyday task. Consider the case of the pedestrian contemplating crossing a busy street. The first step is to identify the risk (speeding automobiles with distracted drivers chatting on cell phones?) and to quantify its magnitude (scrapes? bruises? broken bones? fatali-

ties?). This mundane task is the critical point of departure for one crafting a risk management strategy—remember the old aphorism that "forewarned is forearmed," which is probably the best piece of cheap advice that a risk manager can give.[1]

In the business setting, many kinds of risk are identifiable, even to the most uninitiated. Dangerous machinery or exposed electrical wiring in a factory setting, or slippery floors in an office or retail establishment (squashed grapes on the floor are a grocer's nightmare) are obvious examples. Other types of risk exposures may be less apparent and discernible only to those with experience in a particular area of risk analysis. Much as standing under a tree during a thunderstorm may seem reasonable to those unfamiliar with lightning, risk exposures may not be apparent to an untrained eye.

In the case of the Ford Rouge power plant, for example, there were certainly engineering advantages associated with the consolidation of production of the electricity, steam, and high-pressure air required by the entire Rouge complex. But the risks of this approach also turned out to be substantial, as the events of February 1999 attest.

Perhaps the most insidious risks facing businesses these days, however, come from evolving legal rules, as we have observed in the case of environmental liability and asbestos exposure. The Comprehensive Environmental Response, Compensation and Liability Act, the 1980 Superfund hazardous substance clean-up legislation, introduced strict liability that may involve several entities jointly for cleaning up hazardous waste sites. As a consequence of this new legal reality, a business could have been in full compliance with all applicable laws at the time of the waste disposal, or simply be the current owner of an existing site, yet still be strictly liable for the costs of clean-up. Even partial contributors to the site are fully liable for the entire cost of clean-up, due to joint and several liability,[2] leading to the predictable prospecting for "deep pockets" by enterprising tort attorneys. These liabilities also may be inherited, which makes mergers and acquisitions problematic these days.

Asbestos exposure also provides an instructive example. Fifty years ago, most people had little understanding of the health risks associated with airborne asbestos fibers in the workplace, and exposure standards reflected this. Over time, however, it became increasingly clear that asbestosis (a close cousin of the black lung disease suffered

by coal miners) and mesothelioma (an untreatable cancer of the lung or stomach lining that is both swift and invariably fatal) were associated with workplace exposures. The result has been an explosion of litigation (estimated potential: 1.3 to 3.1 million claims) with expected asbestos liabilities of $200 billion, of which $78 billion will be borne by the affected companies and the rest by their insurers (Parloff 2002).

Litigation has already destroyed the primary producers of asbestos—Johns-Manville, Unarco, and Raybestos Manhattan all declared bankruptcy long ago—and has moved on to bankrupt companies that merely purchased asbestos products, including Babcock & Wilcox, Owens Corning, GAF, and W.R. Grace. Currently in the crosshair of asbestos litigation are Georgia-Pacific (involving gypsum products), 3M (for allegedly failing to warn that the dust masks wouldn't work if improperly used), and Ford (for exposures related to the asbestos used in brakes). Federal-Mogul Corp., an automotive supplier, recently sought Chapter 11 bankruptcy protection because of an asbestos liability inherited from its 1998 acquisition of T&N PLC of Manchester, England, a company that had used asbestos in a separate building supplies business. At the time of the acquisition, Federal-Mogul set aside $2.1 billion in cash to cover the anticipated claims, a sum that in retrospect seems to have been nowhere near enough.

Daniel S. Sobczynki, the former Director of Corporate Insurance for Ford, put it best: "The highest potential risks are those that are unidentified and unmanaged. It is critical to evaluate your risks and to learn from the lessons of others," he says. "The problem of learning from personal experience is that it gives you the lesson after the test has been administered" (*Financial Times* 1999).

MITIGATION AND CONTROL

After the risk exposure has been assessed, the next step is to consider how one deals with it. Continuing with our street-crossing example, one possibility would be to avoid the risk entirely and not cross the street at all (a wise strategy if the road in question were, say, Interstate 94 at rush hour). Alternatively, if we decide to proceed, the question might be the following: do we jaywalk and cross the street now, or

stroll down to the traffic signal and wait for the green light? Each of these alternatives represents an economic decision, weighing the cost of the strategy against the potential benefits.

Generically, mitigating a risk exposure entails the identification of tactics either to reduce the probability of a bad outcome, or to reduce the magnitude of a loss, should a bad outcome occur. The former types of activities, referred to as *loss prevention* measures, would include the cross-at-the-intersection option discussed above, or, in a more mundane industrial setting, the inspection of electrical wiring to reduce the probability of an electrical fire. Indeed, most of the risk mitigation strategies that come easily to mind are designed to keep us out of trouble in the first place—don't put the gasoline can next to the furnace, don't smoke in bed, lock your doors before you retire for the night. *Loss reduction*, on the other hand, describes the class of risk mitigation activities designed to reduce the magnitude of a loss, should one occur. The standard example here would be the installation of sprinklers in a warehouse, which doesn't reduce the probability of a fire starting but, rather, mitigates the damages that result from the fire.

The explosion of boiler number six at the River Rouge powerhouse occurred during a maintenance shutdown. As far as can be determined, a valve unintentionally left open allowed natural gas to flow into the boiler, which was quickly ignited by the electrostatic scrubbers located in the boiler's chimney.

In retrospect, it appears that the tragedy stemmed from a lack of attention paid to issues of risk mitigation during routine episodes of maintenance. Not only was the act of shutting down the boilers rare, but apparently there were no written procedures or checklists to guide the process. Employees who had not been trained in shutting off the boilers and who had last received an equipment manual in 1997, had to shut off over 30 (unlabeled) natural gas valves throughout the powerhouse complex. They missed one, and the rest is history.

We make trade-offs in our personal and business lives between the burden of risk exposure and the cost of risk mitigation. Financing the costs associated with a bad outcome becomes the question. In personal settings, the risk financing strategy generally adopted is that of risk shifting to a third party, usually an insurance company (think about the collision and liability insurance on your car, homeowner's insurance, or the warranty on a new appliance). The problem with this type of risk

transfer, though, is that it creates what is known in economics as a "moral hazard."

A colleague of mine kept a sailboat moored off the end of his dock on Long Island Sound. One day, during casual conversation, I asked about his strategy for dealing with storms and the like—as a boat owner myself, I was aware (risk identification and quantification) of the effects of heavy wave action on a boat banging against a dock. He responded that he wasn't worried because he had insurance and he never took the boat out of the water until the end of the season. The problem here, of course, is that if one is fully insured against a loss, then one has no incentive to take (privately costly) actions to reduce one's risk exposure. Insurance companies, not surprisingly, have figured this out.

When my teen-aged son finally made enough money to purchase a car, it turned out that the machine of his dreams was a 1994 Camaro Z28, with a 5.7 liter V-8 engine and 270 horsepower. You might think that no insurer in their right mind would write coverage in a situation like this, but you would be wrong. An automobile insurer in Michigan was willing to provide liability coverage at a finite premium. But, there was a catch—no coverage for collision damage.[3] Effectively, he has a 100 percent deductible if he wraps the car around a tree.

This retained risk has "incentivized" my son to drive carefully. This is generally the trade-off that you will find in your personal and professional risk financing decisions—increased investment in risk elimination reduces the premiums you pay per dollar of coverage, but the down side is that you are exposed to more risk.

CATASTROPHE PLANNING

Accidents do happen despite the best intentions and most effective efforts to forestall such eventualities. And the response to the bad news is probably the most critical component of any *loss reduction* strategy.

In the immediate aftermath of the Rouge River powerhouse catastrophe, William Clay Ford Jr. dispatched his personal aide, with credit card in hand, to track down the victims' families and do whatever was required to help out. The company worked with its suppliers to procure

electrical switching equipment and to obtain portable boilers for steam. Detroit Edison built an outdoor substation—in a week—to supply the power necessary to get the Rouge River complex back on line. The result was a triumph in loss reduction—a potentially catastrophic business interruption scenario truncated to a one-week hiccup on the production line.

There are many other examples of the importance of catastrophe planning, good and bad. For example, back in 1986, when a still unidentified individual replaced the painkiller in several bottles of Tylenol capsules with cyanide, the result was the death of an innocent consumer. Johnson & Johnson, the maker of Tylenol, didn't attempt to deflect blame (after all, they hadn't adulterated the capsules) or otherwise temporize. They immediately recalled all the capsules from store shelves—even those that were clearly untainted—and then designed the generation of tamper-proof containers still in use today. This is a textbook loss-reduction strategy—timely, aggressive, and (while costly in the short run) effective.

In contrast, consider the strategy of Johns-Manville, once the world's biggest producer of asbestos, which, as we noted earlier, collapsed under the weight of litigation from asbestos claims in 1982. Johns-Manville's apparent decision to ignore the risks of asbestos exposure to its workers, long after the evidence indicated that management may have suspected a link between asbestos exposures in the workplace and worker health, resulted in lives ruined and lost. The cost to Manville and its shareholders was ultimately that of corporate bankruptcy.

Dan Sobczynski offers some sound advice: "Either manage the risk, or it will manage you," he says, "and, when it does, the loss will happen when you are least prepared" (*Financial Times* 1999).

Notes

1. Students of history will recall that Winston Churchill was almost killed by a speeding taxi in New York City during the 1930s. Accustomed to cars driving on the left side of the road, he looked the wrong way while crossing the street, a clear failure in risk identification and quantification.
2. Joint and several liability means, in practice, that even a 1 percent ownership stake in the property can lead to liability for 100 percent of the clean-up costs if

the owners of the other 99 percent interest are financially unable to pay their share.

3. Actually, they would provide such coverage, but at an annual premium effectively equal to the book value of the car!

References

Financial Times. 1999. "Managing Risk before it Manages You." The *Mastering Strategy* series, October 11, pp. 8–9.

Parloff, Roger. 2002. "The $200 Billion Miscarriage of Justice." *Fortune* 145(5): 154–196.

2
States of the World and the State of Decision Theory

Mark J. Machina
University of California, San Diego

TWO EXTRAORDINARY NONSCIENTISTS

Almost 20 years ago, I briefly knew a man by the name of Craig. Although he died about a year after I met him, I've thought about him ever since. Craig had this uncanny ability to converse with a person for a few minutes, and then announce what make and model of car they drove. Neither I, nor anyone I ever spoke to, had ever seen him get it wrong. Craig was never able to explain how he did it, and his unique ability followed him to the grave.

What Craig had perfected was an impressive skill—perhaps even an art—but it was not science. It was not science because it was not a procedure that he could verbally communicate or write down, so that other people in other places or other times could do it also. One of the defining features of scientific activity is that it generates a body of knowledge and techniques that *can* be communicated and utilized by others in this way.

I also knew a woman named Tula with an equally impressive ability. Tula was able to predict how well a person's day would go, based on the shape, size, and color of the *aura* they emitted in the morning. And in contrast to Craig, she could even explain the specifics of her method. For example, if your aura was round and blue, you would have good luck all day. But if it was square and yellow, then you'd best go back home and stay in bed. Tula had prepared a chart with the complete relationship between properties of your aura and the upcoming features of your day, so if you had a copy of the chart, you just needed a daily reading of your aura. Although Tula's success rate wasn't per-

fect (like Craig's was), it still compared favorably with standard medical, meteorological, and macroeconomic predictions, and most of her friends would stop by each morning for a quick reading of their aura, and then go away to consult their chart.

By constructing and distributing her chart, Tula had codified and communicated features of her technique in a way that Craig never could. But since Tula was the only one who could *see* these auras, what she was doing still was not science. An activity is not science unless it involves techniques that others can also apply *as well as* variables that others can observe.

The purpose of this chapter is to examine one of the most important theoretical constructs of modern decision theory—namely, the concept of *states of the world* or *states of nature*—from the point of view of these and similar scientific considerations. Are states of nature inherently descriptive or prescriptive objects? Do individuals making choices under uncertainty *face* these states of nature, or do they *create* them? Are states external and independently observable, like an individual's commodity demand levels, or are they internal and not directly observable, like utility or marginal utility levels? In addressing these questions, I will offer an overview of how researchers have sought to represent the concept of uncertainty, from the original formulation of probabilities and "objective uncertainty" in the seventeenth century, through Leonard Savage's twentieth century formulation of states of nature and "subjective uncertainty," to current work which seeks to eliminate—or at least redefine—the distinction between objective and subjective uncertainty. The following section presents some scientific issues common to all theories of choice, whether under certainty or uncertainty. The next two sections sketch out the current theories of choice under objective and subjective uncertainty. After that, I address the question of whether states of nature should be considered descriptive or prescriptive constructs, and then I consider scientific issues related to the observability and measurement of states of nature. The final section concludes with current work on the relationship between subjective and objective uncertainty.

SCIENTIFIC CONSIDERATIONS IN THE THEORY OF CHOICE

Scientific Modeling "From the Outside In"

The human decision-making process may well be one of the most complicated systematic phenomena in the universe. In terms of the point of view of the scientific observer, it is certainly unique. On the one hand, a scientist trying to model this process is like an anatomist in the days before anesthesia and vivisection—scientists can observe and to some extent even control external influences on a system, and can observe the resulting behavior of the system as a whole, but they cannot "get inside" to observe its constituent parts at work. On the other hand, every scientist is a human decision maker with powers of self-consciousness and self-reflection. However, self-reflection of our decision-making processes has not produced that much more "hard science" than has, say, self-reflection of our breathing or digestive processes.

While advances in neuroscience may ultimately do for decision theory what vivisection did for anatomy, decision theory currently remains very much a "black box" science. Although decision theorists can (and do) use introspection to suggest theories and hypotheses, the rigorous science consists of specifying mutually observable independent variables (in particular, the objects of choice available for selection), mutually observable dependent variables (the selected alternative), and refutable hypotheses linking the two. In other words, choice theory attempts to explain why particular alternatives are selected from a set of available choices.

Issues of Observability

Because decision scientists cannot perform dissection, they are subject to a greater scientific discipline than that required of anatomists. If a decision scientist tried to account for an individual's purchases of bananas as the direct result of something like an "appetite for fruit," we would not know how to test this hypothesis—that is, we would not know how to independently "look for" such an appetite, even if we had a scalpel and an open, anesthetized brain. Such unob-

servable constructs like appetites, utility, and preferences can—and do—play a role in scientific decision theory, but only as inside links in a causal chain that ultimately starts with fully observable independent variables and ultimately ends with fully observable dependent variables. For example, given the joint hypothesis that well-defined commodity preferences exist and are also stable from day to day, standard consumer theory allows us to infer enough information about these preferences from an individual's past demand behavior to be able to make refutable predictions about their future demand behavior, even for some combinations of prices and income never before observed.

In the following sections, we shall see that in passing from choice over certain commodity bundles to choice over uncertain prospects (either "objective lotteries" or "subjective acts"), hypotheses involving the unobservable constructs of commodity preferences and utility functions can be replaced by hypotheses involving the unobservable constructs of risk preferences and beliefs, which also link observable independent to observable dependent variables. Whether the notion of "states of nature" can similarly serve remains to be discussed.

Issues of Classification

In order for a variable or phenomenon to satisfy the criterion of "scientific observability," it is not enough that more than one scientist be able to see it—it is not even enough that a camera be able to record it. Rather, a variable is only scientifically observable if independent observers can agree on their description of what they have just observed. Thus, while a scientist can photograph facial expressions, they cannot be said to have photographed expressions of emotion unless there is a well-defined specification of which expressions correspond to each emotion, and independent observers predominantly agree in their assignment of emotions to each photograph. In other words, scientific observability requires well-defined and commonly accepted classification schemes for the observations, sufficient for grouping and comparing such observations, and relating them to general hypotheses and theories.

Just as different types of variables can have different degrees of observability, different classification schemes will have different degrees of common agreement. Thus, in regular consumer theory, we

are much more prone to classify commodities and define preferences in terms of category schemes like {“fruits,” “vegetables,” “grains”} compared to schemes like {“delicious foods,” “filling foods,” “unpleasant foods”}. Although the latter scheme is in some sense much more directly connected to any given individual’s preferences than the former scheme, the latter scheme cannot be defined independently of the particular consumer being studied. Since foods cannot be classified according to this latter scheme prior to observation of the consumer’s (verbal or choice) behavior, it cannot be used as a classification scheme for independent variables. Categories like “delicious foods,” “unpleasant foods,” etc. can be defined for *dependent variables,* however, either on the basis of the consumer’s verbal expressions, or on the basis of their past purchases or consumption behavior. Thus, whether a given classification scheme does or does not satisfy the criterion of scientific observability may well depend upon whether the scheme is intended to be applied to the independent variables or to the dependent variables of a theory.

Issues of Measurability

The above example of classifying facial photographs into different categories of emotions is an example of a *qualitative* classification of the basic observations. Although qualitative categories and qualitative variables are perfectly valid in the physical, biological, and social sciences, theories and hypotheses are most powerful when they involve *quantitative* independent and dependent variables. Many economists are of the opinion that economics has a more impressive scientific track record than anthropology because economists work with numerical variables such as prices, quantities, and income, rather than with qualitative variables like trust, group identification, or loyalty. Most theories and hypotheses involving quantitative independent and dependent variables are easier to test, to fine tune, and if necessary, to revise, than most theories and hypotheses involving qualitative variables.

Is uncertainty an inherently qualitative or quantitative construct? In the following sections we shall see that one of the two primary methods of representing uncertainty—the so-called “objective approach”—represents uncertainty quantitatively, via numerical probabilities. On the other hand, the other primary method—the so-called “subjective

approach"—has traditionally represented uncertainty in a qualitative manner, via an unstructured set of states of nature. However, in the final section of this paper, we see that taking a measurable, quantitative approach to subjective uncertainty can enhance its power, and in many senses can serve as an almost complete substitute for what may be considered the more ad hoc assumptions made about the world in the objective approach.

CHOICE UNDER OBJECTIVE UNCERTAINTY

Outcomes, Probabilities, and Objective Lotteries

The earliest formal representation of uncertainty came from founders of modern probability theory such as Pascal and Fermat. In this approach, the uncertainty attached to any event is represented by a numerical probability p between 0 and 1. Because probability theory derived from the study of games of chance that involved virtually identical repeated events, such probabilities were held to be intrinsic properties of the events in the sense that an object's mass is an intrinsic property of the object. These probabilities could either be calculated from the principles of combinatorics, for an event such as being dealt a royal flush, or measured by repeated observation, for an event like a bent coin landing heads up.

For an individual making a decision under objective uncertainty, the objects of choice are *objective lotteries* of the form $\mathbf{P} = (x_1,p_1;\ldots;x_m,p_m)$, which yield outcome x_i with objective probability p_i, where $p_1 + \ldots + p_m = 1$. The theory of choice under uncertainty treats lotteries in a manner almost identical to the way it treats commodity bundles under certainty. That is, each individual's preferences over such lotteries can be represented by a real-valued *preference function* $V(\cdot)$, in the sense that for any pair of lotteries $\mathbf{P}^* = (x_1^*, p_1^*;\ldots; x_{m^*}^*, p_{m^*}^*)$ and $\mathbf{P} = (x_1,p_1;\ldots;x_m,p_m)$, the individual prefers $\mathbf{P}^*$ over $\mathbf{P}$ if and only if $V(\mathbf{P}^*) = V(x_1^*, p_1^*;\ldots; x_{m^*}^*, p_{m^*}^*)$ exceeds $V(\mathbf{P}) = V(x_1,p_1;\ldots; x_m,p_m)$, and is indifferent between the two lotteries if and only if $V(\mathbf{P}^*) = (x_1^*, p_1^*;\ldots; x_{m^*}^*, p_{m^*}^*)$exactly equals $V(\mathbf{P}) = V(x_1,p_1;\ldots;x_m,p_m)$.[1]

The Expected Utility Hypothesis

In standard consumer theory, the preference function over commodity bundles is typically assumed to have certain mathematical properties but is typically *not* hypothesized to take any specific functional form, such as the Cobb-Douglas or Constant Elasticity of Substitution form. Specific functional forms are typically only used when absolutely necessary, such as in empirical estimation, calibration, or testing.

In contrast, the standard theory of choice under objective uncertainty typically *does* assume (or does assume axioms sufficient to imply) a specific functional form for the individual's preference function over lotteries, namely the objective expected utility form $V_{EU}(x_1,p_1;...;x_m,p_m) = U(x_1)\cdot p_1 + \ldots + U(x_m)\cdot p_m$ for some von Neumann-Morgenstern utility function $U(\cdot)$. Mathematically, the characteristic features of this functional form are that it is additively separable in the distinct (x_i,p_i) pairs, and also that it is linear in the probabilities. The term "expected utility" arises since it can be thought of as the mathematical expectation of the variable $U(x)$ (the individual's "utility of wealth") if wealth x has distribution $\mathbf{P} = (x_1,p_1;...;x_m,p_m)$. The literature on choice under uncertainty has generated a number of theoretical results linking the shape of the utility function to aspects of the individual's attitudes toward risk, such as risk aversion or comparative risk aversion for a pair of individuals. Excellent discussions of the foundations and applications of expected utility theory can be found in standard graduate level microeconomic texts such as Kreps (1990, Chapter 3), Mas-Colell, Whinston, and Green (1995, Chapter 6), and Varian (1992, Chapter 11).

Violations of the Expected Utility Hypothesis

Although the expected utility model is sometimes viewed as being quite flexible (since the von Neumann-Morgenstern utility function could have any shape), it does generate refutable predictions. Unfortunately, there is a growing body of evidence to suggest that individuals' preferences over lotteries tend to systematically violate some of these predictions. Risk preferences tend to systematically depart from the expected utility property of linearity in the probabilities. The most

notable example of this is the well-known Allais Paradox (Allais 1953), which asks individuals to rank each of the following pairs of lotteries (where \$1M denotes \$1,000,000):

$$a_1 : \{\ 1.00 \text{ chance of } \$1\text{M} \quad \text{versus} \quad a_2 : \begin{cases} 0.10 \text{ chance of } \$5\text{M} \\ 0.89 \text{ chance of } \$1\text{M} \\ 0.01 \text{ chance of } \$0 \end{cases}$$

$$a_3 : \begin{cases} 0.10 \text{ chance of } \$5\text{M} \\ 0.90 \text{ chance of } \$0 \end{cases} \quad \text{versus} \quad a_4 : \begin{cases} 0.11 \text{ chance of } \$1\text{M} \\ 0.89 \text{ chance of } \$0 \end{cases}$$

Experiments by Allais and others have found that the modal (and in some studies, the majority) choices are for a_1 over a_2 in the first pair, and a_3 over a_4 in the second pair. However, a preference for a_1 in the first pair implies that the utility function satisfies the inequality $0.11 \cdot U(\$1\text{M}) > 0.10 \cdot U(\$5\text{M}) + 0.01 \cdot U(\$0)$, whereas a preference for a_3 in the second pair implies $0.11 \cdot U(\$1\text{M}) < 0.10 \cdot U(\$5\text{M}) + 0.01 \cdot U(\$0)$, which is a contradiction.

Although the Allais Paradox was originally dismissed as an isolated example, subsequent work by MacCrimmon and Larsson (1979), Kahneman and Tversky (1979), and others have uncovered a qualitatively similar pattern of departure from the expected utility hypothesis of linearity in the probabilities, over a large range of probability and payoff values (see Machina 1983, 1987 for surveys of this evidence).

NON-EXPECTED UTILITY MODELS OF RISK PREFERENCES

Responses to the above-mentioned violations of the expected utility hypothesis have taken two forms. One branch of the literature has proceeded by positing more general functional forms for the preference function (Edwards 1955, 1962; Kahneman and Tversky 1979; Chew 1983; Fishburn 1983; Quiggin 1982; and Yaari 1987). Such forms accommodate most of the observed departures from linearity in the probabilities, and, given the appropriate curvature assumptions, can

also exhibit standard features like risk aversion, comparative risk aversion, etc.

A second line of work in non-expected utility theory proceeds in a manner closer to that of standard consumer theory—rather than adopting some new functional form, it generalizes the expected utility property of linearity in the probabilities to its natural extension of smoothness in the probabilities (e.g., Machina 1982). That is, it treats the preference function $V(x_1,p_1;...;x_m,p_m)$ as a general smooth function, and studies how properties of its probability derivatives relate to attitudes toward risk. This approach finds that much of expected utility theory is analytically robust to departures from linearity in the probabilities.[2]

CHOICE UNDER SUBJECTIVE UNCERTAINTY

States, Events, Outcomes, and Acts

From a mathematical perspective, the representation of uncertainty by means of additive, numerical probabilities allows us to apply the tremendous body of analytical results of modern probability theory (e.g., Feller 1968, 1971; Billingsley 1986). But from a modeling perspective, the assumption that uncertainty comes prepackaged with well-defined, measurable "objective" probabilities is unrealistic. Outside of the gambling hall, most economic decisions and transactions involving uncertainty—investment decisions, search decisions, insurance contracts, financial instruments—are defined in terms of uncertain events rather than numerical probabilities.

This approach to representing uncertainty and uncertain prospects—formalized by Savage (1954) and now known as the subjective approach—involves the following basic constructs:

$\chi = \{..., x, ...\}$	an arbitrary space of outcomes or consequences.

$S = \{..., s, ...\}$	a space of mutually exclusive and exhaustive states of nature, representing all possible alternative unfoldings of the world.
$E = \{..., E, ...\}$	an algebra of events, each a subset of S.
$f(\cdot) = [x_1 \text{ on } E_1; ...; x_m \text{ on } E_m]$	subjective act yielding outcome x_i in event E_i, for some partition $\{E_1,...,E_m\}$ of S (or equivalently, yielding outcome $f(s)$ in state s).
$A = \{..., f(\cdot), ...\}$	the set of all such subjective acts.
$W(\cdot)$ and $\succ$	the individual's preference function and corresponding preference relation over A.

A wonderful example of the use of this framework to represent an uncertain decision was provided by Savage (1954, pp. 13–15): Say you are making omelets and have already broken five of your six eggs into a mixing bowl. The decision you must make is: Do you break the sixth egg? The uncertainty arises from the fact that this sixth egg has been around for some time and might be rotten. You can either break this egg into the bowl with the other eggs, break it into a separate saucer to inspect it, or throw it away unbroken. Savage represents this problem in terms of states, acts, and outcomes by means of the following table:

	State	
Act	Egg is good	Egg is rotten
Break into bowl	Six-egg omelet	No omelet, and five good eggs destroyed
Break into saucer	Six-egg omelet, and a saucer to wash	Five-egg omelet, and a saucer to wash
Throw away	Five-egg omelet, and one good egg destroyed	Five-egg omelet

The Hypothesis of Probabilistic Sophistication

Although the subjective approach drops the assumption that uncertainty is defined in terms of numerical probabilities, it still allows for

individuals to possess probabilistic beliefs, with the feature that such beliefs may now differ across individuals. Formally, an individual is said to be probabilistically sophisticated, with a subjective (or personal) probability measure $\mu(\cdot)$ over the events E, if their preference function $W(\cdot)$ over subjective acts takes the form

$$W_{PS}(f(\cdot)) = W_{PS}(x_1 \text{ on } E_1; \ldots; x_m \text{ on } E_m) = V(x_1, \mu(E_1); \ldots; x_m, \mu(E_m))$$

for some (not necessarily expected utility) preference function $V(\mathbf{P}) = V(x_1,p_1;\ldots;x_m,p_m)$ over lotteries. That is to say, an individual is *probabilistically sophisticated* if their uncertain beliefs can be completely summarized by a subjective probability $\mu(E)$ attached to each event E, and the individual evaluates each subjective act $f(\cdot) = [x_1 \text{ on } E_1;\ldots;x_m \text{ on } E_m]$ solely on the basis of its implied probability distribution $(x_1,\mu(E_1);\ldots;x_m,\mu(E_m))$ over outcomes. This representation of $W_{PS}(\cdot)$ as the composition of a preference function $V(\cdot)$ over lotteries and a subjective probability measure $\mu(\cdot)$ over events is now referred to as the classical separation of risk preferences from beliefs.

Violations of the Hypothesis of Probabilistic Sophistication

Savage's (1954) joint axiomatization of expected utility risk preferences and probabilistic beliefs, employing an expected utility function for the risk preference function, has been justly termed "the crowning glory of choice theory" (Kreps 1988, p.120). However, the violations of expected utility first observed by Allais were soon matched by violations of probabilistic sophistication, even in situations involving the simplest forms of subjective uncertainty. The most famous of these examples, known as the Ellsberg Paradox (Ellsberg 1961, 2001), involves drawing a ball from an urn containing 30 red balls and 60 black or yellow balls in an unknown proportion. The following table illustrates four subjective acts defined over the color of the drawn ball, when the entries in the table are payoffs or outcomes:

	30 balls	60 balls	
	red	black	yellow
$f_1(\cdot)$	\$100	\$0	\$0
$f_2(\cdot)$	\$0	\$100	\$0
$f_3(\cdot)$	\$100	\$0	\$100
$f_4(\cdot)$	\$0	\$100	\$100

When faced with these choices, most subjects prefer act $f_1(\cdot)$ over $f_2(\cdot)$, on the grounds that the probability of winning \$100 in $f_1(\cdot)$ is guaranteed to be 1/3, whereas in $f_2(\cdot)$ it could range anywhere from 0 to 2/3. Similarly, most subjects prefer $f_4(\cdot)$ over $f_3(\cdot)$, on the grounds that the probability of winning \$100 in $f_4(\cdot)$ is guaranteed to be 2/3, whereas in $f_3(\cdot)$ it could range anywhere from 1/3 to 1. Although this reasoning may well be sound, it is inconsistent with the hypothesis of probabilistic beliefs. That is, there is *no* triple of subjective probabilities $\{\mu(\text{red}), \mu(\text{black}), \mu(\text{yellow})\}$ that can simultaneously generate a preference for $f_1(\cdot)$ over $f_2(\cdot)$ and for $f_4(\cdot)$ over $f_3(\cdot)$, since a probabilistically sophisticated individual would only exhibit the former ranking when $\mu(\text{red}) > \mu(\text{black})$, and only exhibit the latter ranking when $\mu(\text{red}) < \mu(\text{black})$.

Ellsberg also presented what many feel to be an even more fatal example, involving two urns:

	50 balls	50 balls
	red	black
$g_1(\cdot)$	\$100	\$0
$g_2(\cdot)$	\$0	\$100

	100 balls	
	red	black
$g_3(\cdot)$	\$100	\$0
$g_4(\cdot)$	\$0	\$100

In this example, most subjects are indifferent between $g_1(\cdot)$ and $g_2(\cdot)$, are indifferent between $g_3(\cdot)$ and $g_4(\cdot)$, but strictly prefer either of $g_1(\cdot)$ or $g_2(\cdot)$ to either of $g_3(\cdot)$ or $g_4(\cdot)$. It is straightforward to verify that there exist no pair of subjective probabilities $\{\mu(\text{red}), \mu(\text{black})\}$ for the right-hand urn—50:50 or otherwise—that can generate this set of preference rankings. Such examples illustrate the fact that in situations (even simple situations) where some events come with probabilistic information

and some events (termed *ambiguous events*) do not, subjective probabilities do not always suffice to fully encode all aspects of an individual's uncertain beliefs. Since most real-world events do not come with such probabilistic information, Ellsberg's Paradoxes and related phenomenon deal a serious blow to the hypothesis of probabilistic sophistication.

Non-probabilistically Sophisticated Models of Risk Preferences and Beliefs

Just as the Allais Paradox and similar evidence led to the development of non-expected utility models of risk preferences, Ellsberg's Paradoxes and similar phenomena have inspired the development of non-probabilistic models of preferences over subjectively uncertain acts. Such work has also progressed along two lines. One line replaces the subjective expected utility function with more general functional forms.[3]

The second line of research on non-probabilistic models treats $W(x_1 \text{ on } E_1; \ldots; x_m \text{ on } E_m)$ as a general smooth function of the *events* $E_1, \ldots, E_m$, and show how properties of $W(\cdot)$'s *event-derivatives* relate to features of both beliefs and attitudes toward risk, again taking expected utility as its base case. Appendix 2A presents mathematical features of this line of research.

ARE STATES OF NATURE PRESCRIPTIVE OR DESCRIPTIVE?

The second section in this chapter argued that the scientific suitability of a particular theoretical construct—in that case it was a particular classification scheme for food—could depend on whether the construct was meant to be applied to the independent variables of a theory or its dependent variables. This section addresses a similar issue, namely that certain criteria for suitable specification of the states of nature can depend upon whether the states are to be used for positive (that is, descriptive) versus normative (that is, prescriptive) purposes.

Since its inception, expected utility theory has always straddled the boundary between being a descriptive and a prescriptive model of decision making under uncertainty. Even its original presentation by Bernoulli (1738) as a "solution" to the St. Petersburg Paradox can be alternatively interpreted as either a description of why people *don't* assign an infinite certainty equivalent to the Petersburg Game, or a prescription for why an individual *shouldn't* assign an infinite certainty equivalent to the game. Two centuries later, proponents of objective expected utility theory defended it against the Allais Paradox by shifting their emphasis from the alleged descriptive power of the theory to its alleged normative power.

The same points can be made about the particular component of subjective expected utility theory that forms the central topic of this chapter—namely the notion of states of nature. It is one thing to assert that the states of nature approach offers a useful normative framework for decision making. It is quite a different thing to assert that, for the most part, this is how individuals actually do go about making decisions in the absence of probabilistic information. We shall consider each of these two domains in turn—in each case, with the goal of identifying the proper scientific criteria for states.

Criteria for Normative Applications

Savage's omelet example effectively shows how representing nature's underlying uncertainty by a set of "states," then representing one's alterative courses of action as "acts" that map these states into their respective consequences, can serve to organize a decision problem and make it easier to see exactly how one's beliefs (the state likelihoods) and risk preferences should enter into the problem. For proper normative application, this first step—namely, the specification of the states—must satisfy three properties:

1) The alternative states must be mutually exclusive—that is, no two distinct states can simultaneously occur. Thus, it would not have been correct to list "egg is rotten" and "five-egg omelet" as two distinct states, since is it possible that these could simultaneously occur.

2) The family of states must be exhaustive—that is, whatever happens, at least one of the states can be said to have occurred. Although it is at the same logical level as the previous criterion (mutual exclusivity), the exhaustiveness criterion is much more difficult—and some would argue, actually impossible—to guarantee in practice. For example, if you cracked the sixth egg into the bowl and found that it was actually hollow, then neither of the two states in the Savage table could be said to have occurred, since neither of the first-row consequences would be realized (you would not have a six-egg omelet, nor would you have destroyed the other five eggs). When the decision maker has reason to "expect the unexpected," the exhaustivity requirement cannot necessarily be achieved, and the best one can do is specify a final, catch-all state, with a label like "none of the above," and a very ill-defined consequence.

3) The states must represent nature's exogenous uncertainty, so their likelihoods cannot be affected by the individual's choice of act. This issue can be illustrated by a simple example involving the decision whether or not to install a lightning rod on one's house. Naturally, the relevant occurrences are the two mutually exclusive results {"house burns down," "house doesn't burn down"}. But since installing a lightning rod will clearly alter the respective likelihoods of these occurrences, can we really specify states of nature that are independent of the decision maker's action? The answer is illustrated in the following table, which makes it clear that "house burns down" and "house doesn't burn down" are not the states at all, but rather, part of the consequences, and clarifies that the effect of installing a lightning rod—as with any subjective act—is the outcome of an interaction between the act and an exogenous state of nature.

	State		
Act	Big lightning strike	Small lightning strike	No lightning strike
Lightning rod	House burns down, paid for rod	House doesn't burn, paid for rod	House doesn't burn, paid for rod
No lightning rod	House burns down, didn't pay for rod	House burns down, didn't pay for rod	House doesn't burn, didn't pay for rod

Do Decision Makers See the State Space or Do They Construct the State Space?

Is an individual who uses states of nature in normative decision making working with exogenous objects that they observe, or with endogenous objects that they construct? In one sense, this question is either subsidiary to, or equivalent to, the question of whether they are selecting from a menu of alternatives (subjective acts) that they observe as being available to them, or from a menu of alternatives that they have thought up or devised. Viewed in this larger sense, the question of whether the alternatives are observed or constructed is seen to have nothing to do with whether the choice happens to involve uncertainty at all, and indeed, the question may be equivalent to the classic question of whether Alexander Graham Bell discovered the idea for a telephone or invented this idea. In any case, I cannot derive any implications of this issue that pertain to the use of states of nature for normative purposes.

ISSUES OF OBSERVABILITY, CLASSIFICATION, AND MEASUREMENT

Independent Observability and the Exogeneity of States

Although the question of whether states are "exogenous and observed" versus "endogenous and constructed" does not seem to matter in a context of normative decision making, it matters a great deal for their relevance in descriptive science, for the types of reasons discussed in the "Scientific Considerations in the Theory of Choice" section. There we argued that economics had made greater scientific achievements than, say, anthropology because variables like prices and income were easier to measure than variables like trust or group identification. But if for some reason it should turn out that the full price of apples only exists in the eye the consumer and is not independently observable, then this advantage is lost. This might be the case if the acquisition of a commodity involves a time cost, set-up cost, or transaction cost that is observable to the consumer, but not to the outside

observer. Note that an inability to observe the true price—an independent variable—poses scientific problems even if we can still observe the exact amount purchased—the dependent variable—since it impedes our ability to observe the relationship between the two (the true demand function).

In the context of choice under uncertainty, such a problem would arise whenever the state space used by the decision maker did not correspond to the state space hypothesized by the scientific observer. In some sense, this is less likely to happen if the states are exogenous objects that are observed than if they are endogenous objects that must be constructed. But even in the former case, there is the possibility that the decision maker observes either a finer or a coarser set of states than does the scientist. Ultimately, the question reduces the scientist's ability to view the set of actions available to the decision maker—the left-hand columns in the above decision tables—and correctly predict the decision maker's specification (be it an observation or construction) of both the upper row and cell entries. Where this can and cannot be done is an empirical question.

Ex Ante Observability versus *Ex Post* Observability of States

Distinct from the question of whether the scientist can observe the set of states used by the decision maker is the question of whether the scientist can observe the realized state, or exactly when the scientist can observe the realized state. For example, in the case of choice under certainty—say, the demand for apples—it is clearly more important to be able to observe the price of apples that the consumer actually faces upon arriving at the supermarket, than to know the consumer's prior expectations of what this price might to be. But interestingly enough, for choice under uncertainty, the ability to observe the state space before the fact is of much greater importance than the ability to observe the realized state. The reason is that choice under uncertainty is by definition *ex ante*, and only depends upon *ex ante* features of the decision problem, namely the state space and the set of available subjective acts over this space. A scientist who correctly gleans the decision maker's formulation of these concepts, who knows his beliefs over the likelihoods of the states, and who knows his attitudes toward risk, will be able to correctly predict his decision—a decision that by definition

must be made before, and hence cannot be influenced by, the actual realization of the state. In both the omelet and the lightning rod examples, *ex post* knowledge of the realized state is of no further predictive use for the scientist, except for possible future decisions, via its effect on the specification of a state space for some subsequent decision, and/or likelihood beliefs over this space.

Issues of Classification and Measurement

Are qualitative state spaces likely to be more or less subject to the above types of observability issues than quantitative state spaces? As the example of the unobserved apple price illustrates, even real-valued independent variables—real-valued commodity prices or real-valued states of nature—are subject to these issues in principle. On the other hand, decisions where the state space is more naturally quantitative are probably less subject to these specification difficulties than decisions where the state space is more naturally qualitative. For example, compare the uncertainty related to investing in a domestic farming company compared to the uncertainty related to investing in a similar company located in a politically unstable foreign country. In the former case, the state space probably only has few dimensions, all of which are quantitative: the average temperature over the growing season, the average rainfall over the season, and average output prices at harvest. In the latter case, the most significant sources of uncertainty may be subjective—the particular political party that comes to power and its subsequent choice of expropriation policy. There is every reason to think that the scientist will do a much better job of modeling the decision maker's problem formulation in the first case than in the second case. Indeed, in the following section we shall see that measurable, as opposed to qualitative, state spaces can actually serve to bring some mathematical structure of objective uncertainty into a purely subjective setting.

ISSUES OF STRUCTURE: ALMOST-OBJECTIVE UNCERTAINTY

As noted, an important feature of objective uncertainty is that it allows us to apply the analytical tools of probability theory, such as combinatorics, the Central Limit Theorem, the Law of Large Numbers, and Chebyshev's Inequality. Furthermore, since objective probabilities are part of the objects of choice themselves, these types of results can be invoked independently of, and prior to, any knowledge of the individual's attitude toward risk. Thus, for example, the conditions under which the sum of two independent objective lotteries will have the same distribution as (and thus presumably be indifferent to) some third lottery will be the same for all individuals. Such results have the same character as arbitrage results in portfolio theory, which hold independently of risk preferences and hence yield extremely powerful results.

But in some sense, this strength of the objective framework is also its greatest weakness: it imposes too much uniformity of beliefs across individuals, and in many cases, too much structure on each individual's own beliefs. In contrast with preferences over objective lotteries, preferences over real-world subjective prospects are subject to the following three phenomena:

1) Individuals may have different subjective likelihoods for the same event (diverse beliefs).

2) Individuals' beliefs may not be representable by probabilities at all, with some (or all) events being considered ambiguous (absence of probabilistic sophistication).

3) Individuals' outcome preferences may depend upon the source of uncertainty itself (outcome preferences may be state-dependent).

Nevertheless, it turns out that if the state space has a Euclidean structure and preferences are smooth in the events in the sense described in Appendix 2A, then features of "objective" uncertainty will emerge even in a purely subjective setting. In Appendix 2B, we sketch out the intuition of these results—readers wishing a formal development are referred to Machina (2001).

We can summarize the scientific implications of almost-objective uncertainty as follows: In the more traditional approach to uncertainty (e.g., Anscombe and Aumann 1963), the world presented two qualitatively different types of uncertainty: uncertain processes (such as perfectly balanced roulette wheels) that only generated idealized, purely *objective* events for which all agents held common beliefs and probabilistically sophisticated betting preferences; and uncertain processes (such as tomorrow's temperature or rainfall level) that only generated purely *subjective* events, where individuals typically differed in their likelihood beliefs, or had no likelihood beliefs, and could be state-dependent. On the other hand, according to the concepts presented in Appendix 2B, once a purely subjective state space is given a Euclidean structure and preferences are assumed to be smooth in the events, there exist events that arbitrarily closely approximate all the properties of classical "objective events" for all decision makers, in spite of any interpersonal differences in beliefs, lack of probabilistic sophistication, or state-dependence. Furthermore, once standard "objective randomizing devices" are reexamined, they are seen to depend precisely on these type of "almost-objective events."

Given the traditional (e.g., Savage 1954) approach of positing an almost completely unrestricted subjective state space and no event-smoothness, the "Euclidean state space + event-smoothness" approach advocated in the previous paragraph might seem overly strong. But in fact, it is well within standard economic practice. Standard consumer theory under certainty requires no structure at all on a family of objects of choice in order to axiomatize an ordinal utility function over these objects. Debreu's (1954) original topological assumptions were later shown to be unnecessary by Kreps (1988, pp. 25–26). But the workhorse concepts of competitive prices, marginal rates of substitution, demand functions, and the Slutsky equation do not emerge until we assume a Euclidean structure for these objects (vector "commodity bundles" and a Euclidean "commodity space") and/or smooth preferences over this space. Under uncertainty, restricting ourselves to a Euclidean state space amounts to nothing more than restricting ourselves to subjective uncertainty that appears in the form of random variables (such as temperature or random prices). And for the types of reasons discussed earlier in this chapter, real- or vector-valued states of nature are much more likely to be commonly observable and com-

monly measurable than are states of nature that are elements of some more abstract space.

Just as science in general has progressed most rapidly when it has been able to quantify and measure the natural world, research in uncertain preferences and beliefs will further progress most rapidly to the extent we are able to quantify and measure the objects we call "states of nature" or "states of the world."

Notes

I would like to thank Ted Groves, Donald Meyer, and Joel Sobel for helpful comments. This material is based upon work supported by the National Science Foundation under Grant No. 9870894. All opinions and errors are my own.

1. If the outcomes x describe monetary payoffs, then the standard monotonicity assumptions are that $V(x_1,p_1;...;x_m,p_m)$ is increasing in each of the variables $x_1,...,x_m$, and also increasing whenever p_i is increased at the expense of p_j for some pair of outcomes $x_i > x_j$ (that is, whenever probability mass is shifted from a lower to a higher outcome).
2. For example, an expected utility preference function $U(x_1)\cdot p_1 + \ldots + U(x_m)\cdot p_m$ will be risk averse if and only if its *coefficient* with respect to prob(x) (that is, the value $U(x)$) is a concave function of wealth. Correspondingly, a non-expected utility preference function $V(\mathbf{P}) = V(x_1,p_1;...;x_m,p_m)$ will be risk averse if and only if its *partial derivative* with respect to prob(x) (that is, the value $\partial V(\mathbf{P})/\partial \text{prob}(x)$) is a concave function of wealth.
3. Such as the *Choquet expected utility* form

$$W_{Choquet}(x_1 \text{ on } E_1; \ldots; x_m \text{ on } E_m) \equiv \sum_{i=1}^{m} U(x_i)\cdot\left[C\left(E_1\cup\ldots\cup E_i\right) - C\left(E_1\cup\ldots\cup E_{i-1}\right)\right]$$

for some utility function $U(\cdot)$ and *capacity* (monotonic non-additive measure) $C(\cdot)$, where the outcomes are labeled so that $x_1 < \ldots < x_m$ (e.g., Gilboa 1987; Schmeidler 1989; Wakker 1989, 1990; Gilboa and Schmeidler 1994), or the *maxmin expected utility* form

$$W_{maxmin}(x_1 \text{ on } E_1; \ldots; x_m \text{ on } E_m) \equiv \min_{\tau\in T} \int_S U(f(s))\cdot d\mu_\tau(s) \equiv \min_{\tau\in T} \sum_{i=1}^{m} U(x_i)\cdot\mu_\tau(E_i)$$

for some utility function $U(\cdot)$ and *family* $\left\{\mu_\tau(\cdot)\middle|\tau\in T\right\}$ of probability measures on S (e.g., Gärdenfors and Sahlin 1982, 1983; Cohen and Jaffray 1985; Gilboa and Schmeidler 1989).

Appendix 2A

Properties of the Smooth Function Approach to Non-probabilistically Sophisticated Models

This approach starts by equivalently reexpressing each act $f(\cdot) = [x_1 \text{ on } E_1; \ldots; x_n \text{ on } E_n]$ in the form $f(\cdot) = [\ldots; x \text{ on } f^{-1}(x); \ldots] = [\ldots; x \text{ on } E_x; \ldots]$, as x ranges over all possible outcomes $x \in \chi$. The preference functions $W_{SEU}(\cdot)$ *and* $W_{SDEU}(\cdot)$ can then be expressed in the event-additive forms

$$W_{SEU}(\ldots; x \text{ on } E_x; \ldots) \equiv \sum_{x \in \chi} \Phi_x(E_x) \quad \text{where} \quad \Phi_x(E) \stackrel{\text{def}}{=} U(x) \cdot \mu(E)$$

$$W_{SDEU}(\ldots; x \text{ on } E_x; \ldots) \equiv \sum_{x \in \chi} \Phi_x(E_x) \quad \text{where} \quad \Phi_x(E) \stackrel{\text{def}}{=} \int_E U(x|s) \cdot d\mu(s)$$

where the event E_x attached to each outcome x is evaluated by an additive evaluation measure $\Phi_x(\cdot)$, which is the subjective analogue of objective expected utility's probability coefficient $U(x)$.

Just as linearity in a set of variables implies linearity in their changes, event-additive functions like $W_{SEU}(\ldots; x \text{ on } E_x; \ldots) = \sum_{x \in \chi} \Phi_x(E_x)$ and $W_{SDEU}(\ldots; x \text{ on } E_x; \ldots) = \sum_{x \in \chi} \Phi_x(E_x)$ will also be additive in event changes ("growth and shrinkage sets"). That is, their ranking of two acts $f(\cdot) = [\ldots; x \text{ on } E_x; \ldots]$ versus $f^*(\cdot) = [\ldots; x \text{ on } E_x^*; \ldots]$ is determined by the additive formulas

$$W_{SEU}(f^*(\cdot)) - W_{SEU}(f(\cdot)) \equiv \sum_{x \in \chi} \Phi_x(E_x^*) - \sum_{x \in \chi} \Phi_x(E_x) \equiv \sum_{x \in \chi} \Phi_x(E_x^* - E_x) - \sum_{x \in \chi} \Phi_x(E_x - E_x^*)$$

$$W_{SDEU}(f^*(\cdot)) - W_{SDEU}(f(\cdot)) \equiv \sum_{x \in \chi} \Phi_x(E_x^*) - \sum_{x \in \chi} \Phi_x(E_x) \equiv \sum_{x \in \chi} \Phi_x(E_x^* - E_x) - \sum_{x \in \chi} \Phi_x(E_x - E_x^*)$$

where for each x, its *growth set* in going from $f(\cdot)$ to $f^*(\cdot)$, namely the set $E_x^* - E_x$, is evaluated *positively* by x's evaluation measure $\Phi_x(\cdot)$, and its *shrinkage set*, namely the set $E_x - E_x^*$, is evaluated negatively by $\Phi_x(\cdot)$.

Just as differentiability in objective probabilities can be defined as *local linearity in probability changes*, smoothness in subjective events can be defined as *local additivity in event changes*. That is, one can define a general preference function $W(\ldots; x \text{ on } E_x; \ldots)$ to be event-differentiable if at each act $f(\cdot)$ it possesses a family of local evaluation measures $\{\Phi_x(\cdot; f) \mid x \in \chi\}$ such that

$W(\cdot)$ evaluates small event changes from $f(\cdot)$ in the following locally additive manner:

$$W(f^*(\cdot)) - W(f(\cdot)) \equiv \sum_{x \in \chi} \Phi_x(E_x^* - E_x; f) - \sum_{x \in \chi} \Phi_x(E_x - E_x^*; f) + o\big(\delta(f^*(\cdot), f(\cdot))\big)$$

where the distance function $\delta(f^*(\cdot), f(\cdot))$ between acts has the property that it shrinks to zero as the change sets $E_x^* - E_x$ and $E_x - E_x^*$ all shrink to zero, and (as with any definition of differentiability) $o(\cdot)$ denotes a function that is of higher order than its argument. In Machina (2002), I have shown how this calculus of events can be applied to establish the robustness of most of classical state-independent and state-dependent subjective expected utility theory and subjective probability theory to general event-smooth (but not necessarily either expected utility or probabilistically sophisticated) preference functions $W(\cdot)$ over subjective acts.

Appendix 2B
Almost-Objective Uncertainty

PROPERTIES OF PURELY OBJECTIVE EVENTS

We begin by contrasting the three properties of subjective events—diverse interpersonal beliefs, possible absence of probabilistic sophistication, and possible state-dependence—with the following four characteristic properties of idealized, exogenous "purely objective" events:

- *Unanimous, outcome-invariant revealed likelihoods:* In contrast with the above-listed properties of subjective events, all individuals exhibit identical, outcome-invariant revealed likelihoods over purely objective events—corresponding to their objective probabilities.
- *Independence from subjective realizations:* In the presence of joint objective × subjective uncertainty, purely objective events are independent of the realization of subjective events. Thus, the events generated by an exogenous objective coin, die, or roulette wheel are invariant to whether any given subjective event E does or does not occur.
- *Probabilistic sophistication over objective lotteries:* It is almost a truism that all individuals evaluate objective lotteries $\mathbf{P} = (x_1, p_1; ...; x_m, p_m)$ solely according to their outcomes and corresponding objective likelihoods, via some preference function $V(x_1, p_1; ...; x_m, p_m)$.
- *Reduction of objective × subjective uncertainty:* Standard reduction of compound uncertainty assumptions imply that individuals evaluate any objective mixture of subjective acts $\alpha \cdot f(\cdot) + (1-\alpha) \cdot f^*(\cdot) = \alpha \cdot [x_1 \text{ on } E_1; ...; x_m \text{ on } E_m] + (1-\alpha) \cdot [x_1^* \text{ on } E_1^*; ...; x_m^* \text{ on } E_m^*]$ solely according to its induced map $[\ldots; (x_i, \alpha; x_j^*, 1-\alpha) \text{ on } E_i \cap E_j^*; \ldots]$ from events to lotteries.

The above features of objective uncertainty apply to all individuals, whether or not they are expected utility, state-independent or probabilistically sophisticated. The following properties additionally hold for probabilistically sophisticated individuals and expected utility individuals:

- *Under probabilistic sophistication, independence of objective and subjective likelihoods:* If the individual is probabilistically sophisticated with probability measure $\mu(\cdot)$ over subjective events, these likelihoods are independent of exogenous objective events, and vice versa.
- *Under expected utility, linearity in objective likelihoods:* Expected utility is linear in objective probabilities ($V_{EU}(x_1,\ p_1;\ ...;\ x_m,\ p_m) \equiv \sum_{i=1}^{m} U(x_i) \cdot p_i$) and in objective mixtures of lotteries ($V_{EU}(\alpha \cdot \mathbf{P} + (1-\alpha) \cdot \mathbf{P}^*) \equiv \alpha \cdot V_{EU}(\mathbf{P}) + (1-\alpha) \cdot V_{EU}(\mathbf{P}^*)$). Under objective × subjective uncertainty, expected utility is linear in objective mixtures of subjective acts: $W_{SEU}(\alpha \cdot f(\cdot) + (1-\alpha) \cdot f^*(\cdot)) \equiv \alpha \cdot W_{SEU}(f(\cdot)) + (1-\alpha) \cdot W_{SEU}(f^*(\cdot))$), and similarly for $W_{SDEU}(\cdot)$.

ALMOST-EQUALLY-LIKELY EVENTS AND ALMOST-FAIR BETS

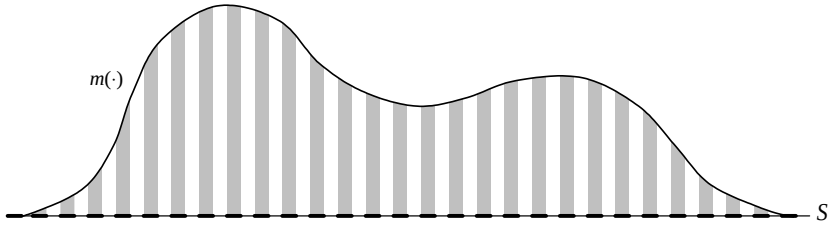
As the above bullet lists indicate, the properties of purely subjective and purely objective events lie in stark contrast. Nevertheless, in a Euclidean state space $S = [\underline{s}, \overline{s}] \subseteq R^1$, some subjective events are closer to being objective than others. We illustrate this by an example which approximates what is surely the "canonical" objective event: namely, the flip of an exogenous, fair coin.

Denoting the events implied by this coin by with standard notation $\{H,T\}$, their characteristic property is that, for any pair of prizes $x^* > x$, all individuals will be indifferent between the bets [x^* on H; x on T] and [x on H; x^* on T]. In contrast, for any subjective event E, ranking of the bets [x^* on E; x on $\sim E$] versus [x on E; x^* on $\sim E$] can differ across individuals (due to diverse beliefs), or can reverse if the prizes $x^* > x$ are replaced by $y^* > y$ (due to state-dependence).

However, consider the event E_n obtained by dividing the state space $S = [\underline{s}, \overline{s}]$ into n equal-length intervals, and defining E_n as the union of the *odd*-numbered intervals (the complementary event $\sim E_n$ thus being the union of the even-numbered intervals). As the following diagram indicates, *regardless* of an individual's particular subjective probability measure $\mu(\cdot)$ over the state space S (indicated by its density function $m(\cdot)$ in the Figure 2.B1), as n approaches infinity, the individual will assign equal subjective probabilities of 1/2 to each of the events E_n and $\sim E_n$, and hence be virtually indifferent between the bets [x^* on E_n; x on $\sim E_n$] versus [x on E_n; x^* on $\sim E_n$]. State-dependent individuals will be similarly indifferent, and as shown in Machina (2001), as $n \rightarrow \infty$, all event-smooth individuals—whether or not they are expected utility, state independent, or even probabilistically sophisticated—will "reveal" E_n and $\sim E_n$ to be equally likely, via their indifference between any two bets of the form [x^* on E_n; x on $\sim E_n$] versus [x on E_n; x^* on $\sim E_n$]. In other words, as $n \rightarrow \infty$,

the purely subjective events E_n and $\sim E_n$—both subsets of the purely subjective state space S—take on the properties of exogenous objective 50:50 events.

Figure 2B.1 Example of a Subjective Probability Density Function



ALMOST-OBJECTIVE EVENTS, ACTS AND MIXTURES

It is clear that by dividing the state space $S = [\underline{s}, \bar{s}]$ into a large number of equal-length intervals and taking the union of every third interval, we could create an subjective event that approximates the properties of an exogenous event of probability 1/3, etc. We can extend and formalize this idea as follows: Given any sufficiently regular (e.g., finite interval union) subset $\wp$ of the unit interval [0,1] and any large n, partition S into n equal-length intervals [0, 1/n), [1/n, 2/n) … [(n–2)/n, (n–1)/n), [(n–1)/n,1], and define the almost-objective event $\wp \underset{n}{\times} S \subseteq S$ by

$$\wp \underset{n}{\times} S \;=\; \bigcup_{i=0}^{n-1} \left\{ \underline{s} + (i+\omega) \cdot \frac{\bar{s} - \underline{s}}{n} \;\middle|\; \omega \in \wp \right\}$$

that is, as the union of $\wp$'s linear images into each of S's n equal-length intervals. Thus, the event E_n illustrated in the previous figure is simply almost-objective event $[0, 1/2] \underset{n}{\times} S$.

By taking a partition $\{\wp_1, ..., \wp_m\}$ of the unit interval we can create almost-objective partitions $\{\wp_1 \underset{n}{\times} S, ..., \wp_m \underset{n}{\times} S\}$ of the state space S, and in turn define almost-objective acts $[x_1$ on $\wp_1 \underset{n}{\times} S; ... ; x_m$ on $\wp_m \underset{n}{\times} S]$. The almost-fair bets of the previous subsection are seen to be the almost-objective acts $[x^*$ on $[0, 1/2] \underset{n}{\times} S; x$ on $(1/2, 1] \underset{n}{\times} S]$ and $[x$ on $[0, 1/2] \underset{n}{\times} S; x^*$ on $(1/2, 1] \underset{n}{\times} S]$. Finally, given two subjective acts $f(\cdot) = [x_1$ on $E_1; ... ; x_m$ on $E_m]$ and $f^*(\cdot) =$

$[x_1^*$ on E_1^*; ... ; $x_m^*{}_*$ on $E_m^*{}_*]$ and any $\wp, \sim\wp \subseteq [0,1]$, we can define the almost-objective mixture $[f(\cdot)$ on $\wp \underset{n}{\times} S$; $f^*(\cdot)$ on $\sim\wp \underset{n}{\times} S]$ of $f(\cdot)$ and $f^*(\cdot)$.

BELIEFS AND BETTING PREFERENCES OVER ALMOST-OBJECTIVE EVENTS

As with the almost-equally likely events defined above, as $n \to \infty$ all event-smooth individuals will exhibit identical revealed likelihood beliefs over any almost-objective event $\wp \underset{n}{\times} S$ essentially treating it as an exogenous objective event, with a probability given by the total length $\lambda(\wp)$ of the subset $\wp \subseteq [0, 1]$. That is to say, given any event-smooth preference function $W(\cdot)$ over subjective acts—whether or not it is expected utility/non-expected utility, state-independent/state-dependent, or probabilistically sophisticated/non-probabilistically sophisticated—outcomes $x^* > x$, disjoint subsets $\wp, \hat{\wp} \subseteq [0,1]$ with $\lambda(\wp) > \lambda(\hat{\wp})$, and subjective act $f(\cdot)$, $W(\cdot)$ will exhibit

$$\lim_{n\to\infty} W(x^* \text{ on } \wp \underset{n}{\times} S; x \text{ on } \hat{\wp} \underset{n}{\times} S; f(s) \text{ elsewhere}) >$$

$$\lim_{n\to\infty} W(x \text{ on } \wp \underset{n}{\times} S; x^* \text{ on } \hat{\wp} \underset{n}{\times} S; f(s) \text{ elsewhere});$$

that is, holding the payoffs elsewhere constant, all event-smooth individuals are unanimous in their preference for staking the greater of two prizes on the event $\wp \underset{n}{\times} S$ and the lesser on $\hat{\wp} \underset{n}{\times} S$, rather than the other way around. Thus, while we have seen that typical subjective events need not have probabilities at all, much less unanimously agreed-upon probabilities, as $n \to \infty$ there *will* be such unanimous agreement on the comparative likelihoods of $\wp \underset{n}{\times}$ versus $\hat{\wp} \underset{n}{\times} S$.

The idea that some subjective events come close to exhibiting objective properties is not new, and precursors of the almost-equal-likelihood example date back at least to Poincaré (1912). Nor are almost-objective events merely a technical curiosum—in fact, most real-world "objective randomization devices" are actually examples of the use of almost-objective events to convert non-probabilistic subjective uncertainty to (almost-) objective uncertainty. To see this, consider the simple example of a game show spinner divided into a large number of alternating red and black sectors of equal angular size. Is it correct to say that the spin of such a wheel is an "objective process"? If so, then it would follow that all individuals would have the same beliefs over all events

defined over this process. But how much agreement will there be on the likelihood of the event that "the wheel spins more than 20 revolutions before finally stopping"?

Viewed from this perspective, the behavior of the wheel—its exact number of revolutions and therefore the color of the sector that finishes opposite the pointer—is a subjective process, where the state of nature is the amount of force applied to the spin. Individuals will surely disagree on their subjective probabilities of an event like "the force will be enough to generate at least 20 revolutions," and some may not be able to attach any subjective probably at all to this event. But if we plot the state (the initial force of the spin) on the horizontal axis of the previous diagram, then an event such as "the force will lead the wheel to stop with the pointer opposite a black sector" is seen to be an almost-objective event of the type illustrated in the figure, which is why even individuals who disagree on the likelihood of "more than 20 spins" will nevertheless agree on the likelihood of "black." In other words, it is not the *process* of spinning the wheel that is either "subjective" or "objective," but rather the different *events* defined on this process that are either subjective or (almost-) objective. A little thought will reveal that virtually all standard physical randomization devices used to generate "objective" likelihoods share this property of being based on a subjectively uncertain (and hence non-probabilistic) state variable (or variables), but working with periodic, "almost-objective" events defined over the state variable.

The above argument shows that with a structured (essentially Euclidean) state space and the property of event-smooth preferences, there exists a substratum of events that arbitrarily closely approximate the first of the four above-listed properties of purely objective events, namely the property of unanimous, outcome-invariant revealed likelihoods. In Machina (2001) I have shown that such events, and the acts and mixtures based on them, also arbitrarily closely approximate the other three listed properties of idealized "purely objective" events. That is, as $n \to \infty$:

- Each individual (probabilistically sophisticated or otherwise) will view all almost-objective events as independent of each purely subjective event, in the sense that for all disjoint $\wp, \hat{\wp} \subseteq [0,1]$ and each $E \subseteq S$, they will have the same revealed likelihood rankings (i.e., betting preferences) over the joint events $(\wp \underset{n}{\times} S) \cap E$ versus $(\hat{\wp} \underset{n}{\times} S) \cap E$ as they do over the events $\wp \underset{n}{\times} S$ versus $\hat{\wp} \underset{n}{\times} S$ (in each case, corresponding to the relative values of $\lambda(\wp)$ versus $\lambda(\hat{\wp})$).

- Although individuals needn't be probabilistically sophisticated over subjective acts in general, they will be probabilistically sophisticated over almost-objective acts. That is, each $W(\cdot)$ will have a corresponding preference function $V_W(\cdot)$ over lotteries such that

$$\lim_{n\to\infty} W\left(x_1 \text{ on } \wp_1 \underset{n}{\times} S; \ldots; x_m \text{ on } \wp_m \underset{n}{\times} S\right) = \lim_{n\to\infty} V_W\left(x_1, \lambda(\wp_1); \ldots; x_m, \lambda(\wp_m)\right).$$

- Each individual (probabilistically sophisticated or otherwise) satisfies the reduction of compound uncertainty property for almost-objective mixtures of acts. Thus if $\left[f_1(\cdot) \text{ on } \wp \underset{n}{\times} S; \ldots; f_m(\cdot) \text{ on } \wp_m \underset{n}{\times} S \right]$ and $\left[\hat{f}_1(\cdot) \text{on } \hat{\wp}_1 \underset{n}{\times} S; \ldots; \hat{f}_{\hat{m}}(\cdot) \text{on } \hat{\wp}_{\hat{m}} \underset{n}{\times} S \right]$ induce almost-objectively equivalent subacts over each event in the common refinement of $f_1(\cdot), \ldots, f_m(\cdot), \hat{f}_1(\cdot), \ldots, \hat{f}_{\hat{m}}(\cdot)$, then

$$\lim_{n\to\infty} W\left(f_1(\cdot) \text{ on } \wp_1 \underset{n}{\times} S; \ldots; f_m(\cdot) \text{ on } \wp_m \underset{n}{\times} S \right) =$$

$$\lim_{n\to\infty} W\left(\hat{f}_1(\cdot) \text{ on } \hat{\wp}_1 \underset{n}{\times} S; \ldots; \hat{f}_{\hat{m}}(\cdot) \text{ on } \hat{\wp}_{\hat{m}} \underset{n}{\times} S \right).$$

The following properties of almost-objective uncertainty additionally hold for probabilistically sophisticated individuals and expected utility individuals:

- Each probabilistically sophisticated individual with subjective probability measure $\mu(\cdot)$ will view *all purely subjective events as independent of each almost-objective event*, in the sense that for all $E, \hat{E} \subseteq S$ and each $\wp \subseteq [0, 1]$, they will have the same revealed likelihood rankings (betting preferences) over the joint events $(\wp \underset{n}{\times} S) \cap E$ versus $(\wp \underset{n}{\times} S) \cap \hat{E}$ as they do over the events E versus $\hat{E}$ (in each case, corresponding to the relative values of $\mu(E)$ versus $\mu(\hat{E})$).
- Each expected utility maximizer will be *linear in almost-objective probabilities and almost-objective mixtures of subjective lotteries*, i.e.,

$$\lim_{n\to\infty} W_{SEU}\left(x_1 \text{ on } \wp_1 \underset{n}{\times} S; \ldots; x_m \text{ on } \wp_m \underset{n}{\times} S\right) = \sum_{i=1}^{m} \lambda(\wp_i) \cdot W_{SEU}\left(x_i \text{ on } S\right)$$

$$\lim_{n\to\infty} W_{SEU}\left(f_1(\cdot) \text{ on } \wp_1 \underset{n}{\times} S; \ldots; f_m(\cdot) \text{ on } \wp_m \underset{n}{\times} S\right) = \sum_{i=1}^{m} \lambda(\wp_i) \cdot W_{SEU}\left(f_i(\cdot)\right)$$

and similarly for $W_{SDEU}(\cdot)$.

References

Allais, M. 1953. "Le Comportement de l'Homme Rationnel devant le Risque, Critique des Postulats et Axiomes de l'Ecole Américaine." *Econometrica* 21(4): 503–546.

Anscombe F., and R. Aumann. 1963. "A Definition of Subjective Probability." *Annals of Mathematical Statistics* 34(1): 199–205.

Bernoulli, D. [1738] 1954. "Specimen Theoriae Novae de Mensura Sortis," *Commentarii Academiae Scientiarum Imperialis Petropolitanae* [Papers of the Imperial Academy of Sciences in Petersburg] V, 175-192. English translation: "Exposition of a New Theory on the Measurement of Risk." *Econometrica* 22(1): 23–36.

Billingsley, P. 1986. *Probability and Measure.* 2d ed. New York: John Wiley and Sons.

Chew, S. 1983. "A Generalization of the Quasilinear Mean With Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox." *Econometrica* 51(4): 1065–1092.

Cohen, M., and J. Jaffray. 1985. "Decision Making in the Case of Mixed Uncertainty: A Normative Model." *Journal of Mathematical Psychology* 29(4): 428–442.

Debreu, G. 1954. "Representation of a Preference Ordering by a Numerical Function." In *Decision Processes*, R. Thrall, C. Boombs, and R. Davis, eds. New York: John Wiley and Sons, pp. 159–165.

Edwards, W. 1955. "The Prediction of Decisions among Bets." *Journal of Experimental Psychology* 50(3): 201–214.

———. 1962. "Subjective Probabilities Inferred From Decisions." *Psychological Review* 69(2): 109–135.

Ellsberg, D. 1961. "Risk, Ambiguity, and the Savage Axioms." *Quarterly Journal of Economics* 75: 643–669.

———. 2001. *Risk, Ambiguity and Decision.* New York: Garland Publishing, Inc.

Feller, W. 1968. *An Introduction to Probability Theory and Its Applications, Volume I.* 3d Ed. New York: John Wiley and Sons.

———. 1971. *An Introduction to Probability Theory and Its Applications, Volume II.* 2d Ed. New York: John Wiley and Sons.

Fishburn, P. 1983. "Transitive Measurable Utility." *Journal of Economic Theory* 31(2): 293–317.

Gärdenfors, P., and N. Sahlin. 1982. "Unreliable Probabilities, Risk Taking, and Decision Making." *Synthese* 53(3): 361–386.

———. 1983. "Decision Making with Unreliable Probabilities." *British Journal of Mathematical and Statistical Psychology* 36(2): 240–251.

Gilboa, I. 1987. "Expected Utility with Purely Subjective Non-Additive Probabilities." *Journal of Mathematical Economics* 16(1): 65–88.

Gilboa, I. and D. Schmeidler. 1989. "Maximum Expected Utility With a Non-Unique Prior." *Journal of Mathematical Economics* 18(2): 141–153.

———. 1994. "Additive Representations of Non-Additive Measures and the Choquet Integral." *Annals of Operations Research* 52: 43–65.

Kahneman, D., and A. Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47(2): 263–291.

Kreps, D. 1988. *Notes on the Theory of Choice.* Boulder, Colorado: Westview Press.

———. 1990. *A Course in Microeconomic Theory.* Princeton: Princeton University Press.

Machina, M. 1982. "'Expected Utility' Analysis Without the Independence Axiom." *Econometrica* 50(2): 277–323.

———. 1983. "Generalized Expected Utility Analysis and the Nature of Observed Violations of the Independence Axiom." In *Foundations of Utility and Risk Theory with Applications*, B. Stigum and F. Wenstop, eds. Dordrecht: D. Reidel Publishing Co., pp. 263–293.

———. 1987. "Choice Under Uncertainty: Problems Solved and Unsolved." *Journal of Economic Perspectives* 1(1): 121–154.

———. 2001. "Almost-Objective Uncertainty." Working paper 2001-12. San Diego: University of California, Department of Economics.

———. 2002. "Robustifying the Classical Model of Risk Preferences and Beliefs." Working paper 2002-06. San Diego: University of California, Department of Economics.

MacCrimmon, K., and S. Larsson. 1979. "Utility Theory: Axioms Versus 'Paradoxes'." In *Expected Utility Hypotheses and the Allais Paradox*, Allais and Hagen, eds. Dordrecht: D. Reidel Publishing Co., pp. 333–409.

Mas-Colell, A., M. Whinston and J. Green. 1995. *Microeconomic Theory.* Oxford: Oxford University Press.

Poincaré, H. 1912. *Calcul des Probabilités.* 2d ed. Paris: Gauthiers-Villars.

Quiggin, J. 1982. "A Theory of Anticipated Utility." *Journal of Economic Behavior and Organization* 3(4): 323–343.

Savage, L. 1954. *The Foundations of Statistics.* New York: John Wiley and Sons.

Schmeidler, D. 1989. "Subjective Probability and Expected Utility without Additivity." *Econometrica* 57(3): 571–587.

Varian, H. 1992. *Microeconomic Analysis*. 3d ed. New York: W.W. Norton & Co.

Wakker, P. 1989. "Continuous Subjective Expected Utility With Non-Additive Probabilities." *Journal of Mathematical Economics* 18(1): 1–27.

———. 1990. "Under Stochastic Dominance Choquet-Expected Utility and Anticipated Utility are Identical." *Theory and Decision* 29(2): 119–132.

Yaari, M. 1987. "The Dual Theory Of Choice Under Risk." *Econometrica* 55(1): 95–115.

3
Gambling with the Future

Economic and Social Perspectives on Casinos in America

William R. Eadington
University of Nevada, Reno

Commercial gaming became a substantial industry in the United States over the second half of the 20th century, generating revenues in 2001 in excess of $64 billion, and having a legal presence in 48 states. Over half of gaming revenues come from commercial and Indian casinos located in more than 30 states.

From strict prohibitions in most states only a generation ago, laws governing casino-style gambling have been entertained or enacted by state governments interested in new sources of tax revenues, new catalysts for job creation and capital investment, new reasons for attracting tourist spending, and—occasionally—in response to citizens' desires to participate in casino-style gambling for the fun of it. The types of gambling authorized include Nevada-style casinos, slot machines at race tracks, and video poker, video lottery terminals, and other electronic gaming devices in bars and taverns. Besides state-authorized gaming, nearly 200 Indian tribes have opened tribal casinos and gaming centers throughout the country, under the general guidance of the Indian Gaming Regulatory Act of 1988.

This chapter examines the major political, social, and economic dynamics that have resulted in the rapid proliferation of permitted gambling—especially casinos and casino-style gambling—in the United States over the past quarter century. This process of legalization and deregulation has created gaming industries of increasing size, sophistication, and presence, which have become—or are quickly

becoming—part of the modern mainstream of commercial entertainment, leisure, and tourism industries in various parts of the country.

Economic benefits notwithstanding, permitted casino-style gaming remains a highly charged political issue. Casino gaming is still considered by some to be an essentially unhealthy activity that has not lost its previous status as a pernicious vice. States that have authorized casino gaming have often done so under conditions of limited competition or by using regional monopoly structures. This approach creates economic rents or monopoly profits that become the objective of ongoing rent-seeking behavior by various special interests in their efforts to capture the rents. Such designed market structures are typically a by-product of desires at the legislative level to control gaming's social impacts through regulatory constraints, geographic isolation, or planned undersupply. Nonetheless, pressures to expand the scope of permitted gaming are found in many jurisdictions, especially when needs for tax revenue generation or job creation are substantial.

It is difficult to make an unambiguous case either in favor of or in opposition to permitting casino-style gaming into any community that previously did not have such activities. Nonetheless, in the first years of the 21st century, it appears that gaming industries will continue to expand in new and diverse ways in many jurisdictions. At minimum, expanding permitted casino-style gaming is now actively on the agenda in many state legislatures, and it is likely to remain so for some time to come.

TRENDS IN GAMING IN THE UNITED STATES

Events of recent years are a continuation of processes toward legalization and a greater presence of permitted gaming that began with Nevada's casino legislation in 1931 and New Hampshire's lottery legalization in 1963. However, the main spread of legal commercial casinos occurred in the first half of the 1990s. Prior to 1988, casinos had been authorized only in Nevada and in Atlantic City, New Jersey. Atlantic City itself was a relatively new addition, with its casinos opening their doors for the first time in 1978. Between 1988 and 1996, a total of nine states[1] authorized new casino industries, some as riverboat

casinos, some as limited-wager casinos in former mining towns, and some as urban casinos. Indian tribal casinos were effectively legalized by a Supreme Court decision in 1987[2] and were provided a statutory framework with the passage of the Indian Gaming Regulatory Act in 1988.[3] Indian casinos spread to nearly 30 states by the early 21st century, with the most significant tribal casinos found in such states as Connecticut, Minnesota, Michigan, New York, and California.

The proliferation of permitted commercial casinos in the United States slowed down after 1993, coinciding with the improvements in performance of the national economy. However, another trend soon emerged: the authorization of gaming devices at race tracks in various states, purportedly to provide the racing industry with a "level playing field" against newly authorized forms of gaming, and a competitive edge over tracks in other states in attracting purses and high-quality race horses. The effect of this development was to create a number of "racinos," where the presence of slot machines would transform race tracks into *de facto* casinos, and typically lead to a high proportion of total revenues for such operations being generated by the gaming devices rather than wagering on racing. Such race track casinos have developed in Iowa, Delaware, West Virginia, New Mexico, Rhode Island, and Louisiana in the 1990s.

Changing economic circumstances, especially recession and substantial fiscal shortfalls at the state level, contribute to the casino debate. In 2003, the United States went through another round of discussion of whether to legalize and expand casinos and casino-style gaming. Economic circumstances in the early years of the first decade of the 21st century parallel the period from 1989 to 1993, when the national economy slowed and then moved into recession, and when many states found themselves financially strapped and desperate for job-creating strategies. With the economic slowdown and recession of 2000–2003, an increasing number of jurisdictions in the United States found themselves in financial difficulty. As such slowdowns occurred, commercial gaming was often one of the strategies put forth for raising government revenues and stimulating local and regional economies. Thus, in 2002 and 2003, debates on casinos, slot machines at race tracks, and even slot machines in bars and taverns took place in legislatures and among political leaders in Hawaii, Kentucky, Maryland,

Pennsylvania, New Hampshire, Rhode Island, Wisconsin, and Minnesota, as well as in other states.

Other events can also have impacts on the debates of whether or not to expand the presence of casinos. In October 2001, shortly after the September 11 terrorist attacks, the State of New York authorized six new tribal casinos and slot machine gaming at eight race tracks. This was motivated in no small measure by the need to close the gap against large impending state deficits, related both to the economic slowdown and to the anticipated economic consequences of the terrorist actions and the subsequent war on terrorism. The debate was hastened by the reality that by 2001, New York was surrounded by successful casino gaming operations in Atlantic City, Eastern Connecticut, and Ontario, Canada.

COMMERCIAL GAMING AND CONTROVERSY

Between 1982 and 2001, total gaming revenues of commercial gaming industries in the United States grew from $10.2 billion to $65.8 billion, with more than half of the 2001 total coming from commercial and tribal casinos. Lotteries, pari-mutuel wagering on racing, and charitable gambling, including bingo, all lost market share as casinos and electronic gaming devices increased their presence and popularity over the past two decades.

However, in spite of rapid economic expansions, general attitudes toward the acceptance of permitted casinos remained at best lukewarm in most jurisdictions. There is growing sentiment in a number of states that—at least in some situations—governments have authorized too much gaming. In such locales, there are pressures to reverse some of the trends that have characterized commercial gaming industries in the past three decades.

In some situations, substantial commercial gaming industries have seen their legal statuses revoked. This occurred when authorization for video poker machines in South Carolina was allowed to expire in 2000, eliminating an industry that was generating gross gaming revenues in excess of $500 million per annum. In 1996, local elections reversed the

legal status of video poker machines in 34 of the 66 parishes in Louisiana.

The issue of when gambling overextends its political welcome can be seen in recent events abroad. Such developments might provide insights into what may lie ahead for American jurisdictions wanting to fully exploit the economic rents from casino-style gambling. Following publication of a 1999 Productivity Commission Report on Gambling (Australian Productivity Commission Report 1999), Australia adopted a number of restrictions on electronic gaming devices after declaring problem gambling to be a public health issue, under a declared strategy of "harm minimization." This followed a decade where the number of electronic gaming devices in Australia expanded from about 70,000 to approximately 190,000, 90 percent of which were located outside of casinos (Monaghan 2001). The Productivity Commission Report claimed, among other findings, that the 2.1 percent of adult Australians who were problem gamblers made up 10 percent of regular players on gaming machines, and generated 42 percent of spending on gaming machines in Australia.

New technological developments in the gaming industries have also become part of the political controversy surrounding gambling. Perhaps the most dramatic of these is Internet gambling, whose legal status has been actively debated in many countries throughout the world, with no clear resolution in the early 21st century in general trends and directions. Internet gambling has a very large potential market and has the capability to bring highly sophisticated gaming products into households everywhere. Based on the spotty evidence that exists on this still largely "gray area" activity, the size of the global Internet gambling market is already measured in the billions of dollars (see, for example, Cabot 2000).

On the other hand, Internet gambling raises social concerns about the potential adverse impacts such ubiquitous gaming opportunities might bring about, especially in the areas of underage gambling and problem and pathological gambling. The activity also poses interesting challenges for jurisdictions on how to regulate and tax the activity, creating a dilemma for governments that are tolerant of permitted gaming primarily because of their ability to extract economic rents from excise taxes on the activity. Because Internet gambling operates with little concern for national borders, and because some jurisdictions have

decided to encourage Internet gaming sites to locate within their borders through offering low tax rates, other jurisdictions will have to match or come close to those tax rates to remain competitive. Furthermore, the United States, at both the congressional level and in various states, has demonstrated little desire to move forward to fully exploit the economic opportunities of Internet gambling.

It is possible that Internet gambling is just the tip of the technological iceberg. Interactive television betting and the use of various handheld computer devices for playing games and making wagers are perhaps the next major gambling developments. However, they will continue to be politically controversial because of the difficulties in exercising social controls over the activities where they take place, and because of the ability of new technologies to outstrip legislative attempts to constrain the presence or availability of gambling in general. A by-product of the new world of Internet and other low-cost and virtually instantaneous communications is likely to be the inability to significantly constrain gambling activities that take place through those media, regardless of the wishes and desires of legislative and parliamentary bodies.

A FRAMEWORK FOR EVALUATING THE BENEFITS AND COSTS OF EXPANDED GAMBLING

One aspect of legalizing new forms of permitted gambling is that such actions create benefits that impact economies—especially local or regional economies—in ways that are generally tangible, measurable, and economic. But an expanded presence of permitted gambling also generates social costs that affect individuals and households in ways that are far less tangible, measurable, and visible. It is extremely challenging to policymakers and social scientists to conceptualize, identify, and measure the social costs that accompany gambling in any meaningful way (see, for example, Walker and Barnett 1999 and Eadington 2003). Furthermore, because of the relative lack of attention to the costs and benefits of gambling prior to the mid 1990s, little serious effort was undertaken to address these issues.[4] It is likely that these dimensions of benefits and costs associated with gambling will remain

at the heart of the debate over the wisdom of expanding or contracting the availability of permitted gambling for some time to come.

Nonetheless, a number of observations can be made about the benefits and costs of permitted gambling in comparison to reasonable alternative states of nature. For those jurisdictions that are still debating the status of gaming within their control, such observations should prove useful.

First, it should be noted that the primary benefit associated with permitted gambling is the creation of *consumer surplus*, the incremental value to consumers from being able to participate in an activity that was previously prohibited. Consumer surplus is generally defined as the difference between what consumers would be willing and able to pay for an activity versus what they actually have to pay for that activity. Such gains accrue predominantly to the consumers of gambling services rather than to producers or the governments who authorize the activity.

However, when permitted gambling is authorized in a manner that prevents the market from expanding to its demand potential, or when the market structure is designed to result in monopoly or otherwise restricted competition, then the price of the activity increases. As a result, a portion of potential consumer surplus is diverted away from consumers and becomes value for someone else. The diverted consumer surplus can be referred to as *economic rents*. Economic rents can be captured by government through the implementation of excise taxes on the activity, or by outright ownership of the gaming franchise. Other economic rents might be captured by companies or organizations that offer gambling services through exclusive or limited franchises. Only when the market is allowed to expand to its demand potential, or when competition from related substitute activities bid down the price of the primary activity to competitive levels, are the economic rents bid away.[5]

As with other activities, most of the costs and benefits associated with permitted gambling are internal to the consumers and producers of the gambling activities. Under the assumption of rational economic actors, consumers choose to spend money on gambling because they derive greater value from participation than the expected or realized cost. Producers provide gambling services because it provides a greater return on their resources than the next best alternatives. As private ben-

efits and costs, there is little reason for public policy considerations to affect the decision processes that generate these allocations.

Public policy intervention is typically justified when negative externalities are associated with the activity.[6] With gambling, the primary negative externalities are linked to problem and pathological gambling. Generally speaking, there have been two major driving forces that have influenced societal decisions to liberalize gambling laws and regulations: 1) a desire on the part of governments to capture economic rents through permitting a previously prohibited activity; and 2) a desire to mitigate the negative side effects (real or perceived negative externalities) associated with the activity by constraining it in ways that would allow for greater control of the adverse side effects.

The combination of these two somewhat conflicting forces has led to a variety of eccentric laws passed in various jurisdictions throughout the world. In the United States, riverboat gambling with mandatory sailing, or mining town casinos with loss limits and restrictions to historic buildings only, reflect states' efforts to capture economic rents while providing protections against people who might overindulge in gambling activities. Voter ratification of Indian gaming, as in the State of California in 1998 and 2000, was a validation of the distribution of economic rents to tribes and tribal members; California's legislature and voters have been reluctant to bestow similar economic rents on other rent seekers, such as the card club industry or the racing industry.

As with other vices such as tobacco, alcohol, illicit drugs, and commercial sex, gambling is perceived as an activity that has a strong realized and latent demand that emanates from a portion of the population. As with the other vices, it also possesses a variety of negative side effects—perceived or real—that are viewed as immoral or otherwise socially damaging by (typically) another subset of the population. Such side effects have served as the impetus for constraints on the permitted offerings of gambling services. As with the other vices, there is no clear consensus on the best approach to regulating and constraining the availability of gambling, and as a result, there has not been much stability on the manner in which legal gambling has been permitted and constrained from one political jurisdiction to another.

The extent of demand for gambling that is realized—as opposed to remaining latent—is partly a function of gambling's legal status. If casino gaming, through the process of legalization or deregulation, is

made more attractive and available to a society's population centers, then the demand for gambling in general, and the total amount of income spent on gambling, will increase. The greater the availability of gambling, and the fewer the constraints that are applied to gaming activities, the larger the realized demand will be. Furthermore, the more that permitted gambling is offered in a competitive market context, the more that demand for gambling will increase. Increased competition will result in lower realized prices to consumers for gambling, and competition will enhance the price and availability of complementary nongaming activities as well. The recent experience of competitive venues such as Las Vegas and Mississippi—in comparison to more supply constrained or monopolistic jurisdictions such as the urban casinos in Detroit and New Orleans, or riverboat jurisdictions in Illinois, Indiana, or Louisiana—clearly demonstrate these effects.

When trying to evaluate social benefits and costs associated with gambling, it is important to evaluate the alternatives to the status of permitted gambling under consideration. If a jurisdiction currently prohibits gambling but has a substantial amount of illegal gambling taking place within its borders, removal of the prohibitions will likely diminish the adverse economic impacts of the illegal industry, and quite possibly will diminish the severity of some of the social costs associated with such illegal activities.

It is also useful to look at the general locational structure under which casino and casino-style gambling is offered, in terms of its potential for delivering benefits and costs. Though it is argued elsewhere that benefits and costs of permitted gambling should be done at the national level (Grinols and Mustard 2001), most policy analysis concentrates on local and regional economic benefits associated with permitted casinos and casino-style gaming. Using that as a starting point, one can create the following categories of casinos and near-casinos:

- Destination resort casinos located away from population centers (such as Las Vegas, Reno, and Lake Tahoe, Nevada; Biloxi, Mississippi; or Atlantic City, New Jersey).
- Rural casinos, located away from population centers (such as Foxwood's in Connecticut and most tribal casinos in the United

States; and the casinos in Deadwood, South Dakota; and in Central City, Blackhawk, and Cripple Creek, Colorado).

- Urban or suburban casinos located in or near major metropolitan areas (such as those found in Detroit, or in and around St. Louis, Kansas City, and Cincinnati), as well as most race track casinos ("racinos").
- Neighborhood casino-style gaming (such as video poker machines, video lottery terminals, and other gaming devices found in bars and taverns in such states as Nevada, Montana, Oregon, and South Dakota). This is sometimes referred to as *convenience gambling.*

If we compute benefits and costs for gambling in the traditional manner, and discount the importance of consumer surplus,[7] we find that jurisdictions that export gambling to citizens of other jurisdictions tend to capture a substantial amount of economic benefit in the form of economic rents and value added by producers and owners of local resources (i.e., the benefits of increased local employment), whereas the social costs associated with problem gambling in particular tend to get exported to the jurisdictions where the gambling consumers reside. In such cases, the ratio of benefits to costs within the jurisdiction is relatively high.

In a similar fashion, benefit/cost ratios for rural casinos are also fairly high, especially if the region for which the impacts are being evaluated includes only the rural area. This is often the case with Indian tribal casinos, where the primary group of interest is the tribe itself, and most of the casino customers are not tribal members.

On the other hand, if urban or suburban casinos are evaluated in this manner, the benefit/cost ratio is considerably lower. Most of the gaming activities provided by such casinos cater to demand in the local market. In such a case, spending on gambling does not stimulate the local economy in the same manner it would if gambling activities were exported. Furthermore, social costs typically remain within the community where the gaming facilities are located. Thus, measured benefits will be lower and social costs will be higher than in either of the first two cases. Nonetheless, such urban/suburban casinos can create significant regional investment and might serve as efficient mechanisms for tax revenue generation. Furthermore, they might bring about

considerable *import substitution* behavior, encouraging local residents who otherwise might travel out of the region to pursue gambling activities to spend their gambling budgets in local casinos instead.

If we consider the situation of convenience gambling—gaming devices in bars and taverns located in neighborhoods—the general tendency is for benefits to be lower and social costs to be higher than in any of the previous situations. Since such facilities generate little in the way of new investment or job creation associated with the gambling activities, economic benefits tend to be lower.[8] Because casino-style gaming is offered in more accessible surroundings than is typical for site-specific casinos, there might be a greater incidence of impulsive gambling and, as a result, of problem and pathological gambling.

The ratio of benefits to costs for a region or jurisdiction is a bellwether to the extent of controversy associated with the various types of permitted gambling. In light of this framework, especially when consumer surplus is given relatively little standing, it is not surprising to see convenience gambling as the most politically vulnerable of the alternatives considered. This thesis is consistent with the recent experiences in Australia noted previously, as well as jurisdictions such as South Carolina and Louisiana, where convenience gambling was eliminated or threatened with elimination because of the political backlashes associated with it.

This framework also carries implications for the new forms of gambling. Unless consumer surplus is given greater standing, Internet gambling and interactive television gambling, for example, will likely prove to be very low on perceived economic benefits and very high on social costs. Furthermore, the competitive and global dimensions of Internet gambling make it very difficult for governments to capture economic rents, especially in the form of taxes on gross gaming revenues. Also, the regulatory challenges of permitted gambling in the home, especially gambling by youth or by those prone to overindulge, imply that the social costs associated with such activities are going to be both socially dangerous and very hard to control without violating other dimensions of personal privacy. Thus, these newest forms of gambling might prove to be the most controversial of all.

CONCLUSION

In summary, the ongoing dynamics of the economic and social impacts of gambling and of permitted gaming industries point out a number of important dimensions characteristic of the activity, the industry, and of public policy processes regarding gambling. Most important of these are:

- Gambling is one of the largest industries whose fundamental economic characteristics are substantially determined by political decisions.
- Political decisions regarding gambling are largely influenced by the ability of competing special interests—including state governments—to capture economic rents associated with liberalizing permitted gaming activities. This is often countered by perceived or real social costs associated with problem and pathological gambling and with an increased availability of gambling in society.
- There is a strong latent demand for casino-style gaming (including gaming within casinos and with electronic gaming devices located outside of casinos), which is manifested when the legal status of gambling is liberalized.
- Technologies have developed over the past two decades that have broadened the appeal of, and the market for, commercial gaming. The same technologies have raised concerns over some adverse social impacts that such an increased presence of gambling in society might bring about. Many of these adverse social impacts are related to problem and pathological gambling behavior.
- Benefit/cost analysis applied to permitted gaming activities is still a relatively primitive science, primarily because of the difficulties in conceptualizing, observing, and measuring social costs. Because of its lingering status as a vice, consumer surplus associated with gambling consumption is often discounted in policy discussions.
- Some types of permitted gambling raise greater social concerns over their impacts than do others. Some categories of venues for

casinos and casino-style gambling are more vulnerable to political controversy and possibly reversal of liberalization of permitted activities than are others. The forms of gambling with the greatest potential for controversy include convenience gambling, Internet gambling, and interactive television gambling.

The debate over the proper role of permitted gambling in society is far from over, though there are some clear long-term trends—visible for much of the past half-century—that have supported increased legalization and deregulation in many jurisdictions. In many respects, these trends reflect society's increased acceptance of gambling as a proper form of (adult) leisure and entertainment.

However, as has been demonstrated in various situations, public attitudes toward gambling can be fickle. Should significant problems arise—such as corruption scandals, the presence of organized crime, or even sensational incidents involving pathological gamblers—gambling might once again come under fire. If the perceptions of social costs associated with gambling become substantial relative to the economic benefits that it is creating, then the political winds can quickly shift harshly against its permitted status. Unless and until respect for gambling as a consumption activity achieves a level comparable with other consumption activities, newer types of permitted gambling will continue to raise public policy debates and remain at the center of political controversies.

Notes

1. The states were South Dakota, Iowa, Illinois, Colorado, Louisiana, Mississippi, Missouri, and Indiana. Only Michigan, where voters authorized three casinos in Detroit in 1996, was added to this list between 1994 and 2001.
2. *California et al. v. Cabazon Band of Mission Indians et al.*, No 85-1708, February 25, 1987; reprinted in Eadington (1990).
3. P.L. 100-497 (1988), 100th Congress; reprinted in Eadington (1990).
4. Until casinos spread beyond Nevada and Atlantic City in the United States, there was little in the way of institutionally funded research on gambling. Similar circumstances prevailed in other countries. Since the 1990s, there have been a number of major national studies undertaken in various countries, including the *Final Report* (National Gambling Impact Study Commission 1999) in the United States, the *Gambling Review Report* (Department for Culture, Media and Sport 2001) in

the United Kingdom, and *Australia's Gambling Industries* (Productivity Commission 1999) in Australia.

5. It should also be noted that the idea of consumer surplus has seldom been an important factor in deliberations regarding legalizing or deregulating gambling. This is probably because of long-standing prejudices that gambling is a tainted activity, and people who participate in gambling are themselves exercising poor judgment in their consumption choices, and should therefore not be given much consideration in deliberations. As a result, most policy deliberation relies primarily on the magnitude and distribution of the economic rents.
6. Negative externalities arise when the market transactions between two parties create costs for third parties who are not involved in the transactions. Without policy intervention, this shifting of costs results in overproduction of the activity that creates negative externalities.
7. See note 6.
8. It should be noted, however, that such gaming devices might be extremely efficient tax collectors.

References

Cabot, Anthony. 2000. *The Internet Gambling Report,* 4th ed. Las Vegas: Trace Publishing.

Department for Culture, Media and Sport. 2001. *Gambling Review Report*. Norwich, U.K.: Stationery Office. Available at: <www:culture.gov.UK/>.

Eadington, William R. 1990. *Indian Gaming and the Law.* Reno: University of Nevada.

———. 2003. "Measuring Costs from Permitted Gaming: Concepts and Categories in Evaluating Gambling's Consequences." *Journal of Gambling Studies* 19(2): 185–212.

Grinols, Earl L., and David B. Mustard. 2001. "Business Profitability Versus Social Profitability: Evaluating Industries with Externalities—The Case of Casinos." *Managerial and Decision Economics* 22(1): 143–162.

Monaghan, Sean. 2001. "Australian Gaming Machine Review: Assessing the Pace and Cost of Harm Minimization." Analyst report, Sydney: Burdett, Buckeridge Young.

National Gambling Impact Study Commission. 1999. *Final Report*. Washington, DC: Government Printing Office.

Productivity Commission. 1999. *Australia's Gambling Industries.* Melbourne; Canberra: Productivity Commission. Available at: <http: www.pc.gov.au/inquiry/gambling/index.html>.

Walker, Douglas M., and A.H. Barnett. 1999. "The Social Costs of Gambling: An Economic Perspective." *Journal of Gambling Studies* 15(3): 181–212.

4
Common Value Auctions and the Winner's Curse

Lessons from the Economics Laboratory

John H. Kagel
Ohio State University

Auctions are of considerable practical and theoretical importance. In practical terms, the value of goods exchanged in auctions each year is huge. Governments routinely use auctions to purchase goods and services, to sell government assets, and to fund the national debt. Private sector auctions are common as well, and are of growing importance in areas such as deregulated utility markets, allocation of pollution rights, and the large variety of items now being sold via Internet auctions. Auctions are commonly employed when one party to the exchange (for example, the seller) is uncertain about the value that buyers place on the item; they provide a mechanism, absent middlemen, to establish value in such situations. Auctions play a prominent role in the theory of exchange, as they remain one of the simplest and most familiar means of price determination in the absence of intermediate market makers. In addition, auctions serve as valuable illustrations, and one of the most prominent applications, of games of incomplete information, as bidders' private information is the main factor affecting strategic behavior (Wilson 1992).

There are at least two distinct types of risk in auctions. In private value auctions, where bidders know the value of the item to themselves with certainty, there is uncertainty regarding other bidders' values. In first-price sealed-bid auctions, in which buyers simultaneously submit sealed bids with the high bid winning the item at the price bid, bidders

face a strategic trade-off: the lower their bid the higher their surplus conditional on winning, but the lower their probability of winning.[1] Further, the famous (within economic circles, at least) Vickrey auction (Vickrey 1961), in which the high bid wins but pays the second-highest bid price, was designed with the specific purpose of eliminating this strategic uncertainty. Bidders in the Vickrey auction have a dominant strategy of bidding their valuations, so they do not have to consider this strategic trade-off. (Similarly, in an open outcry English auction in which bidding starts out low and the auctioneer gradually raises the price, bidders have a dominant strategy to remain active until the price reaches their valuation. Hence, there are no strategic trade-offs here as well.)

Common value auctions, the other canonical type of auction, introduce a whole new risk dimension. In a pure common value auction, the value of the item is the same to all bidders. What makes the auction interesting is that bidders do not know the value at the time they bid. Instead they receive signal values that are correlated—or, more technically, affiliated (Milgrom and Weber 1982)—with the value of the item, so that bidders must estimate the common value based on their private information signals, while still wrestling with the strategic issues associated with private value auctions. Mineral rights auctions, particularly the federal government's outer continental shelf (OCS) oil lease auctions, are typically modeled as pure common value auctions. There is a common value element to most auctions. For example, bidders for an oil painting may purchase for their own pleasure, constituting a private value element, but they may also bid for investment and eventual resale, reflecting the common value element.

There are no efficiency issues in pure common value auctions, as all bidders place equal value on the item.[2] What has been of overriding concern to both theorists and practitioners for these auctions are the revenue-raising effects of different auction institutions. A second key issue, one that has provided much of the focus for both experimental and empirical work on common value auctions, is the winner's curse, an unpredicted effect that was initially postulated on the basis of field data, and whose existence has often been hotly debated among economists.

The winner's curse story begins with Capen, Clapp, and Campbell (1971), three petroleum engineers who claimed that oil companies had

suffered unexpectedly low rates of return in the 1960s and 1970s on OCS lease sales "year after year."[3] They argued that these low rates of return resulted from the fact that winning bidders ignored the informational consequences of winning. That is, bidders naively based their bids on the unconditional expected value of the item (their own estimates of value), which, although correct on average, ignores the fact that you only win when your estimate happens to be the highest (or one of the highest) of those competing for the item. But winning against a number of rivals following similar bidding strategies implies that your estimate is an overestimate of the value of the lease conditional on the event of winning. Unless this adverse selection effect is accounted for in formulating a bidding strategy, it will result in winning bids that produce below normal or even negative profits. The systematic failure to account for this adverse selection effect is commonly referred to as the winner's curse: you win, you lose money, and you curse.

Terminological aside: When discussing the winner's curse, many economists, particularly theorists, unfortunately use the term to refer to the difference between the expected value of the item conditional on the event of winning and the naive expectation (not conditioning on the event of winning). Further, their use of the term typically refers to players who fully account for this winner's curse, rather than those who fall prey to it.

The idea that oil companies suffered from a winner's curse in OCS lease sales was greeted with skepticism by many economists, as it implies that bidders repeatedly err, violating basic economic notions of rationality and contrary to equilibrium predictions.[4] An alternative and simpler explanation as to why oil companies might claim that they fell prey to a winner's curse lies in cartel theory, as responsiveness to the winner's curse claim could serve as a coordination device to get rivals to reduce their bids in future sales. Nevertheless, claims that bidders fell prey to the winner's curse have arisen in a number of field settings. In addition to the oil industry (Capen, Clapp, and Campbell 1971; Lorenz and Dougherty 1983 and references cited therein), claims have been made in auctions for book publication rights (Dessauer 1981), professional baseball's free agency market (Cassing and Douglas 1980; Blecherman and Camerer 1998), corporate takeover battles (Roll 1986), and in real estate auctions (Ashenfelter and Genesore 1992).

It is exceedingly difficult to support claims of a winner's curse using field data because of reliability problems with the data and because alternative explanations for overbidding are often available. For example, Hendricks, Porter, and Boudreau (1987) found that in early OCS lease sales, average profits were negative in auctions with seven or more bidders. They note that one possible explanation for this outcome is the increased severity of the adverse selection problem associated with more bidders. However, they note that the data could also be explained by bidder uncertainty regarding the number of firms competing on a given tract (their preferred explanation). That is, since most tracts received less than six bids, it seems likely that firms would expect this number or less. As a result, although firms might have fully accounted for the adverse selection effect based on the expected number of firms bidding on a tract, they would nevertheless be incorrect for tracts that attracted above average numbers of bidders, and overbid on those tracts.

The ambiguity inherent in using field data, in conjunction with the controversial nature of claims regarding a winner's curse, provided the motivation for experimental studies of the winner's curse. Early laboratory experiments showed that inexperienced bidders are quite susceptible to the winner's curse (Bazerman and Samuelson 1983; Kagel and Levin 1986; Kagel et al. 1989). In fact, the winner's curse has been such a pervasive phenomenon in the laboratory that most of these initial experiments have focused on its robustness and the features of the environment that might attenuate its effects. Additional interest has focused on public policy issues—the effects of public information regarding the value of the auctioned item and the effects of different auction institutions on sellers' revenue.

This survey begins with a brief analysis of the first experimental demonstration of the winner's curse (Bazerman and Samuelson 1983). This is followed by summaries of experiments investigating bidding in common value auctions using an experimental design that I helped develop. These experiments also demonstrate the existence of a winner's curse even when allowing for extensive feedback and learning from past auction outcomes. They also address policy issues such as the effects of public information and different auction institutions (e.g., first-price sealed-bid auctions versus open outcry English auctions) on sellers' revenue. I conclude with a brief summary of the empirical find-

ings from the experimental literature and the role experiments have played in the successful sale of government airwave rights (the spectrum auctions). In reviewing the experimental work on common value auctions, I hope to show how experiments proceed by successively narrowing down plausible explanations for the question at hand. This is done through a series of experiments rather than any single "critical" experiment; it is based on sorting out between competing explanations, and on following up on the logical implications of behavior observed in earlier experiments.

AN INITIAL EXPERIMENT DEMONSTRATING THE WINNER'S CURSE

Bazerman and Samuelson (1983) conducted the first experiment demonstrating a winner's curse. Using M.B.A. students at Boston University, the experiment was conducted in class, with students participating in four first-price sealed-bid auctions. Bidders formed their own estimates of the value of each of four commodities—jars containing 800 pennies, 160 nickels, 200 large paper clips each worth four cents, and 400 small paper clips each worth $0.02. Unknown to subjects, each jar had a value of $8.00. (Subjects bid on the value of the commodity, not the commodity itself.) In addition to their bids, subjects provided their best estimate of the value of the commodities and a 90 percent confidence bound around these estimates. A prize of $2.00 was given for the closest estimate to the true value in each auction. The number of bidders varied between 4 and 26. Their analysis focused on bidder uncertainty about the value of the commodity and the size of the bidding population.

The average value estimate across all four commodities was $5.13 ($2.87 below the true value). As the authors note, this underestimation should reduce the likelihood and magnitude of the winner's curse. In contrast to the mean estimate, the average winning bid was $10.01, resulting in an average loss to the winner of $2.01.[5] The average winning bid generated losses in over half of all the auctions.

Estimated bid functions, using individual bids as the unit of observation, showed that bids were positively, and significantly, related to

individual estimates so that bidders indeed faced an adverse selection problem, only winning when they had higher estimates of the value of the item. Bids were inversely related to the uncertainty associated with individual estimates, but this effect was small (other things equal, a $1.00 increase in the 90 percent confidence interval reduced bids by $0.03). Numbers of bidders had no significant effect on individual bids.

In contrast, regressions employing the average winning bid showed that these bids were positively, and significantly, related to the winning bidder's estimate of uncertainty and to the number of bidders in the auction. This suggests that winning bidders are substantially more aggressive than other bidders. Indeed, Bazerman and Samuelson note that average winning bids were sensitive to a handful of grossly inflated bids.

The results of this experiment show that the winner's curse is easy to observe. However, many economists would object to the fact that subjects had no prior experience with the problem and no feedback regarding the outcomes of their decisions between auctions, so that the results could be attributed to the mistakes of totally inexperienced bidders. The robustness of these results is even more suspect given their sensitivity to a handful of grossly inflated bids, which one might suppose would be eliminated as a result of bankruptcies or learning in response to losses incurred in earlier auctions. Common value auction experiments conducted by Kagel and Levin (1986) and their associates explore these issues, along with a number of public policy implications of the theory.

SEALED-BID AUCTIONS

Kagel and Levin and their associates conducted experiments in which bidders participated in a series of auctions with feedback regarding outcomes. Bidders were given starting cash balances from which losses were subtracted and profits were added. Bidders whose cash balances became negative were declared bankrupt and were no longer permitted to bid. Unlike the Bazerman and Samuelson experiment, Kagel and Levin controlled the uncertainty associated with the value of the

auctioned item rather than simply measuring it. They did this by conducting auctions in which the common value, x_o, was chosen randomly each period from a known uniform distribution with upper and lower bounds $[\underline{x}, \overline{x}]$. In auctions with a symmetric information structure, each bidder is provided with a private information signal, x, drawn from a uniform distribution on $[x_o - \varepsilon, x_o + \varepsilon]$, where ε is known. In first-price sealed-bid auctions, bids are ranked from highest to lowest with the high bidder paying the amount bid and earning profits equal to $x_o - b_1$, where b_1 is the highest bid. Losing bidders neither gain nor lose money.

In this design, the strategy of bidding, max $[x - \varepsilon, \underline{x}]$, is a risk-free strategy that fully protects a bidder from negative earnings since it is the lower bound estimate of x_o. This lower bound estimate for x_o was computed for subjects along with an upper bound estimate of x_o, (min $[x + \varepsilon, \overline{x}]$**)**. Bidders were provided with illustrative distributions of signal values relative to x_o, and several dry runs were conducted before playing for cash. Following each auction period, bidders were provided with the complete set of bids, listed from highest to lowest, along with the corresponding signal values, the value of x_o, and the earnings of the high bidder.

Surviving bidders were paid their end-of-experiment balances in cash. To hold the number of bidders fixed while controlling for bankruptcies, $m > n$ subjects were often recruited, with only n bidding at any given time (who bids in each period was determined randomly or by a fixed rotation rule). As bankruptcies occur, m shrinks but (hopefully) remains greater than or equal to the target value n.

Some Initial Experimental Results: Inexperienced Bidders

Auctions with inexperienced bidders show a pervasive winner's curse that results in numerous bankruptcies. Table 4.1 provides illustrative data on this point. For the first nine auctions, profits averaged –$2.57, compared to the risk neutral Nash equilibrium (RNNE) prediction of $1.90, with only 17 percent of all auctions having positive profits. Note: this is after bidders had participated in two or three dry runs, with feedback of signal values, x_o, and bids following each auction, so that the results cannot be attributed to a total lack of experience. The negative profits are not a simple matter of bad luck either, or a handful of grossly inflated bids, as 59 percent of all bids and 82

Table 4.1 Profits and Bidding in First Nine Auctions for Inexperienced Bidders

Experiment	Percent of auctions with positive profits	Average actual profits (*t*-statistic)	Average predicted profits under RNNE (S_m)	Percent of all bids $b>E[x_o/X=x_{ln}]$	Percent of auctions won by high signal holder	Percentage of high bids $b_1>E[x_o/X=x_{ln}]$	Percentage of subjects going bankrupt[a]
1	0.0	–4.83 (–3.62)***	0.72 (0.21)	63.4	55.6	100	50.0
2	33.3	–2.19 (–1.66)	2.18 (1.02)	51.9	33.3	88.9	16.7
3	11.1	–6.57 (–2.80)**	1.12 (1.19)	74.6	44.4	88.9	62.5
4	11.1	–2.26 (–3.04)***	0.85 (0.43)	41.8	55.6	55.6	16.7
5	33.3	–0.84 (–1.00)	3.60 (1.29)	48.1	44.4	88.9	50.0
6	22.2	–2.65 (–1.53)	2.55 (1.17)	67.3	66.7	100	33.3
7	11.1	–2.04 (–2.75)**	0.57 (0.25)	58.5	88.9	66.7	50.0
8	11.1	–1.40 (–2.43)**	1.59 (0.34)	51.9	55.6	55.6	16.7
9	44.4	0.32 (0.30)	2.37 (0.76)	35.2	88.6	66.7	16.7

10	0.0	–2.78 (–3.65)***	3.53 (0.74)	77.2	66.7	100	20.0
11	11.1	–3.05 (–3.53)***	1.82 (0.29)	81.5	55.6	88.9	37.5
Average	17.2	2.57	1.90	59.4	59.6	81.8	41.1

NOTE: S_m = standard error or mean. **significant at the 5% level, two-tailed test; ***significant at the 1% level, two-tailed test.
[a] For all auctions.
SOURCE: From Kagel et al. (1989). Reprinted with permission from Western Economic Association International.

percent of the high bids were above $E[x_o|X = x_{1n}]$; the expected value of x_o conditional on having the highest signal x. Further, 40 percent of all subjects starting these auctions went bankrupt. In short, the winner's curse is a genuinely pervasive problem for inexperienced bidders. It is remarkably robust being reported under a variety of treatment conditions (Kagel et al. 1989; Lind and Plott 1991; Goeree and Offerman 2000) and for different subject populations, including professional bidders from the commercial construction industry (Dyer, Kagel, and Levin 1989).

Auctions with Moderately Experienced Bidders and the Effects of Public Information on Sellers' Revenue

Kagel and Levin (1986) report auctions for moderately experienced bidders (those who had participated in at least one prior first-price common value auction experiment). Treatment variables of interest were the number of rival bidders and the effects of public information about x_o on revenue. Table 4.2 reports some of their results. For small groups (auctions with 3–4 bidders), the general pattern was one of positive profits averaging $4.32 per auction, which is significantly greater than zero but still well below the RNNE prediction of $7.48 per auction. In contrast, for these same bidders bidding in larger groups (auctions with 6–7 bidders), profits averaged –$0.54 per auction compared to the RNNE prediction of $4.82. Thus, the profit picture had improved substantially compared to the inexperienced bidders discussed in the previous section.

However, comparing large and small group auctions, actual profit decreased substantially more than profit opportunities as measured by the RNNE criteria. This implies that subjects were bidding more aggressively, rather than less aggressively, as the number of rivals increased, contrary to the RNNE prediction. This is confirmed in regressions using individual subject bids as the dependent variable. Higher individual bids in response to increased numbers of rivals is often considered to be the hallmark characteristic of a winner's curse. Thus, although bidders had adjusted reasonably well to the adverse selection problem in auctions with 3–4 bidders, in auctions with 6–7 bidders, with its heightened adverse selection effect, the winner's curse reemerged as subjects confounded the heightened adverse selection

effect by bidding more aggressively with more bidders. This result also suggests that the underlying learning processes are context-specific rather than involving some sort of "theory absorption" that readily generalizes to new environments.[6]

Public information was provided to bidders in the form of announcing the lowest signal value, x_L. For the RNNE, public information about the value of the item raises expected revenue. The mechanism underlying this outcome works as follows: All bidders evaluate the additional public information assuming that their signal is the highest since, in equilibrium, they only win in this case. Evaluating additional information from this perspective, together with affiliation, induces all bidders other than the highest signal holder to, on average, revise their bids upward after an announcement of unbiased public information. This upward revision results from two factors:

1) Bidders without the highest signal treat the public information as "good news." These bidders formulated their bids on the assumption that they held the highest private information signal and would win the auction. As such, with affiliation, the public information tells them that, on average, the expected value of the item is higher than they had anticipated (i.e., the private information signal they are holding is somewhat lower than expected, conditional on winning, for this particular auction), which leads them to increase their bids.

2) Bidders respond to this anticipated increase in bids from lower signal holders by raising their bids. The bidder with the highest signal is not, on average, subject to this first force. Thus, she does not, on average, revise her estimate of the true value. Nevertheless, she raises her bid in anticipation of other bidders raising their bids; the "domino" effect of bidders with lower signals raising their bids.

These strategic considerations hold for a wide variety of public information signals (Milgrom and Weber 1982). There are, however, several methodological advantages to using x_L. First, the RNNE bid function can be readily solved for x_L, provided low signal holders are restricted to bidding x_L, so that the experimenter continues to have a benchmark model of fully rational behavior against which to compare

Table 4.2 Profits and Bidding by Experiment and Number of Active Bidders: Private Information Conditions (profits measured in dollars)

Auction series (no. of periods)	Number of active bidders	Average actual profit (*t*-statistic)[a]	Average profit under RNNE (standard error of mean)	Percent of auctions won by high signal holder	Percent of high bids $b_1 > E[x_o / X=x_{1n}]$
6		3.73	9.51		
(31)	3–4	(2.70)**	(1.70)	67.7	22.6
2		4.61	4.99		
(18)	4	(4.35)***	(1.03)	88.9	0.0
3 small		7.53	6.51		
(14)	4	(2.07)	(2.65)	78.6	14.3
7 small		5.83	8.56		
(19)	4	(3.35)***	(2.07)	63.2	10.5
8 small		1.70	6.38		
(23)	4	(1.56)	(1.21)	82.6	39.1
1		2.89	5.19		
(18)	5	(3.14)***	(0.86)	72.2	27.8
3 large		–2.92	3.64		
(11)	5–7	(–1.49)	(0.62)	81.8	63.6
7 large		1.89	4.70		
(18)	6	(1.67)	(1.03)	72.2	22.2
4		–0.23	4.78		
(25)	6–7	(–0.15)	(0.92)	69.2	48.0

5 (26)	7	–0.41 (–0.44)	5.25 (1.03)	42.3	65.4
8 large (14)	7	–2.74 (–2.04)	5.03 (1.40)	78.6	71.4
Small market average	3–4	4.32 (5.55)***	7.48 (0.77)	75.2	19.0
Large market average	6–7	–0.54 (0.87)	4.82 (0.50)	62.9	53.9

NOTE: **Significant at the 5% level, two-tailed *t*-test; ***significant at the 1% level, two-tailed *t*-test.

[a] Tests null hypothesis that mean is different from 0.0.

SOURCE: From Kagel and Levin (1986). Reprinted with permission of the American Economic Association.

actual bidding. Second, x_L provides a substantial dose of public information about x_0 (it cuts expected profit in half), while still maintaining an interesting auction. As such it should have a substantial impact on prices, regardless of any inherent noise in behavior. Finally, the experimenter can always implement finer, more subtle probes of public information after seeing what happens with such a strong treatment effect.[7]

Kagel and Levin (1986) found that in auctions with small numbers of bidders (3–4), public information resulted in statistically significant increases in revenue that averaged 38 percent of the RNNE model's prediction. However, in auctions with larger numbers of bidders (6–7), public information *reduced* average sellers' revenue by $1.79 per auction, compared to the RNNE model's prediction of an increase of $1.78. Kagel and Levin attribute this reduction in revenue to the presence of a relatively strong winner's curse in auctions with large numbers of bidders. If bidders suffer from a winner's curse, the high bidder consistently overestimates the item's value, so that announcing x_L is likely to result in a downward revision of the most optimistic bidders' estimate. Thus, out of equilibrium, public information introduces a potentially powerful offset to the forces promoting increased bids discussed earlier, and will result in reduced revenue if the winner's curse is strong enough. This hypothesis is confirmed using detailed data from auctions with 6–7 bidders, which shows that the RNNE model's prediction of an increase in sellers' revenue is critically dependent on whether or not there was a winner's curse in the corresponding private information market.

Is the Winner's Curse a Laboratory Artifact? Limited Liability for Losses

Results of experiments are often subject to alternative explanations. These alternative explanations typically provide the motivation for subsequent experiments that further refine our understanding of behavior. This section deals with one such alternative explanation and the responses to it.

In the Kagel and Levin (1986) design, subjects enjoyed limited liability as they could not lose more than their starting cash balances. Hansen and Lott (1991) argued that the overly aggressive bidding reported in Kagel and Levin may have been a rational response to this

limited liability rather than a result of the winner's curse. In a one-shot auction, if a bidder's cash balance is zero, so that they are not liable for *any* losses, it indeed pays to overbid relative to the Nash equilibrium bidding strategy. With downside losses eliminated, the only constraint on more aggressive bidding is the opportunity cost of bidding more than is necessary to win the item. In exchange, higher bids increase the probability of winning the item and making positive profits. The net effect, in the case of zero or small cash balances, is an incentive to bid more than the Nash equilibrium prediction. Hansen and Lott's argument provides a possible alternative explanation to the overly aggressive bidding reported in Kagel and Levin (1986) and in Kagel et al. (1989).

Responses to the limited-liability argument have been twofold. First, Kagel and Levin (1991) reevaluated their data in light of Hansen and Lott's arguments, demonstrating that for almost all bidders cash balances were *always* large enough so that it *never* paid to deviate from the Nash equilibrium bidding strategy in a one-shot auction. Second, subsequent empirical work has demonstrated a winner's curse in experimental designs where limited liability for losses could not logically account for overbidding. This provides experimental verification that limited-liability forces do not account for the overly aggressive bidding reported.

Kagel and Levin's design protects against limited-liability problems since bidding $x - \varepsilon$ insures against all losses and bidders have their own personal estimate of the maximum possible value of the item ($\min [x + \varepsilon, \ \overline{x}]$). The latter implies that it is never rational, limited liability or not, to bid above this maximum possible value in a first-price auction. Further, cash balances only have to be a fraction of the maximum possible loss for the limited-liability argument to lose its force in a first-price auction. For example, Kagel and Levin (1991) report simulations for auctions with 4 or 7 bidders, with $\varepsilon = \$30$ and cash balances of \$4.50 (which 48 out of the 50 bidders always had), for which unilateral deviations from the RNNE bid function were not profitable even when fully accounting for bidders' limited liability. Further, limited-liability arguments imply more aggressive bidding in auctions with fewer rather than larger numbers of bidders, just the opposite of what the data show.[8] As such, overbidding in the Kagel and Levin experi-

ment must be explained on some other grounds, such as the judgmental error underlying the winner's curse.

Empirical work on this issue has proceeded on several fronts. Lind and Plott (1991) replicated Kagel and Levin's results in auctions where bankruptcy problems were almost completely eliminated. One experimental treatment involved conducting private value auctions where subjects were sure to make money simultaneously with the common value auctions, thereby guaranteeing a steady cash inflow against which to charge any losses incurred in the common value auctions. A second treatment involved sellers' markets in which bidders tendered offers to sell an item of unknown value. Each bidder was given one item with the option to keep it and collect its value or to sell it. Lind and Plott's results largely confirm those reported by Kagel and Levin and their associates.

Cox, Dinkin, and Smith (1998) conducted auctions using Kagel and Levin's design in which, under one treatment, they reinitialize bidders' cash balances in each auction period, with balances large enough that subjects could not go bankrupt even if biding well above their signal values. In contrast to this unlimited-liability treatment, their other treatments employed procedures where cash balances fluctuated, bidders could go bankrupt, and in some treatments, bidders with negative cash balances were permitted to continue to bid. Using data for all treatments and all levels of bidder experience, Cox, Dinkin, and Smith find no significant differences in individual bid patterns in the unlimited-liability treatment, contrary to Hansen and Lott's argument. Further, restricting their analysis to experiments with experienced subjects, and dropping data from an entire experiment if even one subject adopted a pattern of high bids when having a negative cash balance, Cox, Dinkin, and Smith find that the unlimited-liability treatment significantly *increased* individual bids, the exact opposite of Hansen and Lott's hypothesis. This unexpected outcome is, however, consistent with Kagel and Levin's (1991) argument that in a multi-auction setting, where cash balances carry over from one auction to the next, there is a potentially powerful offset to any limited-liability forces present in a one-shot auction: Overly aggressive bidding due to low cash balances may be offset by the risk that such bids will result in bankruptcy, thereby preventing participation in later auctions with their positive expected profit opportunities. Unfortunately, it is also consis-

tent with the artifactual explanation that because subjects were paid off in only a few of the unlimited-liability auctions (in order to keeps costs to a manageable level), subjects treated these auctions differently than those in which they were paid as a result of each outcome.[9]

Summary

Even after allowing for some learning as a result of feedback regarding past auction outcomes, a strong winner's curse is reported for inexperienced bidders in sealed-bid common value auctions. High bidders earn negative average profits and consistently bid above the expected value of the item conditional on having the high signal value. Further, this is not the result of a handful of overly aggressive bidders but applies rather broadly across the sample population. Similar results are reported in low-bid wins, supply auctions with both student subjects and professional bidders drawn from the commercial construction industry (Dyer, Kagel, and Levin 1989). Arguments that these results can be accounted for on the basis of limited liability for losses have been shown to be incorrect. Further, a clever experiment by Holt and Sherman (1994) (also see Avery and Kagel 1997) is able to rule out the idea that the winner's curse is a result of an added thrill, or extra utility, from winning.

Note that the overbidding associated with the winner's curse is not simply a matter of miscalibrated bidders, it is associated with fundamental breakdowns of the comparative static predictions of the rational bidding model: With a winner's curse public information reduces revenue, contrary to the theory's prediction, as the additional information helps high bidders to correct for overly optimistic estimates of the item's worth. In second-price sealed-bid auctions, increased numbers of bidders produce no change in bidding, contrary to the robust Nash equilibrium prediction that bids will decrease (Kagel, Levin, and Harstad 1995).

We are still left with the puzzle, first expressed by Lind and Plott, that although many experiments report a clear winner's curse (negative profits), comparing between the symmetric RNNE and totally naive bidding models offered in the literature (all players treat their signals as if they are private values and go on to bid as if in a private value auction; Kagel and Levin 1986), bidding is closer to the RNNE. One

promising explanation for this phenomenon appears to be that bidders are cursed to different degrees. That is, agents may make partial, but incomplete, adjustments for the adverse selection effect associated with common value auctions, with the perfectly rational and perfectly naive bidding models being polar cases. Depending on the extent to which players are "cursed," they may suffer losses, but bidding can, in fact, still be closer to the symmetric RNNE bidding model than the totally naive bidding model. (See Eyster and Rabin 2000 for a formal model of this sort.)

ENGLISH AUCTIONS AND FIRST-PRICE AUCTIONS WITH INSIDER INFORMATION

My colleagues and I have also studied English auctions and first-price auctions with insider information (one bidder knows the value of the item with certainty and this is common knowledge). These experiments were initially motivated by efforts to identify institutional structures that would eliminate, or mitigate, the winner's curse for inexperienced bidders. The experiments also investigate the comparative static properties of Nash equilibrium bidding models for very experienced bidders. In both institutional settings, the winner's curse is alive and well for inexperienced bidders, although it is clearly less severe in English than in first-price auctions.

In contrast, comparative static predictions of the Nash equilibrium bidding model are largely satisfied for more experienced bidders. However, in the case of English auctions, the information processing mechanism that the Nash bidding model specifies is not satisfied. Rather, bidders follow a relatively simple rule of thumb that results in almost identical prices and allocations as the Nash model's predictions for the distribution of signal values employed in the experiment. In the insider information auctions, less-informed bidders (outsiders) have some proprietary information (i.e., the insider knows the value of the item with certainty but does not know the outsiders' signals). This results in marked differences in predicted outcomes compared to the standard insider information model in which the insider has a double informational advantage—she knows the value of the item and the sig-

nals the outsiders have (Wilson 1967; Weverbergh 1979; Englebrecht-Wiggans, Milgrom, and Weber 1983; Hendricks, Porter, and Wilson 1994). Most notably, in our model the existence of an insider generates higher average revenue than in auctions with a symmetric information structure, a prediction that is satisfied in the data for experienced bidders. In contrast, in the double informational advantage model the existence of an insider reduces average revenue.

English Auctions

Levin, Kagel, and Richard (1996) implement an irrevocable exit, ascending-price (English) auction. Prices start at $\underline{x}$, the lowest possible value for x_o, and increase continuously. Bidders are counted as actively bidding until they drop out of the auction and are not permitted to reenter once they have dropped out. The last bidder earns a profit equal to x_o less the price at which the last bidder dropped out. Bidders observe the prices at which their rivals drop out of the bidding. Auctions of this sort have been run in Japan (Milgrom and Weber 1982; Cassady 1967). The irrevocable exit procedure, in conjunction with the public posting of drop-out prices, insures that in equilibrium bidders can infer their rivals' signal values from their drop-out prices.

In a symmetric RNNE, the bidder with the low signal value (x_L) drops out of the auction once the price reaches his signal value.[10] The price at which the low bidder drops out of the auction reveals his signal value to the remaining bidders. Thus, the public information, x_L, that was provided by the experimenters in Kagel and Levin (1986) is provided endogenously here (at least in theory) by the first drop-out price. Given the uniform distribution of signal values around x_o, in a symmetric equilibrium, for any remaining bidder j, $(x_L + x_j)/2$ provides a sufficient statistic for x_o conditional on x_j being the highest signal, so that drop-out prices other than x_L contain no additional information and should be ignored. This sufficient statistic is the equilibrium drop-out price for j (d_j) in the symmetric RNNE

$$d_j = (x_L + x_j)/2.$$

This represents the maximum willingness to pay, conditional on all the information revealed by earlier drop-out prices and conditional on win-

ning. As in first-price auctions with x_L publicly announced, expected profit in the English auction is sharply reduced (by about a half) compared to first-price auctions with strictly private information (as long as $n > 2$). As such, in equilibrium, the English auction is predicted to significantly raise average sellers' revenue compared to first-price sealed-bid auctions.

The key difference between the English auction and a first-price sealed-bid auction with x_L publicly announced is that in the English auction information dissemination is endogenous, rather than exogenous. Higher signal holders must be able to recognize and process the relevant information, and low signal holders must recognize the futility of remaining active once the price exceeds their signal value. As such, we would expect the information dissemination process to be noisier than with x_L publicly announced. Nevertheless, if bidders are able to correctly recognize and incorporate the public information inherent in other bidders' drop-out prices, we would predict two results: 1) for inexperienced bidders, *contrary* to the Nash equilibrium bidding model's prediction, English auctions will reduce average sellers' revenues compared to first-price sealed-bid auctions, as losses will be sharply reduced, or even be eliminated, on average, in the English auctions, and 2) for more experienced bidders, where negative average profits have been largely eliminated in the sealed-bid auctions, the English auctions will raise average revenue, as the theory predicts. The second prediction is the standard, equilibrium prediction. The first prediction follows directly from our experience with first-price auctions with x_L publicly announced.

Table 4.3 shows averages of predicted and actual changes in revenue between English and first-price auctions for inexperienced bidders, as well as averages of predicted and actual profit, with the results classified by numbers of bidders and ε.[11] Average revenue is predicted to be higher in the English auctions in all cases, for the set of signal values actually drawn, with significantly higher average revenue predicted for all values of ε with $n = 4$ and for $\varepsilon = \$12$ with $n = 7$.[12] However, for these inexperienced bidders, with the exception of $n = 4$ and $\varepsilon = \$24$, actual revenue is lower in the English auctions in all cases, with significantly lower average revenue for $n = 4$ and 7 with $\varepsilon = \$6$, and with the reduction in revenue barely missing statistical significance (at the 10 percent level) with $n = 7$ and $\varepsilon = \$12$. Further, the revenue increase

Table 4.3 Inexperienced Bidders: Actual versus Theoretical Revenue Changes and Profit Levels[a] in English versus First-Price Auctions

	n = 4							*n* = 7						
	Average change in revenue: English less first-price			Average profit				Average change in revenue: English less first-price			Average profit			
				First-price		English					First-price		English	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
ε	Actual	Theo-retical	Differ-ence	Actual	Theo-retical	Actual	Theo-retical	Actual	Theo-retical	Differ-ence	Actual	Theo-retical	Actual	Theo-retical
$6	–1.54**	1.54***	–3.08***	–2.13	2.76	–0.58	1.23	–1.98**	0.10	–2.08**	–3.85	0.99	–1.87	0.89
	(0.72)	(0.49)	(0.71)	(0.52)	(0.38)	(0.50)	(0.30)	(0.87)	(0.34)	(0.78)	(0.71)	(0.19)	(0.51)	(0.29)
					[29]		[28]					[18]		[18]
$12	–0.54	2.76***	–3.30***	–1.32	5.01	–0.78	2.25	–1.95*	1.08	–3.03***	–3.75	2.76	–1.80	1.68
	(1.25)	(0.92)	(0.84)	(0.79)	(0.60)	(0.95)	(0.69)	(1.19)	(0.65)	(0.92)	(0.89)	(0.53)	(0.77)	(0.40)
					[41]		[45]					[30]		[43]
$24	1.09	8.10***	–7.01**	1.20	9.83	0.11	1.73	ND	ND	ND	ND		ND	
	(3.29)	(2.32)	(3.05)	(1.93)	(1.25)	(2.64)	(2.14)							
					[25]		[13]							

NOTE: Standard errors are in parentheses. Bracketed terms are the number of auction periods. ND = no data. *The null hypothesis that the value is greater than or equal to zero can be rejected at the 10% significance level; **the null hypothesis that the value is greater than or equal to zero can be rejected at the 5% level; ***the null hypothesis that the value is greater than or equal to zero can be rejected at the 1% level.

[a] All values reported in dollars.

SOURCE: From Levin, Kagel, and Richard (1996). Reprinted with permission of the American Economic Association.

with $n = 4$ and $\varepsilon = \$24$ is statistically insignificant, and is well below the predicted increase.

These perverse revenue effects in terms of Nash equilibrium bidding theory are associated with negative average profit in both the first-price and English auctions. The negative average profits reported in Table 4.3 indicate that inexperienced bidders suffered from a winner's curse in both auction institutions, but that the curse was relatively stronger in the first-price auctions. These results serve to generalize those reported for first-price sealed-bid auctions with x_L publicly announced: Given a relatively strong winner's curse in sealed-bid auctions, public information reduces rather than raises sellers' average revenue. The two major differences between the present results and the first-price auctions with x_L publicly announced are: 1) here, public information is generated endogenously in the form of drop-out prices, and 2) average profits in the English auctions were negative, but with the exogenous release of public information in the first-price auctions they were positive. This last result suggests that information dissemination in the English auction is noisier than with x_L publicly announced.[13]

For more experienced bidders, English auctions are capable of raising average sellers' revenue, as the data in Table 4.4 demonstrate. With $n = 4$, actual revenue is higher in the English auctions for both values of ε, with a statistically significant increase for $\varepsilon = \$18$. However, for $n = 7$, there is essentially no difference in revenue between the first-price and English auctions. The significant increase in revenue in English auctions with $n = 4$ and $\varepsilon = \$18$ is associated with elimination of the worst effects of the winner's curse in the first price auctions, as bidders earned a substantial share (more than 50 percent) of predicted profit. The importance of eliminating the winner's curse for the revenue-raising prediction of the theory to hold is reinforced by the absence of any revenue increase with $n = 7$, in conjunction with the relatively low share of expected profit (21 percent) that was earned in these first-price auctions.

Levin, Kagel, and Richard (1996) develop an econometric model to characterize how bidders process information in the English auctions. As noted, the Nash bidding model predicts that bidders with higher signal values will average their own signal value with the first drop-out price observed, ignoring all intermediate drop-out prices. What Levin, Kagel, and Richard found, however, is that bidders placed

Table 4.4 Super-Experienced Bidders: Actual versus Theoretical Revenue Changes and Profit Levels[a] in English versus First-Price Auctions

	n = 4							*n* = 7						
	Average change in revenue: English less first-price			Average profit				Average change in revenue: English less first-price			Average profit)			
				First-price		English					First-price		English	
ε	(1) Actual	(2) Theo-retical	(3) Differ-ence	(4) Actual	(5) Theo-retical	(6) Actual	(7) Theo-retical	(8) Actual	(9) Theo-retical	(10) Differ-ence	(11) Actual	(12) Theo-retical	(13) Actual	(14) Theo-retical
$18	2.21**	3.96***	–1.75**	3.37	6.77	1.16	2.82	–0.25	2.85***	–3.10***	0.76	3.86	1.01	1.01
	(0.95)	(0.73)	(0.68)	(0.50)	(0.48)	(0.88)	(0.53)	(0.86)	(0.61)	(0.59)	(0.65)	(0.50)	(0.56)	(0.37)
			[163]		[107]						[75]		[96]	
	1.20	2.98	–1.78	8.45	11.27	7.25	8.29							
$30	(3.10)	(2.30)	(2.19)	(1.28)	(1.34)	(2.76)	(l.93)		ND			ND		ND
				[31]		[33]								

NOTE: Standard errors are in parentheses. Bracketed terms are the number of auction periods. ND = no data. **The null hypothesis that the value is greater than or equal to zero can be rejected at the 5% significance level; ***the null hypothesis that the value is greater than or equal to zero can be rejected at the 1% significance level.

[a] All values reported in dollars.

SOURCE: From Levin, Kagel, and Richard (1996). Reprinted with permission of the American Economic Association.

weight on their own signal value and the immediate past drop-out price, ostensibly ignoring x_L and any earlier drop-out prices. Further, as more bidders dropped out, subjects placed less and less weight on their own signal value, and more weight on the last drop-out price. This pattern, although inconsistent with the Nash model, is consistent with bidders acting as if they were averaging their own signal value with the signal values underlying the drop-out prices of *all* earlier bidders. Levin, Kagel, and Richard attribute the adoption of this signal averaging rule in favor of the Nash rule to the fact that it is easy and quite natural to use, and that it yields results similar to the Nash rule without requiring that bidders explicitly recognize the adverse selection effect of winning the auction and/or knowing anything about sufficient statistics.

Auctions with Insider Information

Kagel and Levin (1999) investigate bidding in first-price sealed-bid auctions with an asymmetric information structure (AIS). The asymmetry is introduced by choosing one bidder at random in each auction period—the insider (I)—to receive a private information signal x equal to x_o and being told that $x = x_o$. Each of the other bidders, the outsiders (Os), receive a private information signal from a uniform distribution on $[x_o - \varepsilon, x_o + \varepsilon]$, as in the auctions with a symmetric information structure (SIS). The insider does *not* know the realizations of Os' private information signals. Os know that they are Os, that there is a single I who knows x_o, and the way that all other Os got their private signals.

Note that this information structure differs substantially from the "standard" insider information model employed in the economics literature in which the insider has a double informational advantage—I knows x_o and Os only have access to public information about x_o (Engelbrecht-Wiggans, Milgrom, and Weber 1983; Hendricks and Porter 1988). In contrast, in our design Os have some proprietary information, which permits them to earn positive expected profit in equilibrium. In the double informational advantage model, Os earn zero expected profit in equilibrium.

This experimental design has a number of interesting comparative static predictions that contrast sharply with the double informational

advantage model. First and foremost, the existence of an insider benefits the seller by increasing expected revenue relative to auctions with an SIS. In contrast, in the double informational advantage model, the existence of an insider unambiguously reduces sellers' expected revenue.[14] Second, increases in the number of Os results in Is bidding more aggressively in our model. In contrast, in the double informational advantage model, I's bidding strategy is unaffected by increases in the number of Os. Finally, both models imply that Is earn substantially larger expected profit than Os (zero profit for Os in the double informational advantage model), and that Is earn higher expected profit, conditional on winning, than in SIS auctions, although the predicted increase in profit is relatively small in our design.

Kagel and Levin (1999) conjecture that for inexperienced bidders the existence of an insider might attenuate the winner's curse. Os in the AIS auctions who win against better informed Is face a stronger adverse selection effect than in SIS auctions. However, it is entirely plausible that the need to hedge against the existence of an insider is more intuitive and transparent than the adverse selection problem resulting from winning against symmetrically informed rivals. Thus, at least for inexperienced bidders, having an insider may actually reduce the severity of the winner's curse. This would be true, for example, if Os view the situation as similar to a lemon's market (Akerlof 1970), where it seems reasonably clear there is no rampant winner's curse (our culture warns us to beware of used car salesmen). On the other hand, inexperienced subjects may bid higher in order to make up for their informational disadvantage, thus exacerbating the winner's curse.

Kagel and Levin employ two alternative definitions of the winner's curse for Os in the AIS auctions. The first, very conservative definition concerns bidding above the expected value conditional on having the highest signal value among Os (ignoring I's bid). If all Os bid this way, and Is best respond to these bids, then Os would earn average *losses* of more than $1.50 per auction, conditional on winning. The second definition accounts for Is best responding to Os' bids, and solves for the zero expected profit level for Os. Table 4.5 reports results for inexperienced bidders in these auctions. The data clearly indicate that the winner's curse is alive and well for inexperienced Os. Consider auctions with $\varepsilon = \$6$, which were used to start each session. With $n = 4$, almost 60 percent of the high Os' bids were above the conservative measure of

Table 4.5 Inexperienced Bidders: Auctions with Asymmetric Information Structure (AIS)

		Outsiders' bids								Insiders' bids	
				Frequency of winner's curse (raw data)							
				Against outsiders only		Against outsiders and insiders					
Number of bidders	ε	Average earnings conditional on winning (S_m)	Frequency of outsiders winning (%) (raw data)	High outsider bid (%)	All bids (%)	High outsider bid (%)	All bids (%)	Average bid factor[a] (S_m)	Frequency high outsider bid from high outsider signal holder (%) (raw data)	Average earnings conditional on winning (S_m)	Average bid factor (S_m)
		–1.68	70.6	58.8	39.2	94.1	70.6	1.16	52.9	0.71	1.46[b]
4	6	(0.93)	(12/17)	(10/17)	(20/51)	(16/17)	(36/51)	(0.62)	(9/17)	(0.35)	(0.26)
		–1.40	65.2	39.1	23.2	65.2	47.8	6.00	73.9	2.74	2.25
	12	(0.50)**	(15/23)	(9/23)	(16/69)	(15/23)	(33/69)	(0.77)	(17/23)	(0.77)**	(0.35)
		–6.56	71.4	28.6	14.3	85.7	57.1	11.61	100	5.05	5.09
	24	(3.07)	(5/7)	(2/7)	(3/21)	(6/7)	(12/21)	(2.78)	(7/7)	(3.50)	(1.27)
		–3.68	100	100	85.2	100	92.6	–0.61[c]	66.7	—	1.09[b]
7	6	(0.61)***	(9/9)	(9/9)	(46/54)	(9/9)	(50/54)	(0.62)	(6/9)		(0.29)
		–2.47	78.9	89.5	69.7	89.5	79.8	4.85	73.7	1.93	1.91[b]
	12	(1.03)**	(15/19)	(17/19)	(78/112)	(17/19)	(91/114)	(1.03)	(14/19)	(0.61)***	(0.33)

NOTE: S_m = standard error of the mean. ** Significantly different from 0 at the 5% level, two-tailed test; *** significantly different from 0 at the 1% level, two-tailed test.

[a] High bids only.

[b] A single outlier bid less than $x_o - \varepsilon$ was dropped.

[c] In this treatment, high Os actually bid *above* their signal values, on average.

SOURCE: From Kagel and Levin (1999). Reprinted with permission of the Econometric Society.

the winner's curse, so that these bids would have lost money, on average, just competing against other Os. Further, considering the behavior of both Is and Os (the second winner's curse measure), 94 percent of the high O bids were subject to the winner's curse. With $n = 7$, there is an even stronger adverse selection effect, with the result that the winner's curse was more pervasive: 100 percent of the high O bids and 85.2 percent of *all* O bids fell prey to the winner's curse, even with no accounting for I's bids. The net result, in both cases, was large negative profits for Os when they won (–$1.68 per auction with $n = 4$; –$3.68 with $n = 7$). Although somewhat diminished in frequency, a strong winner's curse is also reported for higher values of ε as Os continued to earn negative profits throughout, with at least 47 percent of all bids subject to the winner's curse for any value of ε (when accounting for both Is' and Os' bids). Finally, regressions comparing bid functions for inexperienced Os in AIS auctions versus inexperienced bidders in SIS auctions show no significant difference between the two treatments. Thus, contrary to Kagel and Levin's original conjecture, the introduction of an insider did not induce significantly less aggressive bidding for inexperienced Os compared to SIS auctions.

Table 4.6 reports data for super-experienced bidders (subjects who had participated in at least two prior first-price sealed-bid auction sessions). For these bidders the winner's curse has been largely eliminated and the comparative static predictions of the theory are generally satisfied. Is earned significantly greater profits conditional on winning than did Os. For example, with $e = \$18$ and $n = 7$, Os earned average profits of around $0.50 per auction conditional on winning. In contrast, Is earned around $3.25 per auction, conditional on winning. Further, Os earned substantially lower profits than in corresponding SIS auctions, for which profits averaged around $2.25 per auction. Also, as the theory predicts, Is increased their bids in the face of greater competition from more Os.

Last, but not least, as the theory predicts, for more experienced bidders, auctions with insider information consistently raised average sellers' revenue compared to SIS auctions (Table 4.7). The intuition underlying this prediction for our model is as follows: The seller would be unambiguously worse off in the AIS auction relative to the SIS auction if Is in the AIS auction won *all* the time while bidding according to the prescribed (AIS) equilibrium. However, Is do not win all the time,

Table 4.6 Super-Experienced Bidders: Auctions with Asymmetric Information Structure (AIS)

		Outsiders' bids						Insiders' bids	
				Frequency of winner's curse: Against outsiders and insiders (raw data)					
Number of bidders	ε	Average earnings conditional on winning (S_m)	Frequency of outsiders winning (%) (raw data)	High outsider bid (%)	All bids (%)	Average bid factor[a] (S_m)	Frequency high outsider bid from high outsider signal holder (%) (raw data)	Average earnings conditional on winning (S_m)	Average bid factor (S_m)
4	12	0.65 (0.43)	53.7 (29/54)	9.3 (5/54)	4.9 (8/162)	10.05 (0.23)	92.6 (50/54)	3.30 (0.23)***	3.60[c] (0.19)
	18	0.87 (0.68)	63.3 (19/30)	3.3 (1/30)	1.1 (1/90)	15.29 (0.26)	93.3 (28/30)	4.13 (0.37)***	5.80[c] (0.50)
	30	3.67 (2.32)	42.1 (8/19)	5.3 (1/19)	3.5 (2/57)	27.04 (0.65)	94.7 (18/19)	7.94 (0.69)***	8.24 (0.61)
7[b]	18	0.52 (0.34)	64.5 (49/76)	22.4 (17/76)	17.2 (77/453)	15.86 (0.26)	86.8 (66/76)	3.24 (0.36)***	4.35 (0.26)
	30	3.90 (3.07)	41.7 (5/12)	16.7 (2/12)	19.4 (14/72)	26.95 (0.85)	83.3 (10/12)	4.95 (0.80)***	5.98 (0.67)

NOTE: S_m = standard error of the mean. ***Significantly different from 0 at the 1% level, two-tailed *t*-test.

[a] High bids only.

[b] Includes several auctions with $n = 6$.

[c] A single outlier bid less than $x_o - \varepsilon$ was dropped.

SOURCE: From Kagel and Levin (1999). Reprinted with permission of the Econometric Society.

Table 4.7 Change in Seller's Revenue: AIS versus SIS Auctions with Super-Experienced Bidders

	$n = 4$			$n = 7$		
	Change in revenue: AIS less SIS auctions	Mean profits (Σ^2)		Change in revenue: AIS – SIS auctions	Mean profits (Σ^2)	
		AIS	SIS		AIS	SIS
	1.759	2.063	3.822	0.739	1.492	2.231
ε = \$18	(2.057)**	(8.561)	(49.972)	(1.573)*	(6.770)	(19.221)
	2.734	6.148	8.876	0.919	4.517	5.436
ε = \$30	(1.097)	(24.334)	(59.731)	(0.425)	(17.978)	(15.839)

NOTE: *t*-statistics are in parentheses. *Significantly different from 0 at $p < 0.10$, one-tailed test; **significantly different from 0 at $p < 0.05$, one-tailed test.

SOURCE: From Kagel and Levin (1999). Reprinted with permission of the Econometric Society.

and when Os win (with their equilibrium bid), they win with relatively high signal values, yielding more revenue than when Is win. Further, the existence of the insider helps to "protect" the seller's revenue compared to an SIS auction when Os would have won with relatively low signal values in the SIS auction, since in this case I wins and pays more than O would have paid in the SIS auction. The net result is higher revenue for the seller and reduced variance in seller's revenue (holding x_o constant) compared to SIS auctions.[15]

The increase in revenue resulting from an insider in our model is counterintuitive for those whose intuition has been honed on the double informational advantage model. This reversal of the double informational advantage model's prediction rests critically on the fact that less informed bidders have some proprietary information. Many "real world" cases are more realistically modeled with Os having some proprietary information and not just public information. In these circumstances, it may well be the case that the introduction of a single well-informed insider increases average sellers' revenue, and that both Is and Os earn economic rents. This potential for insider information to raise average sellers' revenue had not been explicitly recognized in the auction literature prior to this.[16]

CONCLUSION

Summary and Policy Implications

Experimental studies of common value auctions have been going on for more than 15 years now, paralleling the profession's interest in the theoretical and practical properties of these auctions. This research has established several facts about behavior relative to the theory.

For inexperienced bidders, Nash equilibrium bidding theory does not predict well. Inexperienced bidders suffer from a winner's curse, earning negative average profits and with relatively large numbers of bidders going bankrupt. Overbidding here represents a fundamental breakdown in the theory, resulting in the reversal of a number of important comparative static predictions: Bidding does not decrease in response to increased numbers of bidders in second-price auctions as the theory predicts, and public information about the value of the item reduces, rather than raises, revenue in the presence of a winner's curse. This perverse effect of public information in the presence of a winner's curse extends to the endogenous release of public information in English clock auctions.

Experienced bidders in the lab eventually overcome the worst effects of the winner's curse, rarely bidding above the expected value of the item conditional on winning and earning positive average profits. Super-experienced bidders also satisfy key comparative static predictions of the theory: Release of public information in sealed-bid auctions raises revenue, and English clock auctions raise more revenue than do sealed-bid auctions. Further, average revenue increases in an experimental design where the existence of an informed insider is predicted to raise revenue compared to auctions with symmetrically informed bidders. Nevertheless, these super-experienced bidders still earn well below equilibrium profits and, in the overwhelming majority of cases, are not best responding to rivals' bids (they are bidding far more aggressively than they should; Kagel and Richard 2001).

It is worth noting that these very experienced bidders in the lab have learned how to overcome the worst effects of the winner's curse in an environment with strong information feedback, substantially stronger than is likely to be present in field settings. As such, learning

might not proceed as quickly in field settings. Further, there are dynamics of interactions within organizations that may retard adjustment to the winner's curse. These include payments of large salaries to petroleum geologists to estimate likely reserves, and then having to recognize that these estimates still have a very large variance and are not very precise; transfers of personnel within the firm and between firms prior to receiving feedback about the profitability of bids; and gaming that goes on within organizations.[17] Finally, even assuming that the winner's curse will be eliminated in the long run in field settings, it often takes some time before this happens, so this out-of-equilibrium behavior is important in its own right.

The winner's curse extends to a number of other settings as well: bilateral bargaining games (Samuelson and Bazerman 1985; Ball, Bazerman, and Carroll 1991), blind-bid auctions (Forsythe, Isaac, and Palfrey 1989), markets where quality is endogenously determined (Lynch et al. 1986, 1991), and voting behavior (the swing voters curse; Feddersen and Pesendorfer 1998, 1999).[18]

Experimental studies of auction markets have played a significant role in the design and execution of the recent wave of spectrum (air wave rights) auctions carried out in this country and abroad.[19] Auction experiments have served two principle functions in this work: 1) as a "wind tunnel" to test out the auction software, which implements a relatively complicated set of bidding rules (see, for example, Plott 1997), and 2) as a test bed against which to compare theory with behavior. In the latter role, a central design element has been to use ascending-price auctions (with price feedback for bidders) to both minimize the presence of the winner's curse and to generate increased revenue in the absence of a winner's curse, central insights derived from the interaction between common value auction theory and experiments:

> An ascending auction ought to remove another common problem with auctions, the "winner's curse." This strikes when a successful bidder discovers too late that his prize is not worth what he paid for it. Some critics of the scale of the bids seem to see the curse at work [in Britain's third generation sales]. Yet the winner's curse is much likelier in sealed-bid auctions, where bidders lack an important piece of information about the value of the asset: the valuations of other, perhaps better-informed, bidders. In an

> ascending auction, however, that information is clearly revealed. (*The Economist* 2000, p. 21)
>
> . . . by allowing bidders to respond to each other bids, [an ascending-price auction] diminishes the winner's curse: that is, the tendency for naive bidders to bid up the price beyond the licenses's actual value, or for shrewd bidders to bid cautiously to avoid over paying. (McAfee and McMillan 1996, p. 161)

Notes

Research support from the Economics and DRMS Divisions of NSF, the Sloan Foundation, and the Russell Sage Foundation are gratefully acknowledged. Special thanks to my colleagues and my coauthors, especially Dan Levin, who have taught me so much. Much of the material here is taken from my paper with Dan Levin titled "Bidding in Common Value Actions: A Survey of Experimental Research," which appears as Chapter 1 in the collection of our published papers investigating common value auctions: John H. Kagel and Dan Levin, *Common Value Auctions and the Winner's Curse*, Princeton University Press.

1. Here, I am assuming that buyers are competing to purchase an item. Similar remarks hold for procurement auctions in which sellers compete to offer services at the lowest cost. In this case, however, the trade-off is inverted; the higher their bids, the larger the surplus conditional on winning, but the lower the probability of winning.
2. However, once the seller uses a minimum bid requirement, and/or we consider entry to be determined endogenously, different auctions may induce different probabilities of an actual sale. Thus, efficiency may become an issue (Levin and Smith 1994).
3. Unless, of course, one argues that the Groucho Marx statement "I do not wish to join any club that accepts me," is an earlier recognition of the winner's curse.
4. See, for example, the exchange between Cox and Isaac (1984, 1986) and Brown (1986).
5. Winning bidders paid these losses out of their own pockets or from earnings in other auctions.
6. There is a whole body of psychological literature indicating the difficulty of learning generalizing across different contexts (see, for example, Gick and Holyoak 1980; Perkins and Salomon 1988; Salomon and Perkins 1989).
7. Kagel and Levin (1986) did not restrict low signal holders to bidding x_L, failing to recognize that without this restriction there is no pure strategy Nash equilibrium, but a much more complicated mixed strategy equilibrium so that their benchmark calculations are incorrect. However, the correct benchmark yields an even higher increase in revenue from announcing x_L so that the conclusions reached regarding

public information receive even stronger support with the correct benchmark (Campbell, Kagel, and Levin 1999).

8. The greater the number of rivals, the lower the probability of winning as a result of more aggressive bidding; hence, the less likely it is to pay to deviate from the Nash strategy even with limited liability. See also the calculations reported in Kagel and Richard (2001).
9. For a completely different approach to the limited liability problem, see Avery and Kagel (1997).
10. The intuition is roughly as follows: Given symmetry, the low signal holder knows that those remaining in the auction have higher signal values. But the low signal holder can't profit from this additional information since it is only revealed once the price is greater than these remaining signal values; i.e., price is already greater than the expected value of the item to the low signal holder. The analysis is confined in the interval $\underline{x} + \varepsilon \leq x \leq \overline{x} - \varepsilon$.
11. Common value auctions involve pure surplus transfers so that revenue differences are calculated as: $[\pi_E - \pi_F]$ where π_E and π_F correspond to profits in English and first-price auctions, respectively. In this way we have effectively normalized for sampling variability in x_o by subtracting it from the price.
12. *t*-tests are conducted for predicted revenue increases to measure the reliability of the prediction for Levin, Kagel, and Richard sample data. One-tailed *t*-tests are used here since the symmetric RNNE makes unambiguous predictions regarding revenue increases. Two-tailed *t*-tests are used for determining statistical significance of actual revenue changes, since in practice there are force promoting lower revenues in English auctions and we often observe this outcome.
13. To further investigate this question, we have conducted some additional sessions with inexperienced bidders in which x_L was publicly announced prior to bidding in the English auction. In auctions with six bidders and ε = \$12, average profits in the standard English auction (where x_L was not announced) were –\$1.55, with average profits in auctions with x_L announced of \$1.56 (t = 1.46, d.f. = 30, $p <$ 0.10, one-tailed test; Kagel and Levin 2002).
14. Although one can readily demonstrate that increased revenue is *not* a general characteristic of AIS auctions in which Os have some proprietary information, it is a natural element in our design and can be found in other AIS structures as well (Campbell and Levin 2000).
15. In our design, the increase in revenue going from SIS to AIS varies with *n*, with revenue differences increasing starting from low *n*, reaching a maximum revenue differential for intermediate levels of *n*, and decreasing thereafter.
16. These results motivated Campbell and Levin (2000) to further investigate the role of insider information in first-price auctions compared to homogeneous information environments. This chapter connects the revenue raising effects of an insider to more general propositions regarding the revenue raising effects of increased bidder information found in Milgrom and Weber (1982).
17. A friend of mine in Houston who was a geologist for a major oil company told me that there was such a broad range of legitimate value estimates for most tracts that

when the bidding department started reducing bids relative to value estimates to the point that they were winning very few auctions, the geologists simply raised their estimates. (Geologists love to drill, and failure to win tracts means they can't drill.)

18. See Kagel and Levin (2002) for reviews of this work, or better yet, consult the original publications.
19. Led by the Federal Communications Commission, the U.S. government has conducted a number of sales to date raising a total of $23.9 billion and selling over 10,000 licenses between July 1994 and July 2000. Even more spectacular, in an auction ending in April 2000, the British government raised £22.5 billion ($35.53 billion) from the sale of "third generation" mobile phone licenses. See Klemperer (2000) and McAfee and McMillan (1996) for reviews and evaluations of these auctions.

References

Akerlof, G. 1970. "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84(3): 488–500.

Ashenfelter, O., and D. Genesore. 1992. "Testing for Price Anomalies in Real Estate Auctions." *American Economic Review: Papers and Proceedings* 82(2): 501–505.

Avery, C., and J.H. Kagel. 1997. "Second-Price Auctions with Asymmetric Payoffs: An Experimental Investigation." *Journal of Economics and Management Strategy* 6(6): 573–604.

Ball, S.B., M.H. Bazerman, and J.S. Carroll. 1991. "An Evaluation of Learning in the Bilateral Winner's Curse." *Organizational Behavior and Human Decision Processes* 48(7): 1–22.

Bazerman, M.H., and W.F. Samuelson. 1983. "I Won the Auction But Don't Want the Prize." *Journal of Conflict Resolution* 27(4): 618–634.

Blecherman, B., and C.F. Camerer. 1998. "Is There a Winner's Curse in the Market for Baseball Players?" Photocopy, Brooklyn Polytechnic University, Brooklyn, NY.

Brown, K.C. 1986. "In Search of the Winner's Curse: Comment." *Economic Inquiry* 24(5): 513–515.

Capen, E.C., R.V. Clapp, and W.M. Campbell. 1971. "Competitive Bidding in High-Risk Situations." *Journal of Petroleum Technology* 23(6): 641–653.

Campbell, C., and D. Levin. 2000. "Can the Seller Benefit from an Insider in Common Value Auctions?" *Journal of Economic Theory* 91(1): 106–120.

Campbell, C., J.H. Kagel, and D. Levin. 1999. "The Winner's Curse and Public Information in Common Value Auctions: Reply." *American Economic Review* 89(1): 325–334.

Cassady, Ralph. 1967. *Auctions and Auctioneering*. Berkeley, CA: University of California Press.

Cassing, J., and R.W. Douglas. 1980. "Implications of the Auction Mechanism in Baseballs' Free Agent Draft." *Southern Economic Journal* 47(1): 110–121.

Cox, J.C., and R.M. Isaac. 1984. "In Search of the Winner's Curse." *Economic Inquiry* 22(4): 579–592.

———. 1986. "In Search of the Winner's Curse: Reply." *Economic Inquiry* 24(3): 517–520.

Cox, J.C., S.H. Dinkin, and V.L. Smith. 1998. "Endogenous Entry and Exit in Common Value Auctions." Photocopy, University of Arizona, Tucson, AZ.

Dessauer, J.P. 1981. *Book Publishing*. New York: Bowker.

Dyer, D., J.H. Kagel, and D. Levin. 1989. "A Comparison of Naive and Experienced Bidders in Common Value Offer Auctions: A Laboratory Analysis." *Economic Journal* 99(394): 108–115.

Engelbrecht-Wiggans, R., P.R. Milgrom, and R.J. Weber. 1983. "Competitive Bidding and Proprietary Information." *Journal of Mathematical Economics* 11(2): 161–169.

Eyster, E., and M. Rabin. 2000. "Cursed Equilibrium." Photocopy, University of California at Berkeley, Berkeley, CA.

The Economist. 2000. "Comments." 356(8181): 21.

Feddersen, T., and W. Pesendorfer. 1998. "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting." *American Political Science Review* 92(7): 23–36.

———. 1999. "Elections, Information Aggregation, and Strategic Voting." *Proceedings of the National Academy of Science* 96(19): 10572–10574.

Forsythe, R., R.M. Isaac, and T.R. Palfrey. 1989. "Theories and Tests of 'Blind Bidding' in Sealed-Bid Auctions." *Rand Journal of Economics* 20(2): 214–238.

Gick, M.L., and K.J. Holyoak. 1980. "Analogical Problem Solving." *Cognitive Psychology* 12(3): 306–355.

Goeree, J.K., and T. Offerman. 2000. "Efficiency in Auctions with Private and Common Values: An Experimental Study." Photocopy, University of Virginia, Charlottesville, VA.

Hansen, R.G., and J.R. Lott, Jr. 1991. "The Winner's Curse and Public Information in Common Value Auctions: Comment." *American Economic Review* 81(1): 347–361.

Hendricks, Kenneth, and Robert H. Porter. 1988. "An Empirical Study of an Auction with Asymmetric Information." *American Economic Review* 76: 865–833.

Hendricks, K., R.H. Porter, and B. Boudreau. 1987. "Information, Returns, and Bidding Behavior in OCS Auctions: 1954–1969." *The Journal of Industrial Economics* 35(4): 517–542.

Hendricks, K., R.H. Porter, and C.A. Wilson. 1994. "Auctions for Oil and Gas Leases with an Informed Bidder and a Random Reservation Price." *Econometrica* 62(6): 1415–1444.

Holt, C.A. Jr., and R. Sherman. 1994. "The Loser's Curse." *American Economic Review* 84(3): 642–652.

Kagel, J.H., and D. Levin. 1986. "The Winner's Curse and Public Information in Common Value Auctions." *American Economic Review* 76(5): 894–920.

———. 1991. "The Winner's Curse and Public Information in Common Value Auctions: Reply." *American Economic Review* 81(1): 362–369.

———. 1999. "Common Value Auctions with Insider Information." *Econometrica* 67(5): 1219–1238.

———. 2002. "Bidding in Common Value Auctions: A Survey of Experimental Research." In *Common Value Auctions and the Winner's Curse*, J.H. Kagel and D. Levin, eds. Princeton: Princeton University Press, pp. 1–84.

Kagel, J.H., D. Levin, and R.M. Harstad. 1995. "Comparative Static Effects of Numbers of Bidders and Public Information on Behavior in Second-Price Common-Value Auctions." *International Journal of Game Theory* 24: 293–319.

Kagel, J.H., D. Levin, R. Battalio, and D.J. Meyer. 1989. "First-Price Common Value Auctions: Bidder Behavior and the Winner's Curse." *Economic Inquiry* 27(2): 241–258.

Kagel, J.H., and J.F. Richard. 2001. "Super-Experienced Bidders in First-Price Common Value Auctions: Rules of Thumb, Nash Equilibrium Bidding and the Winner's Curse." *Review of Economics and Statistics* 83(3): 408–419.

Klemperer, P. 2000. "What Really Matters in Auction Design." Photocopy, Nuffield College, Oxford University, Oxford, U.K.

Levin, D., J.H. Kagel, and J.F. Richard. 1996. "Revenue Effects and Information Processing in English Common Value Auctions." *American Economic Review* 86(3): 442–460.

Levin, D., and J.L. Smith. 1994. "Equilibrium in Auctions with Entry." *American Economic Review* 84(3): 585–599.

Lind, B., and C.R. Plott. 1991. "The Winner's Curse: Experiments with Buyers and with Sellers." *American Economic Review* 81(1): 335–346.

Lorenz, J., and E.L. Dougherty. 1983. "Bonus Bidding and Bottom Lines: Federal Off-Shore Oil and Gas." SPE 12024, 58th Annual Fall Technical Conference, San Francisco, October 5–8.

Lynch, M., R.M. Miller, C.R. Plott, and R. Porter. 1986. "Product Quality, Consumer Information and 'Lemons' in Experimental Markets." In *Empir-*

ical Approaches to Consumer Protection Economics, P.M. Ippolito and D.T. Scheffman, eds. Washington, DC: FTC Bureau of Economics, pp. 251–306.

———. 1991. "Product Quality, Informational Efficiency, and Regulations in Experimental Markets." In *Research in Experimental Economics*, R. Mark Isaac, ed. Greenwich, CT: JAI Press, p. 4.

McAfee, R.P., and J. McMillan. 1996. "Analyzing the Airwaves Auction." *Journal of Economic Perspectives* 10(1): 159–176.

Milgrom, P., and R.J. Weber. 1982. "A Theory of Auctions and Competitive Bidding." *Econometrica* 50(5): 1089–1122.

Perkins, D.N., and G. Salomon. 1988. "Teaching for Transfer." *Educational Leadership* 46(1): 22–32.

Plott, C. 1997. "Laboratory Experimental Test Beds: Application to the PCS Auction." *Journal of Economics and Management Strategy* 6(3): 605–638.

Roll, R. 1986. "The Hubris Hypothesis of Corporate Takeovers." *Journal of Business* 59(2, Part 1): 197–216.

Salomon, G., and D.N. Perkins. 1989. "Rocky Roads to Transfer: Rethinking Mechanisms of a Neglected Phenomenon." *Education Psychologist* 24(2): 113–142.

Samuelson, W.F., and M.H. Bazerman. 1985. "The Winner's Curse in Bilateral Negotiations." In *Research in Experimental Economics*, Vol. 3, V.L. Smith, ed. Greenwich, CT: JAI Press, pp. 105–138.

Vickrey, W. 1961. "Counterspeculation, Auctions, and Competitive Sealed Tenders." *Journal of Finance* 16(1): 8–37.

Weverbergh, M. 1979. "Competitive Bidding with Asymmetric Information Reanalyzed." *Management Science* 25: 291–294.

Wilson, R. 1967. "Competitive Bidding with Asymmetric Information." *Management Science* 13(11): 816–820.

———. 1992. "Strategic Analysis of Auctions." In *Handbook of Game Theory with Economic Applications*, Vol. 1, R.J. Aumann and S. Hart, eds. Amsterdam: Elsevier Science Publishers.

5
Sharing Very High Risks

How Government Can Make Health Insurance Markets More Efficient and More Accessible

Katherine Swartz
Harvard School of Public Health

Between 40 and 44 million Americans—one in six nonelderly—do not have any form of health insurance, according to the 2001 Census. Why they do not have health insurance involves a variety of reasons, many of which are often present in any particular person who lacks coverage. We can make two generalizations, however. First, a majority simply cannot afford to purchase health insurance unless it is heavily subsidized, which currently means subsidized by an employer that sponsors group coverage. About two-thirds of the uninsured have family incomes below $35,000, which is generally too low to be able to afford health insurance unless an employer pays a large share of the group premium. The second generalization is that health insurance markets, especially the small group and individual (nongroup) markets, are subject to market failure. The market failure is caused by insurers' fear of adverse selection. Carriers know from experience that people who know or suspect they will have expensive health care needs in the coming year are also more likely to apply for insurance coverage than people who do not expect such expenses. Such people make up a disproportionate fraction of the people who apply for coverage every year. As a result, insurers are especially likely to either refuse to insure an applicant or set a high premium for anyone who they perceive to be likely to incur higher medical expenditures. People who fall into this category are generally over the age of 45, female, working in particular types of occupations, and have had medical problems in the past. For these people, health insurance is also either unaffordable—given the high premiums relative to their incomes—or simply unavailable.

Both of these explanations of why people lack health insurance provide rationales for government taking a role in health insurance. Medicaid and the State Children's Health Insurance Program (SCHIP) were created largely to deal with the fact that low-income people cannot afford to purchase health insurance at existing prices. Current efforts to incrementally expand eligibility for Medicaid and new, subsidized buy-in programs are similarly grounded in the tradition that we use government to redistribute resources in our society to make sure that low-income or otherwise deserving people receive goods and services deemed necessities.

When markets break down in the absence of full information (as with adverse selection), economic theory argues for government to intervene to counter the problem with the objective of making the market competitive and thereby efficient.[1] In this chapter, I develop the idea that in the case of health insurance markets, government intervention in the form of being responsible for the very highest-cost individuals every year would reduce insurers' fear of adverse selection. In turn, this would reduce inefficiency caused by insurers spending enormous effort to predict whether or not an individual will be likely to have high medical costs, and premiums ought to be lower as a result. In addition, if insurers do not need to bear the risks of very high-cost people because such risks have been shifted to government—and society at large—then accessibility to health insurance should be greater.

The plan of the chapter is as follows. In the next section, I briefly describe who lacks health insurance in the United States. In the third section, I describe how health insurance markets work and how insurers compete in the individual, nongroup market. In the fourth section, I discuss the proposal to have government shift the risk of very high-cost people from insurers to the general population and how it could increase efficiency and accessibility in individual and small group insurance markets. I also provide some examples of government taking the role of reinsurer and "backstopper" of markets so that they function. Finally, I offer some concluding comments.

WHO DOES NOT HAVE HEALTH INSURANCE?

The uninsured are a cross section of Americans—children, young adults, and middle-aged people who generally work full time but do not earn more than $30,000 per year (in part because they have no more than a high school diploma and do not have specific skills). Because they have low incomes and no health insurance, they frequently cannot afford their share of health insurance premiums when an employer does sponsor coverage and have debts for emergency medical care that they are working to pay down. Some of the adults are widowed or divorced, with young children, so the income they earn does not enable them to pay for nongroup health insurance. Many uninsured adults are self-employed or working in small, family-run businesses that cannot afford to sponsor health insurance. About 9.2 million of the uninsured are children, and perhaps as many as 3 million of these children are eligible for Medicaid or the SCHIPs. However, parents either do not realize their children are eligible for the programs or they find the process of applying for public coverage "unpleasant" (Kaiser Commission on Medicaid and the Uninsured 2000a). A majority of uninsured are white, but African Americans and Hispanics comprise a disproportionate share of the uninsured.

The Henry J. Kaiser Family Foundation's Commission on Medicaid and the Uninsured has conducted lengthy interviews with seven families and one 52-year-old grandmother (Kaiser Commission on Medicaid and the Uninsured 2000b). Two common threads run throughout their stories. One, the adults work hard but do not earn high incomes, so even when they have the option of obtaining health insurance through an employer, they feel that they cannot afford the employee share of the premium. Second, all of the uninsured families have incurred medical debts as a result of being uninsured. The debts are for very treatable medical problems that would not cause an insured person to think twice about seeing a physician or going to the emergency room with a sick child. But the uninsured bills for such care—running between $1,000 and $6,000—leave the uninsured families both strapped for cash to pay for health insurance and in daily fear of further medical bills.

When we examine demographic and socioeconomic characteristics of the uninsured, the multidimensional stories of real people are often overshadowed. Nonetheless, knowing more about the distributions of characteristics of the uninsured helps when developing public policies to increase access to health insurance. I will draw upon the March 1999 Current Population Survey (CPS) for most of what follows. The March 1999 CPS showed that there were almost 44 million nonelderly Americans without any form of health insurance. According to the March 2001 CPS, the number of uninsured declined to about 38.4 million, largely as a result of the booming economy and small increases in the number of people with employer-sponsored coverage. However, the mild recession in 2001 through early 2002, combined with the increase in unemployment, has most analysts believing that the number of uninsured in early 2003 will be closer to the number in 1999, so I will use 1999 data.

Age

The uninsured are generally young—64 percent are younger than 35—making them relatively inexpensive in terms of expected medical care use (Table 5.1). A quarter of the uninsured are children under the age of 18. The 11 million uninsured children account for 15.4 percent of all children. Two decades ago, about a third of the uninsured were children, and close to 20 percent of all children were uninsured, so the decline in the number of uninsured children is a reflection of the impact of the expanded Medicaid eligibility criteria for children. Young adults (18–24) and adults between 25 and 34 have much higher chances of being uninsured—30 percent of young adults and 24 percent of 25–34-year-olds lack coverage.

Income

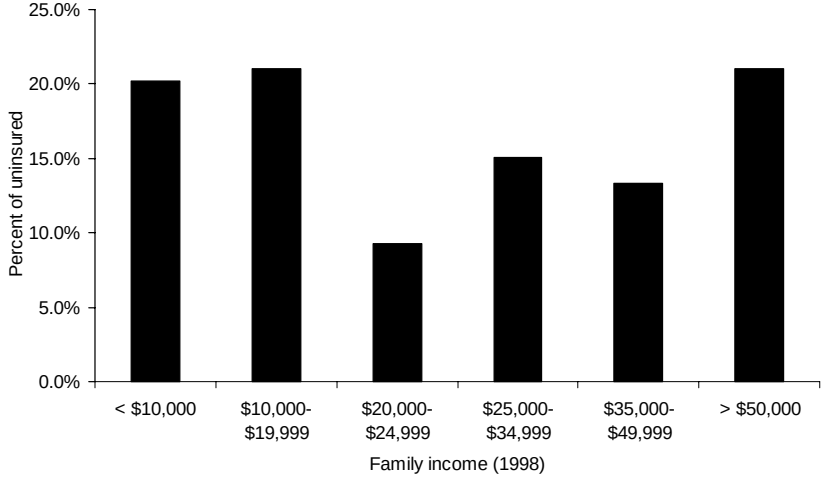
Just over half of the uninsured in 1999 had family incomes in the previous year of under $25,000 (Figure 5.1). (For comparison, in 1999 the median household income for all Americans was $42,100.) Another 15 percent had family incomes between $25,000 and $35,000. Thus, two-thirds of the uninsured in 1999 had incomes below $35,000. Another way of looking at family income is to adjust it for family size

Table 5.1 Uninsured by Age Cohort, 1999

Age cohort	Number	% of uninsured	% of age cohort
< 18	11.073	25.0	15.4
18–24	7.776	17.6	30.0
25–34	9.127	20.6	23.7
35–44	7.708	17.4	17.2
45–64	8.239	18.6	14.2
65 +	0.358	0.8	1.1
Total	44.281	100.0	16.3

SOURCE: March 1999 CPS.

Figure 5.1 Income Distribution of Nonelderly Uninsured, 1999



SOURCE: March 1999 CPS.

and compute it relative to the poverty level by family size. In data not shown here, two-thirds of the uninsured had incomes below 250 percent of the poverty level. These incomes are simply too low for people to afford to purchase health insurance unless it is heavily subsidized by an employer that sponsors group coverage. The 21 percent of the uninsured who have family incomes above $50,000 reflects two changes in the uninsured over the 1990s. One is the growing economy and tight labor market by the end of the 1990s. This caused many people with part-time or part-year jobs (that do not include health insurance as part of the compensation) to work more hours per week and/or more weeks during the year, enabling them to earn incomes above $50,000. This was especially true in two-earner families where each adult might have earned less than $20,000 in weaker economic times. A second factor that explains some of the uninsured with incomes above $50,000 is that a little more than half of these people live with family members who are not part of their "nuclear" or insurance family unit. That is, they live with parents, grown children, or siblings, and because they are all relatives, their "family" income is higher than it would be for an insurance definition of family. Nonetheless, it is worrisome that an increasing number of uninsured people have family incomes that we think of as solidly in the middle-class section of the income distribution. We do not know how much of this growth reflects people being offered health insurance where they work but declining it for themselves or their dependents because they cannot afford the employee share of the premium.

Labor Force Status of Adults

More than two-thirds of uninsured adults are in the labor force, with 60 percent of uninsured adults working and another 8 percent unemployed and looking for work. When we count all the dependents of working uninsured adults, a little more than four out of five uninsured live with someone who works (71 percent live with someone who works full time and 12 percent live with someone who works part time, according to the Urban Institute's analysis of CPS data for the Kaiser Commission on Medicaid and the Uninsured 2000c).

Health Status

The uninsured are in relatively good health, with only 7 percent saying they are in "fair" health and another 2 percent saying they are in "poor" health. One reason more of the uninsured are not in poor health is that some of the population in poor health qualify for Medicaid or Medicare (the latter by virtue of long-term disability). Moreover, the vast majority of young people and people who work generally do not have serious medical conditions. The vignettes of the uninsured collected by the Kaiser Commission on Medicaid and the Uninsured showed people who were not in poor health in spite of the fact that they often had medical debts of $1,000 or more. The medical bills were for treatable medical episodes that occurred in emergency rooms because the people were uninsured (e.g., strep throat, childhood asthma attacks), or events such as unexpected caesarian section deliveries.

Basic Policy Dilemma

This picture of the uninsured illuminates a basic policy dilemma. On the one hand, health insurance coverage in the United States is based on employer-sponsored coverage, and we assume that working people will obtain insurance through an employer group. Employer competition for high-skill labor has forced compensation for high-skill jobs to include higher wages and fringe benefits, including health insurance. On the other hand, we have an economy where many jobs do not require higher education and/or special skills. Such jobs generally have low wages and no health insurance. (Although low-skill jobs in large firms are more likely to provide health insurance as a fringe benefit, in 1999 a quarter of all uninsured adults worked for firms with more than 500 employees.)

The fact that the labor market for low-skill workers is not tight enough to cause employers to offer health insurance for low-skill jobs is a large part of the explanation for why 60 percent of the uninsured adults are working but uninsured. Most uninsured adults do not have more than a high school education and are not skilled enough to be in high-skill jobs. This problem is further compounded by the fact that almost half of the uninsured workers are employed by firms that have fewer than 25 employees.

As I discuss below, small firms face much higher per person premiums than do large firms, which have much larger numbers of people for pooling risks of medical expenditures. Because small firms generally have small profit margins, they cannot afford to increase the compensation of low-skill workers with the relatively high cost health insurance available to them.

Thus, unless we want to radically alter the labor market for low-skill workers and the economic conditions in which small firms operate, we need to develop two concurrent policies to expand health insurance coverage. One policy would provide heavily subsidized quasi-public coverage to people with incomes below some level, such as 250 percent of the poverty level, or $35,000. The second would increase access to private health insurers for higher-income uninsured individuals by reducing the risk to insurers of covering people who do not have employer-sponsored coverage. Developing such a policy would provide a way for private insurers to continue to be the primary source of health insurance in the United States and cover more of the uninsured. To see why requires an understanding of how insurers view the uninsured and how they compete for business, the subject that we turn to next.

HOW HEALTH INSURANCE COMPANIES COMPETE

To understand the health insurance markets in the United States, we start with the fact that the majority of people obtain coverage through employers. Approximately 63 percent of the population (of all ages) have employer-sponsored group coverage.[2] Those with employer-sponsored coverage pool their individual risks of high medical care costs. Almost everyone in large employer groups participates in the employer-sponsored health insurance, so there is only a small proportion of each group who are likely to have unexpectedly high medical expenses. But people who do not have access to such pooling of risks—the uninsured and the people who obtain individual coverage—face insurance markets in which adverse selection is a major problem.

Three Interconnected Health Insurance Markets

Health insurance is sold in the United States in several interconnected markets. We can loosely distinguish between large employer group, small group, and individual (or nongroup) insurance markets. Some indemnity insurers and managed care plans (hereafter referred to collectively as carriers) actively sell coverage in all three markets, but most do not. More often, we observe large carriers selling coverage to large employer groups, and smaller carriers selling in the small group and individual markets. In addition to these three types of markets, every state regulates how insurance is sold within its borders. The states have different regulations governing facets of insurance ranging from what benefits must be covered by insurance policies to how rates are determined to requirements about financial reserves. As a result, there are 51 different submarkets within each of the three distinct markets. Many carriers, particularly smaller carriers, offer policies only in those states with similar regulations so they do not have to keep track of and respond to many regulatory changes.

One result of this is that in the individual markets in 1997, the number of carriers selling individual policies ranged from only two or three (in Delaware, Idaho, and Alaska) to more than 40 (in New York and Texas) (Chollet, Kirk, and Chow 2000). New York's relatively large number of carriers selling individual coverage is due to the requirement that all HMOs sell individual coverage. In 1997, just under 700 carriers sold individual policies in the United States; by comparison, 2,450 carriers sold policies in the large and small group markets (Chollet, Kirk, and Chow 2000). In spite of this difference, the individual and group markets are characterized by a small number of carriers having at least half of the total number of policies sold in each type of market in each state (Chollet, Kirk, and Chow 2000).

Large employers have avoided state regulations and state taxes on health insurance by self-insuring (or self-financing) their employees' health care costs. The Employees Retirement and Income Security Act of 1974 (ERISA) exempts self-insured employers from state regulations and taxes on policies sold within a state. Most self-insured employers pay a fee to a third-party administrator (almost always a carrier) to administer the claims from medical care providers, and the

employees are usually unaware that the third-party administrator is not their insurer as well.

Health coverage is sold and priced quite differently in the three types of health insurance markets (ignoring for the moment the 50 different jurisdictions' regulations). The selling practices and pricing differences largely reflect the extent to which carriers fear adverse selection in each of the markets. In the large group market, adverse selection at the group level is uncommon since almost all employees in a large company generally enroll for coverage. If an employer offers a choice of plans, then carriers may be concerned about adverse selection if they are the choice of a small proportion of the group (Buchmueller and Feldstein 1997; Cutler and Reber 1998). Employees and their dependents in large groups pay average premiums based on the total expected costs of the group; a particular person's expected medical care costs are not factored into the premium he or she pays. Usually, the employer also negotiates with several carriers as to the out-of-pocket cost sharing and benefits covered, and trade-offs between these and the premiums.

Small groups (typically, groups with less than 50 employees) and individuals face very different markets. Per policy premiums are substantially higher in these markets than in the large group market; it is not unusual to find premiums for single or family policies to be more than twice as expensive for small groups or individuals than for large groups. The primary reason for these higher premiums is that pooling of risks occurs over much smaller groups of people in the small group and individual markets. As a result, the variance on the expected costs is much larger. This creates a greater risk that actual costs will exceed expected costs by a wide margin. Carriers respond to this in two ways. First, they set higher premiums for small group and individual policies because the risk per policy is higher and they need to be compensated for bearing greater risk. Second, they try to insure only people who they expect will have lower medical costs and to avoid insuring people who they perceive to be high-cost users of medical care. Carriers go to great expense to selectively insure people who they perceive to have low risks of high medical care costs. The costs of the risk-selection mechanisms used by carriers are a large component of the higher premiums for small group and individual policies.

Information Asymmetry Shapes the Form of Competition between Carriers

Carriers cannot discern from applicant information whether an applicant will have high medical care use in the coming year. But they believe that people who apply for insurance coverage are disproportionately comprised of people who expect to have high medical care use in the near future—perhaps because they or a close relative had a medical condition in the past. The problem for carriers is that they usually cannot obtain this information; there is an asymmetry of information between what the carriers know and what the insurance applicants know. When there is asymmetric information in a market, the market cannot be competitive and inefficiency will result. In the case of health insurance markets, the carriers have the disadvantage in terms of the asymmetry of information.

Carriers' fear of adverse selection among applicants in the small group and individual markets motivates their behaviors. Carriers fear adverse selection because it causes them to underestimate premium revenues needed for expenditures and thus risk substantial financial losses. To avoid adverse selection, many carriers adopt selection mechanisms to screen out applicants whom they suspect will use expensive medical care (Swartz and Garnick 1999, 2000a,b; Chollet and Kirk 1998). Such mechanisms include medical underwriting practices,[3] refusing to issue or renew a policy, excluding coverage of services for preexisting medical conditions, and differentiating their policies from their competitors' by generously covering some types of services (e.g., preventative) but limiting coverage of other services (e.g., substance abuse treatment) (Stone 1993; Frank et al. 1997).[4]

Thus, competition in insurance markets, especially the small group and individual markets, focuses on how well carriers use mechanisms to identify which firms or individuals might be high-risk versus low-risk. As Newhouse pointed out in the context of risk adjustment models, a carrier only needs to be a little better than its competitors in the use of selection mechanisms to make more of a profit (Newhouse 1994). When carriers are not constrained in their ability to set different premiums for people who they believe have different probabilities of using expensive medical care, then carriers compete in large part in terms of the accuracy of their models for predicting a person's (or

firm's) medical expenses. These models are generally known as actuarial models because they are based on actuarial tables of likelihoods of using different amounts of medical care by many different demographic and socioeconomic characteristics as well as health status and prior use of health care.[5] Different carriers will then price their health insurance policies to people and small firms based on the individual's or firm's expenditures predicted by each carrier's actuarial model. Usually, the models are used to determine how the premiums might be underwritten for particular individuals or firms. That is, if a small firm is predicted to have a high risk of high medical expenses in the next year because several people in the group had high expenses in the last year, the carrier may agree to offer insurance only if the firm pays a substantially higher premium. The additional premium amount underwrites the basic premium for the policy.

Underwriting principles might also cause a carrier to deny coverage completely or exclude coverage for a condition to a group or person on the basis of information known by the carrier. Most states allow exclusion of coverage for a preexisting condition (such as cancer, osteoarthritis, or allergies) for a limited time period—typically 12 months. As a result, carriers more often simply deny an application if a person has had quite serious conditions, such as angina or a myocardial infarction (Chollet and Kirk 1998). In some states, underwriting of premiums is not permitted because it is viewed as a selection mechanism that discriminates against people if they are perceived to have high risks of expensive medical care. When underwriting is not permitted or its use is restricted, carriers turn to other selection mechanisms to avoid insuring high-risk people.

A frequently used mechanism for separating high- and low-risk applicants consists of differentiating the benefits (or medical services) covered by a policy. If a carrier is able to identify a health care benefit that is particularly attractive to low-risk people but not high-risk people, then it can design policies that cause people to voluntarily reveal that they are likely to be low- or high-risk people. Carriers' use of differences in benefits packages is a mechanism for getting individuals (or groups) to reveal information that separates them in terms of risk levels for nominally unpredictable expensive medical events. Thus, for example, if a person knows that cancer runs in his or her family—which the carriers do not know—the person might choose a policy that

has high upper limits on covered expenses, provides for cancer screening tests, and includes first-rate cancer centers in the list of providers. By choosing such a policy, the person is revealing information to the carrier regarding his or her risk expectations. Carriers have invested in substantial efforts to understand how differences in benefits packages can be used to attract low-risk people to some policies and high-risk people to other policies.

Carriers also have developed monopolistic market niches in the small group and individual markets as another mechanism for avoiding adverse selection (Swartz and Garnick 2000a,b). In the individual markets, for example, some carriers specialize in marketing to individuals who have left the armed services; others specialize in policies attractive to very small firms of professionals (e.g., lawyers or financial advisors) or only to individuals who are self-employed. As a result, few carriers in a state market actively compete for business among all consumers seeking individual policies, and people whom insurers perceive as high-risk have few, if any, options for obtaining health insurance (Pollitz, Sorian, and Thomas 2001; GAO 1996).

The differences in states' regulations of the insurance markets within their borders permit the greater or lesser use of these mechanisms or different combinations of the strategies to avoid insuring high-risk people. States that have attempted to block carriers' use of such preferential selection mechanisms, particularly in the small group or individual markets, have almost always set up regulations that block the use of only one or two of these mechanisms. State regulations, for example, might mandate that all policies sold in the state must cover substance abuse treatment so as to inhibit carriers' ability to avoid high-risk people who may want coverage of care for substance abuse. Some states have enacted regulations requiring carriers to accept any applicant ("guaranteed issue") so a carrier cannot turn down an applicant it views as high-risk.[6] Of course, if a state has only one or two of these regulations in place, the carriers can use other mechanisms that are not proscribed to accomplish the same objective. A common example is when a state requires carriers to accept any applicant but does not also have a regulation governing the way in which premiums can be set, we observe what should be a totally expected outcome: high-risk people are indeed offered coverage but at an extraordinarily high premium. Similarly, when states require community rating of premiums

(say, in the small group insurance market) but do not standardize the benefits to be covered in policies sold in the market, carriers can use differences in what benefits are covered under different policies to try to separate high-risk firms from low-risk firms.

In summation, the information asymmetries in health insurance markets cause the markets, particularly the small group and individual markets, to be inefficient. Inefficiency reflects the fact that enormous efforts and expense are spent in developing and applying selection mechanisms to avoid covering people who are likely to use expensive medical care. Carriers compete with each other not in terms of producing insurance per se at the lowest possible cost, but in terms of insuring as high a proportion of low-risk people as possible in order to keep costs low. Thus, the usual competitive market forces that cause producers to seek profits by reducing their costs of production and increasing market share have been altered by the fear of adverse selection in insurance markets. In insurance, carriers seek to minimize their risk of unexpected high costs by competing to have very high shares of low-risk people among the people they insure. The competition among carriers consists of trying to do better than other carriers at selecting low-risk people, which involves efforts that do not contribute to producing insurance. The costs of creating and using selection mechanisms are a measure of the inefficiency that exists in health insurance markets.

A ROLE FOR GOVERNMENT: COVER VERY HIGH-COST PEOPLE EVERY YEAR

The market failure caused by carriers' fear of adverse selection leaves us with two outcomes. One is that risk selection activities cause premiums to be substantially higher in the small group and individual market than in the large group market, making health insurance relatively unaffordable for most people who do not have access to employer-sponsored coverage. The second outcome is that a substantial number of people do not have access to health insurance, especially in the individual market, because they have some characteristic that causes a carrier to perceive them as high-risk.

The inefficiency due to expenditures on risk selection could be substantially reduced if government were to shift responsibility from the carriers to the general population for the costs of people who, each year, have very high costs—that is, people who have health care costs in the top 1–3 percent of the distribution of medical expenditures. Currently, if a carrier has enrollees with unexpectedly high costs, those costs are borne by the other people insured by the carrier and whatever stockholders the carrier may have. If the carrier has to substantially increase premiums to recover from losses due to unexpectedly high costs of some enrollees, there is a high probability that some number of enrollees who have low costs will leave the carrier in response to the premium increase. This leaves the carrier with a risk pool that has a higher average expected cost. If the following year there are again unexpectedly high costs, the cycle will repeat itself; if it continues, we have what the insurance industry calls a "death spiral," where the particular policy has to be closed down and abandoned or the carrier is forced out of business. This outcome places all the burden of insuring high-cost people on the individuals who have had health insurance from the carrier—and who have to pay higher premiums or drop their coverage —and the shareholders of the carrier.

If the costs of very high-cost people were shifted instead to the government—and thus to the entire population—carriers' fears of adverse selection and a death spiral would be substantially reduced. The burden of such costs would be redistributed from the carriers that encountered adverse selection. As a result, carriers would no longer have an incentive to use and develop risk selection mechanisms, and the inefficiency present in the small group and individual insurance markets would be greatly reduced. This would also enable people to purchase health insurance policies rather than being denied coverage.

What I am suggesting is that government—most likely the federal government, but it could be state governments—take on the role of reinsurer for carriers that have insured people who have very high medical bills in a year. That is, the government could pay a portion of the costs of those individuals whose total annual medical costs exceed some threshold—say, $30,000—or an amount that places a person's medical expenditures above the 98th or 99th percentile of the distribution of medical expenses of the entire population. Carriers often purchase reinsurance to protect themselves from the risk that an insured's

claims will exceed $50,000. Instead, if the government acted as the reinsurer for the high-cost claims, the carriers would then have far less incentive to avoid insuring people they expect to have high expenditures.

Examining the distribution of medical expenditures for the U.S. population shows why this proposal would greatly reduce carriers' incentives to use selection mechanisms. According to preliminary estimates from the 1996 Medical Expenditure Panel Survey, Monheit predicts that 68 percent of the population had medical expenditures below $1,000.[7] He further estimates that 4.5 percent of the population had expenditures between $5,000 and $9,999, while just 4 percent of the population had expenditures above $10,000. It is very difficult to predict who will have expenditures between $5,000 and $10,000 per year. But so long as a carrier is not responsible for costs of people with expenditures above, say, $30,000, then it is not worth the expense for a carrier to use risk selection methods to avoid people with expenditures in the 90th to 96th or 98th percentile of the expenditures distribution. It is simply too difficult to distinguish between people who will have expenditures at the 30th percentile and those who will be in the 5–10 percentiles below the threshold for reinsurance. Moreover, while there is some correlation between a person's medical expenditures from one year to the next, that correlation falls away when a longer period of time is considered (McCall and Wai 1983; Welch 1985; Goodman et al. 1991; Gornick, McMillan, and Lubitz 1993). Thus, we should expect that different people each year would have very high medical expenditures that would qualify for the government reinsurance.

Reinsurance almost always requires the original insurer (the carrier) to bear some portion of the costs above the threshold where reinsurance picks up insuring events. This cost-sharing is built into the reinsurance structure so the original carrier will retain an incentive to manage the health care of high-cost people. It would be important to maintain this incentive if the government were to reinsure the very high medical care expenses. In addition, for any person who has health care expenditures over the reinsurance threshold level each year, the government could cover either a portion of the costs above the threshold or a portion of all of the person's costs. In either case, the share of costs that the government would cover also could vary over different levels of expenditures. For example, the government could cover 90

percent of the costs above the eligibility threshold up to two times the threshold, 80 percent of the costs from two times the threshold up to three times the threshold, and then 100 percent of the costs above that.

Having government take on the role of reinsurer would make the small group and individual insurance markets function more efficiently. This would immediately provide what economists call a "welfare" gain to everyone who purchases health insurance in the small group or individual insurance markets, since the premiums for insurance will decline in proportion to the reduction in use of selection mechanisms. Moreover, high-risk people who currently cannot obtain coverage from all carriers also would benefit because carriers would no longer deem them undesirable. High-risk people would have greater access to carriers and policies in insurance markets.

The welfare gains caused by the increased efficiency in the insurance markets are not "free," of course. This requires government revenues to pay all or some of the medical care costs of the designated high-cost people. A political advantage of using the income tax and sources of revenues for the general revenue funds is that they do not require implementation of a new tax to pay for either a new insurance program for high-cost people or a reinsurance fund to pay carriers for high-cost claims. On the other hand, when a program is competing for general revenue funds along with high-visibility government programs—such as education, highway maintenance and construction, or homeland security—then it is vulnerable to pressures to cut the budget. This is particularly true for programs that benefit everyone, but may appear to assist only a small number of people—in this case, those individuals with high-cost claims. The argument has to be made that both of the government options for high-cost individuals increase the efficiency of insurance markets, thereby providing benefits to everyone.

Implementing an institutional structure to permit the government to take responsibility for the health care expenses of the very high-cost individuals also would require some standardization of health policies sold in the small group and individual markets. Standardizing the benefits covered by policies would make it possible to compare medical expenditure patterns of people and then to identify those people who have the very highest medical expenses. Without such standardization, it would be quite difficult to know whether a person had high expendi-

tures because of a very generous insurance policy as opposed to being quite ill.

New York State created a subsidized health insurance program for low-income individuals and small firms with low-wage workers that is very close to my proposed plan. "Healthy New York" was developed during 2000 and began enrolling individuals in February 2001. Under Healthy New York, the state pays as much as 90 percent of the costs of claims between $30,000 and $100,000 for people who have claims in a calendar year that exceed $30,000 (Swartz 2001). The money for the pool of funds that pay for these costs comes from the state's tobacco settlement funds. To ensure transparency of why people have high-cost claims, currently there is only one standardized benefits package for the Healthy New York policies. Premiums under Healthy New York for eligible low-income individuals are about 50 percent less than the premiums for individual coverage in the regular individual market; for small firms the premiums are about 15–30 percent below premiums for comparable policies in the small group market.

In sum, if government were to redistribute the risk of very high medical care costs from carriers to the broader population, efficiency would be increased in the small group and individual insurance markets, enabling more people to obtain health insurance. Premiums would be reduced because carriers would reduce their efforts to identify high-risk people whom they do not want to insure. As a result, relatively low-risk people would be more likely to obtain and retain coverage. Higher-risk people, who currently have great difficulty finding carriers willing to insure them, would have more choice of policies and carriers since there would be sharply reduced incentives for carriers to avoid higher-risk enrollees.

CONCLUSION AND POLICY IMPLICATIONS

When risk is present in markets, such as health insurance markets, market failure can be especially likely because of information asymmetry and the potential for adverse selection. Risk also can cause markets to fail to form. If government acts to take care of or remove the worst risks in such markets, the inefficiency in the markets would be

greatly reduced, and markets that otherwise could not even start up would be able to function.

There are precedents in other markets with risk where the federal government has taken responsibility for the worst risks, thereby enabling markets to function and grow. A market for reinsurance for catastrophes has developed in the United States because there has been a history (including, most recently, the response to the terrorist attacks of September 11, 2001) of the federal government stepping in to pay large fractions of the costs of catastrophes. Indeed, the creation of the Federal Emergency Management Agency in 1978 formally acknowledged the federal government's role in assisting with recovery from catastrophes. The secondary mortgage market in the United States, which enables lenders of mortgage money to replenish their capital, was established because the federal government has taken responsibility for the worst-risk mortgages since 1954. The Federal Housing Authority (FHA) and the Veterans Administration (VA) shifted the risk of default from mortgage lenders to the federal government for people who otherwise would not have qualified for mortgage loans. The FHA mortgage insurance and the VA mortgage guarantee program set minimum standards for what properties were eligible for mortgages and what types of financial information were needed from borrowers. This standardization of information permitted mortgages to be resold on a national basis because standardized information made it easier for lending institutions that were not local to perform due diligence investigations of mortgages that were offered for resale in the secondary mortgage market. In addition, very high-risk mortgages are backed by federal guarantees. It is unlikely that either the reinsurance market or the secondary mortgage market would function without the government backstopping them by covering the worst risks.

Similarly, if government were to reinsure the costs of those individuals with the highest medical expenditures each year, the risk of very high costs would be shifted from carriers to the general population. This would cause carriers in the small group and individual insurance markets to spend substantially less on efforts to avoid insuring people they perceive to be likely to have high costs. In turn, this would reduce the rates for health insurance faced by people who purchase insurance in these markets and enable a much larger set of people to obtain health coverage.

Having the government act as reinsurer, along with backstop carriers in the individual and small group markets, will help about a third of the people who currently are uninsured. The remaining uninsured do not have sufficient incomes to afford health insurance unless it is heavily subsidized. As we noted earlier, many of the low-income uninsured have medical debts for highly treatable episodes of care. Such debts would be far lower if the people had obtained medical care in settings other than hospital emergency departments. To facilitate the use of more efficient settings for medical care by the low-income uninsured, government should either create more community health care centers or extend eligibility to adults for public programs similar to the SCHIPs. Such government moves also would increase efficiency in the provision of health care to the very low-income uninsured.

Finally, the rising costs of medical care mean that health insurance premiums will also increase, along with increased cost-sharing required when people use medical care. If the past is any indication of how this will affect people's decisions to purchase insurance or take up employer-sponsored coverage, the rising costs will lead to greater numbers of uninsured as more people come to view health insurance as unaffordable. As we have seen in the last decade, the uninsured are increasingly people with lower middle-class incomes. We need to rethink both how we provide and finance health insurance if we are to avoid rising numbers of uninsured—such rethinking could begin with the government taking on the role of reinsurer for small group and individual health insurance markets. The government as reinsurer provides a mechanism for public funds to enable private health insurance markets to operate efficiently and be accessible to more people.

Notes

The author gratefully acknowledges the financial support of the Robert Wood Johnson Foundation under Grant Number 033818. Opinions expressed are those of the author and may not represent the opinions of the Robert Wood Johnson Foundation or Harvard University.

1. In the case of public goods, the argument is that government should produce the goods because a market cannot be sustained.
2. Based on estimates by the Census Bureau from the March 2001 Current Population Survey. The estimates by type of coverage are not mutually exclusive because

people can be covered by more than one type of health insurance during the year, and in some cases at the same time (for example, some people have both Medicare and Medicaid coverage). See <www.census.gov/hhes/hlthins/hlthin01/fig03.gif>.

3. Medical underwriting is the process by which carriers set the premium for an applicant based on the person's expected medical care costs. Thus, if a person has poor health status, actuarial underwriting practices would yield a higher premium than that for a similar person in excellent health. The underwriting process essentially determines whether a person pays an additional amount plus the base premium for the policy.
4. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) has been sometimes mistakenly assumed to restrict these selection practices in the individual insurance market. HIPAA does not prohibit carriers from applying selection practices to the great majority of individuals who seek coverage in the individual insurance markets. See Nichols and Blumberg (1998) for details.
5. Applicants in both the small group and individual markets generally have to respond to questionnaires about their health status, use of medications and medical care in the past, and health risk behaviors. It is not unheard of for small groups to be offered coverage for most but not all of the members of the group, with the rejected members being denied coverage because carriers believe they will have high medical expenditures.
6. For example, Washington State, New York, and New Jersey's individual insurance markets are required to guarantee issue of policies to any applicant regardless of the applicant's health status, age, gender, or place of residence.
7. Communication between Alan Monheit and the author, Spring 2001. Monheit and Marc Berk have analyzed the distribution and concentration of the population's medical expenditures. See, for example, Berk and Monheit (2001).

References

Berk, M., and A. Monheit. 2001. "The Concentration of Health Expenditures, Revisited." *Health Affairs* 20(2): 9–18.

Buchmueller, T.C., and P.J. Feldstein. 1997. "The Effect of Price on Switching among Health Plans." *Journal of Health Economics*. 16(2): 231–247.

Chollet, D.J,. and A.M. Kirk. 1998. *Understanding Individual Health Insurance Markets: Structure, Practices, and Products in Ten States*. Menlo Park, CA: The Henry J. Kaiser Family Foundation.

Chollet, D.J., A.M. Kirk, and M.E. Chow. 2000. *Mapping State Health Insurance Markets: Structure and Change in the States' Group and Individual Health Insurance Markets, 1995–1997*. Washington, DC: Academy for Health Services Research and Health Policy.

Cutler, D.M., and S. Reber. 1998. "Paying for Health Insurance: The Trade-off between Competition and Adverse Selection." *Quarterly Journal of Economics* (113)2: 433–466.

Frank, R.G., T.G. McGuire, J.P. Bae, and A. Rupp. 1997. "Solutions for Adverse Selection in Behavioral Health Care." *Health Care Financing Review* 18(3): 109–122.

General Accounting Office (GAO). 1996. *Private Health Insurance: Millions Relying on Individual Market Face Cost and Coverage Trade-offs*. GAO/HEHS-97-8.

Goodman, M.J., D.W. Roblin, M.C. Hornbrook, and J.P. Mullooly. 1991. "Persistence of Health Care Expense in an Insured Working Population." In *Advances in Health Economics and Health Services Research (Risk-Based Contributions to Private Health Insurance)*, Vol. 12, R.M. Scheffler and L.F. Rossiter, eds. Greenwich, CT: JAI Press, pp.149–173.

Gornick, M., A. McMillan, and J. Lubitz. 1993. "A Longitudinal Perspective on Patterns of Medicare Payments." *Health Affairs* 12(2): 140–150.

Kaiser Commission on Medicaid and the Uninsured. 2000a. *Making It Simple: Medicaid for Children and CHIP Income Eligibility Guidelines and Enrollment Procedures.* Menlo Park, CA: The Henry J. Kaiser Family Foundation.

———. 2000b. *In Their Own Words: The Uninsured Talk About Living Without Health Insurance*. Menlo Park, CA: The Henry J. Kaiser Family Foundation.

———. 2000c. *Health Insurance Coverage in America 1999 Data Update*. Menlo Park, CA: The Henry J. Kaiser Family Foundation.

McCall, N. and H.S. Wai. 1983. "An Analysis of the Use of Medicare Services by the Continuously Enrolled Aged." *Medical Care* 21: 567–585.

Newhouse, J.P. 1994. "Patients at Risk: Health Reforms and Risk Adjustment." *Health Affairs* 13(1): 132–146.

Nichols, L.M. and L.J. Blumberg. 1998. "A Different Kind of 'New Federalism'? The Health Insurance Portability and Accountability Act of 1996." *Health Affairs* 17(3): 25–42.

Pollitz, K., R. Sorian, and K. Thomas. 2001. *How Accessible Is Individual Health Insurance for Consumers in Less-than-Perfect Health?* Menlo Park, CA: The Henry J. Kaiser Family Foundation.

Stone, D.A. 1993. "The Struggle for the Soul of Health Insurance." *Journal of Health Politics, Policy and Law* 18(2): 287–317.

Swartz, K. 2001. *Healthy New York: Making Insurance More Affordable for Low-Income Workers.* Report 484, New York: The Commonwealth Fund.

Swartz, K., and D.W. Garnick. 1999. "Can Adverse Selection be Avoided in a Market for Individual Insurance?" *Medical Care Research and Review* 56(3): 373–388.

———. 2000a. "Lessons from New Jersey's Creation of a Market for Individual Health Insurance." *Journal of Health Politics, Policy and Law* 25(1): 45–70.

———. 2000b. "Adverse Selection and Price Sensitivity When Low-Income People Have Subsidies to Purchase Health Insurance in the Private Market." *Inquiry* 37(1): 45–60.

Welch, W.P. 1985. "Medicare Capitation Payments to HMOs in Light of Regression Towards the Mean in Health Care Costs." In *Advances in Health Economics and Health Services Research (Biased Selection in Health Care Markets)*, Vol. 6, R.M. Scheffler and L.F. Rossiter, eds. Greenwich, CT: JAI Press, pp. 75–96.

6
Risk and Agriculture

Some Issues and Evidence

Rulon D. Pope
Brigham Young University

As a subfield in economics, agricultural economics has an unusual genesis and hence an unusual orientation. In production, its roots are found in the study of agronomy and horticulture. Out of these disciplines grew studies and training in farm management. As Marshall's (1920) marginal analysis reached its climax, agricultural economics was just beginning to emerge as a discipline in land-grant colleges throughout the United States. It embraced marginal economic analysis, comparative advantage, and competition as important insights into market behavior. A hard-fought view began to emerge that the behavior of those involved in agriculture throughout the world was consistent with these basic economic concepts (summarized nicely in Schultz 1980). Yet, agricultural economics has always strived to help family farms (in the United States or abroad) understand more fully their economic environment. Thus, there has always been a normative dimension to agricultural economics as well (similar, perhaps, to finance in a business school). In most other fields of economics, economists are not so presumptuous as to suggest to economic agents how they should optimize—unless it is the government.

Today, agricultural economics considers a broad set of issues and behavior about resources, consumers, the environment, and policy about food and fiber using the full range of current economic concepts and methods. Likely second only to finance, agricultural economics has embraced risk concepts as an essential ingredient to understanding and prescribing behavior. It was an early entry into experimental economics by measuring individual risk preferences and subjective probabilities across a relatively broad set of agents (see Young et al. 1979;

Nelson and Bessler 1989). The purpose of this chapter is to highlight a few selective but central concepts, issues, and contributions about agricultural behavior under risk. Although the central paradigm of economic behavior is called *expected utility maximization*, the distinctions among various models of behavior under risk will not be important. Indeed, the relevant concepts and issues can be portrayed as a choice among distributions based upon the mean (a measure of central tendency) and the variance (a measure of dispersion). Such models can often be rationalized as maximization of expected utility (Meyer 1987). The main normative and positive issues are about choices that reduce risk but, even more fundamentally, raise expected utility. It would be impossible to cover all of the relevant topics, but this chapter will address some issues central to agricultural economic research on risk (see Robison and Barry 1987; Just and Pope 2002; and Caswell 1995 for more in-depth discussions).

It is useful to state at least my perception of a few generally relevant economic facts about agriculture that serve as background:

1) Agricultural production is atomistic and is generally placed on international markets. However, demand for raw agricultural products is much more concentrated than final consumer demand for food products. This implies that farms and final consumers are generally price takers with international shocks readily transmitted to agricultural markets. There is often an underlying suspicion by many agricultural producers that markets are unfair to them because of this alleged asymmetric market power.

2) Farm products have relatively price-inelastic demand and supplies. Income elasticities of demand for many raw food goods are relatively low compared to manufactures and services. Much has been made of the inelastic demands and supplies in agriculture, implying that shocks have greater price and income consequences than in many sectors.

3) Production is heavily constrained by biological processes that have long lags between the point in time in which a decision is made and its ultimate consequences. This is particularly notable in livestock production but is prevalent throughout agriculture.

4) Production is heavily seasonal with definitive intra-seasonal stages of production.

5) Investment decisions tend to have long physical and economic lives. Land often has low alternative uses, except near cities.

6) Weather, disease, and pests (vicissitudes of nature) are direct and pervasive in agriculture.

7) Government policy is omnipresent and often intrusive in market outcomes (e.g., the Common Agricultural Policy in the European Union, target prices, and subsidies in the United States). In developed countries, policy generally attempts to raise farm incomes and often raises consumer prices. In developing countries, policy often attempts to lower consumer prices.

8) Most of the demands on factors of production (inputs) in crops are inherently spiked rather than distributed uniformly throughout the season. This may imply an incentive to choose productive activities that don't compete for resources at a given point in time.

9) Institutions for the ownership of factors of production and the organization of production vary widely throughout the world.

10) Evidence seems to suggest that yields are generally increasing over time but that deviations about this trend are random (not bunchy). However, prices are highly correlated (thus, bunchy) in adjacent time periods.

It is also useful to briefly state some stylized facts about U.S. agriculture.

1) Production occurs in predominantly single or family-run enterprises. Despite ever-increasingly larger farms, the last agricultural census shows that over 85 percent of farms are "family farms," and true corporations (beyond small family-held corporations) make up only 0.4 percent of producers (Allen and Lueck 1998).

2) Structural changes in livestock production have been dramatic and often resemble manufacturing with large scale and substantial division of labor. There is substantial contract farming where

farmers supply only some of the inputs and are paid incentive contracts for producing.

3) Production is increasingly specialized during the post-war period, but multi-output production is still common. There appear now to be substantial returns to specialization of production. There is widespread innovation with continual technical progress, and there are large numbers of strains of a given crop or livestock available for production with different inherent characteristics.

4) Many farms rent and own land; thus, contracting for the services of land is ubiquitous.

5) Crop insurance and disaster relief have been the center of policy debates in recent decades; price supports and production controls were central in earlier times.

6) An interesting aspect of U.S. agricultural data is that there are little farm level data available to researchers. There is a small set of selective (not random) panels in a few states, but the data are often of limited value for the questions studied and they are not widely available. This constrains the kind of evidence that is accumulated (Just and Pope 2001).

DIVERSIFICATION AS A RESPONSE TO RISK

Since at least the early 1950s, risk reduction through diversification has received considerable attention. On the prescriptive side, agricultural economists studied and proposed various diversification strategies to reduce risk. However, there was a rather serious policy aspect to this research. If farmers have significant opportunities to reduce their risk, then perhaps some of the rationale for agricultural policy needs to be rethought. The basic incentive for diversification is widely known and can be discovered with a simple thought experiment. Suppose that the variance of the net income from a 1,000-acre corm farm is a number labeled σ^2, while the expected return from the farm is labeled μ. If another crop exists—say, soybeans, which has an

identical and independent distribution to corn—then as Samuelson (1967) has shown, the optimal choice for a risk-averse individual is to plant 500 acres in each crop. Using variance to illustrate, this perfect diversification will reduce the variance of income from σ^2, when specialized in corn, to $\sigma^2/2$ when the farm is diversified. The variance of farm income is reduced in half by diversification, and a risk-averse producer would presumably find this attractive. This is the incentive for diversification: low returns in one enterprise may be mitigated by high returns in the second enterprise. Indeed, if there are N identically and independently distributed enterprise returns, the variance can be further reduced to σ^2/N by putting $500/N$ acres in each enterprise (this assumes that there are no economies or diseconomies of scale). In cases of more general distributional settings (uneven means, variances, and nonzero covariances), there is a marginal benefit from diversification (reduction of the variance) and a marginal cost (reduction in expected or average income by not specializing in the activity or enterprise with the largest expected return).

The standard approach to economic behavior up to 1950 implied specialization: choose the enterprise with the highest expected return. This is equivalent to maximizing any increasing function of expected wealth or maximizing

$$(6.1)\quad U = u(W_0 + \mu),$$

where μ is expected net income from farming and W_0 is initial certain wealth. Thus, if confronted with the choice of producing corn, which is expected to yield $25 per acre, and hay, which is expected to yield $15 per acre, a prudent farmer maximizing expected wealth would specialize in the production of corn. However, in 1952, E.O. Heady argued that farmers likely had distaste for risk (risk aversion) as measured by the variability or variance of net income. Given estimates of individual enterprise variances and covariances, farmers can analytically choose the crop or enterprise combination that minimizes total farm variance of income. This procedure focuses on the benefits from diversification and highly favors diversification rather than specialization as an optimal decision. Knowing that enterprise expected returns will likely be unequal, Heady also discussed choices that minimize variance for a

given expected income appropriate for any mean-variance utility function of the form

(6.2) $U = u(W_0 + \mu, \sigma^2)$,

where μ is expected farm income and enters utility u positively, σ^2 is the variance of farm income, and U is utility (expected utility). In this chapter, an individual is "risk responsive" when one includes the variance in a maximization such as in Equation (6.2). A person is "risk-averse" when increased variance reduces utility. In the summary in Young et al. (1979), most of the individual farmers whose preferences were elicited were not risk-neutral for some decisions—a majority were risk-averse, and some were mixed, meaning that for some decisions a person might be risk-averse and for others risk preferring or neutral.

In Heady's analysis, expected farm income is: $\mu = h_1\pi_1 + h_2\pi_2$, where h is the proportion of total land or investment in enterprise 1 and π_1 and π_2 are expected returns per unit of land or investment on enterprises 1 and 2 respectively. Similarly, the variance of farm income for two enterprises is

$$(6.3)\quad \sigma^2 = h^2\sigma_1^2 + (1-h)^2\sigma_2^2 + 2h(1-h)\sigma_{12},$$

where σ_1^2 is the variance of enterprise, 1 income and σ_2^2 is the variance of enterprise 2 income, and σ_{12} is the covariance of the two incomes. Heady found that if a typical Iowa corn farm diversified by halving corn acreage and correspondingly increasing hay production, variance of income could be reduced substantially without a significant reduction of expected income. This is because the correlation between hay and corn income is relatively small, 0.45. Thus, large random draws in one crop's income are often offset with low random draws in the other crops income. This incentive becomes most pronounced for enterprises whose outcomes tend to be independent or are negatively correlated with similar means. When expected returns are very different in the two enterprises, then specialization of production becomes more likely.

Subsequent writers added the possibility of renting in or out land (Johnson 1967). In this case, the square root of the variance or the stan-

dard deviation of income can be linearly reduced by choosing to produce fewer (or less) risky enterprises and engaging in more safe activities. Examples of the latter are cash lease of land and investing in a risk-free asset in the case of capital. In cases where the cash lease or risk-free rate exceeds the expected return from risky enterprises, then specialization in the risk-free asset is predicted under risk aversion (see Equation (6.2)). In cases where the expected risky return exceeds the cash lease rate, firms combine the two according to their tastes for risk (risk aversion). A similar conclusion can be obtained for investment capital among risk-free and risky assets.

Many authors attempted to use quadratic programming or an equivalent mathematical programming model to identify the risk-efficient (minimizing variance for a given mean) set of enterprise choices for farms, regions, or countries. The main advantage is that quadratic programming models of farms could integrate many production constraints on firm behavior. For example, perhaps machinery and labor supplies were limited throughout the months of a growing season. In addition, they could include many policy constraints or incentives, such as land set-asides. However, the normative and positive content (what farms should do and what they do) of these models is only as good as the models themselves. Failure to reflect individual preferences, beliefs, or constraints will yield recommendations or insights that may be irrelevant to a decision maker. Additional effort is needed to understand what decision makers actually do in their response to risk.

Lin, Dean, and Moore (1974) attempted to test whether programming models incorporating response to risk (variance) were better than risk-neutral models. Using elicitation techniques, the preference functions of a small set of farmers were estimated. Many of these preference functions implied that the mean and variance of incomes should enter into farmers' objective functions. Mathematical programming models using these general objective functions were superior at predicting what farmers actually did when compared to models based solely on maximizing expected farm income. Though by today's standards the techniques and evidence used to advance their argument might be rather unconvincing, it was and is an important paper in positive economics, convincing many that risk was fundamental to understanding behavior in agriculture. This paper confirmed empirically

what other researchers had suspected: “risk aversion was superior to risk neutrality for explaining behavior” (e.g., Officer and Halter 1968).

Spatial Diversification

Soon after Heady’s work, Emery Castle (1954) noted that area diversification was almost as important as crop diversification. Indeed, spatial diversification apparently has been a successful strategy for risk reduction since medieval times (McCloskey 1976). Formulae for the variance across farms always involve the covariance (see Equation (3)) or correlation, which is the covariance divided by the product of the square root of the variances. Spatial diversification becomes particularly useful if the correlation across farms is sharply reduced as distance between the plots increases.

Jensen (1961) argued that spatial diversification was an important managerial technique open to dryland farms in the Great Plains because of idiosyncratic weather across areas. Thompson and Wilson (1994) argued that one of the primary reasons that Mexican ejido communal farmers resisted privatization of grazing land is that yields are variable with highly idiosyncratic weather patterns. Farmers could readily reduce the variance of their yields by scattering production spatially. Of course, spatial diversification has a cost in terms of expected return (increased travel costs), but apparently the benefits are sufficient to make it viable.

Davis et al. (1997) found that the correlation between yields of different peach orchards decreased 2.28 percent for each mile of separation, which could be a significant factor in the pattern of operation. It should also be mentioned that larger farms often have a significant advantage due to very subtle advantages in diversification. Many farms, such as orchards, have different responses at different elevations. Thus, a farm can in some cases gain a significant reduction in the variance while having contiguous plots by diversifying by elevation. However, more research is required to know how extensively spatial diversification techniques are used.

Econometric Models of Risk Response

There was an increasing awareness in the decade of the 1970s that the evidence via programming models of risk-averse or risk-responsive behavior was not on sound statistical footing. In mathematical programming models, parameters are usually estimated and treated as exact. Hence, there was not a readily deducible metric to decide when something like the null hypothesis of "no aversion or risk response" could be rejected. As with other fields of econometrics, programming models gave way to the search for econometric evidence. These models often made very simple assumptions about constraints, but the results were more easily amenable to inference. The ability to incorporate more complex constraints on behavior in econometric models may imply that more of the old programming constraints will find their way into econometric models (Andrews 2001).

For over four decades, agricultural economists had been using computers to estimate short run supply functions essentially of the form

$$(6.4)\quad A = b_0 + b_1\mu_p + b_2 z,$$

where μ_p is the expected price (or yield, or both) of the crop or livestock, A would be acreage or supply, z represents other variables, and the $b_0 - b_2$ are constants to be estimated econometrically. One prominent example of such an approach is the adaptive expectation model discussed in undergraduate econometrics texts. The coefficient of b_1 is presumed to be positive and the larger the magnitude of b_1, the more elastic is the supply.

Around 1970, some argued that this approach was limiting because it didn't capture risk response. Behrman (1968) incorporated risk response in agriculture as he studied crop production in Thailand. This is a large and careful study. More importantly, a regression was estimated of the following form:

$$(6.5)\quad A = b_0 + b_1\mu_p + b_2\sigma_p^2 + b_3 z,$$

where μ_p is an estimate of expected price (yield or both), σ_p^2 is an estimated of the variance (or standard deviation) of price, yield, or both, z represents other variables (for example, the means and variances of

substitutes and complements), and the $b_0 - b_3$ are constants to be estimated econometrically. Behrman found that in a preponderance of the cases, b_1 was estimated to be positive and statistically different from zero, indicating that supply curves are upward-sloping in expected price. In a majority of cases, b_2 was estimated to be negative and statistically different from zero. This was particularly true for upland crops that are sold on the market, unlike rice, which is often consumed by the farm family. Behrman concludes that "The estimated responses to the relative standard deviations do provide further support, however, for the hypothesis that the agricultural sectors in underdeveloped countries respond negatively to risks" (p. 336). Six years later, another work on risk response was very influential. Just (1974), using a Bayesian approach, formalized the estimation and specification of the mean and variance of revenue, including complementary and substitute crops, and estimated an acreage response model like Equation (6.4) for counties in California. He concluded that there was convincing statistical evidence that b_2 is negative for many crops. Thus, it appeared that risk response was not limited to developing countries.

A number of papers during the next three decades sought to determine whether a model like Equation (6.5) captures something that Equation (6.4) does not. Indeed, there has been mounting evidence of risk-responsive behavior across many commodities, countries, and aggregations. Table 6.1 summarizes a sample of these studies. Though the elasticities measuring risk response are often low in absolute value (column 5), they usually have the expected sign (negative), and risk coefficients are statistically significant (column 3). In some cases, the response to the risk of competing crops can be captured (column 4). For example, more corn acreage may be planted when the risk in soybeans increases. Various measures of risk can be constructed (column 6), and this issue continues to be a matter of research and controversy. In many cases, both μ_p and σ_p^2 are estimated using weights of past observations. This is called *adaptive* in the table. In this case, risk is measured by a backward looking mechanism; surprises in the past affect the expectation of the future variance of price, yields, or revenue per acre.

To illustrate, μ_p might be a weighting of the previous three years of prices with weights summing to one. Similarly, σ_p^2 is estimated by weighting the last three years of squared deviations about μ_p (Chavas

and Holt 1996). More sophisticated single-equation approaches use long memory geometrically declining weights (Just 1974), ARCH/GARCH with conditional or time varying variances, and/or rational risk (Aradhyula and Holt 1989; Myers 1989; Holt and Aradhyula 1990, 1998; Holt and Moschini 1992). Rational risk implies that the mean and variances implied by the model match the market data given available information. One must build up a structural model of the supply and the demand side of the market to yield expected price and the variance of price given available information. Then, the restrictions implied by the rational expectation hypothesis must be imposed. One of the most impressive but complex applications of rational risk is found in Holt and Aradhyula (1998), where a carefully specified model of the broiler market is estimated. Risk-responsive behavior was evident. More complicated still would be to estimate a complete model of production or supply and inputs demanded (factor demands), such as chemicals, labor, land, and machinery and product supply using rational expectations of the first two moments of price. A number of authors have estimated such models without explicit complicated expectational schemes and found evidence of risk-responsive behavior (Antle 1987; Chavas and Holt 1996; Love and Buccola 1991; Saha, Shumway, and Talpaz 1994; Coyle 1999).

In summary, the available econometric evidence suggests that firms rebalance their production portfolios such that when the perceived risks of an enterprise increase, farms substitute toward less risky enterprises. Taken as a whole, this evidence is very persuasive that these models capture something. However, for some, there are still reservations about the explanation of risk aversion for these risk effects. That is, is it possible that Equation (6.5) merely picks up a nonlinearity, lags, or aggregation problems (e.g., Pope 1981)? Part of the reason for this skepticism is the very success of the approach. When Equation (6.5) is applied to highly aggregated data where risk measures are substantially compromised and/or in markets where reasonably good futures markets exist, it still seems to work well. The question then is not one of insufficient evidence, but of interpretation of the evidence.

Table 6.1 Selected Risk Response Studies

Author	Dependent variable	Significant negative own risk coefficient[a]	Significant cross risk coefficient	Own risk elasticity (short run)[b]	Risk measure[c]
Behrman (1968)	Rice, corn, casaba, A[d]	Yes	N.A.[e]	Small but negative	3-year std. dev. (moving)
Just (1974)	Grain, cotton, A	Yes	Yes	Not calculated	Adaptive like (infinite)
Lin (1977)	Wheat, A	Yes	N.A.	Not calculated	3-year std. dev.
Estes et al. (1981)	Potato, A	Yes	N.A.	Not calculated	Adaptive
Hurt and Garcia	Sow farrowing	Yes	Yes	–0.47 to –0.56	Adaptive
Brorsen, Chavas, and Grant (1985)	Wheat margins (f-m & m-r)[f]	Yes	N.A.	Not calculated	Adaptive
Aradhyula and Holt (1989)	Broilers	Yes	Yes	–0.045	GARCH rational
Holt and Aradhyula (1990)	Broilers	Yes		0.232, –0.012, –0.046	GARCH
Chavas and Holt (1996)	Corn and soybean, A	Yes	No	Not reported	Adaptive
Love and Buccola (1991)	Corn and soybean system, A	Yes	No	N.A.	Yes
Pope and Just (1991)	Potato and sugar beet, A	Yes	Yes	Not reported	Adaptive
von Massow and Weersink (1993)	White beans, corn soybeans, wheat, A	Yes	Yes	–0.073 to –0.220	Adaptive
Saha, Shumway, and Talpaz (1994)	Wheat system	Yes	N.A.	N.A.	N.A.

Holt (1994)	Corn, A	Yes	Yes	–0.018	Rational
Duffy, Shalishali, and Kinnucan (1994)	Cotton, corn, and soybean, A	Yes	No	Not reported	Adaptive
Krause, Lee, and Koo (1995)	Wheat, A	Yes	N.A.	–0.062 to 0.003	Adaptive
Krause and Koo (1996)	Wheat, barley, flaxseed, and oil sunflower, A	Yes	Yes	–0.05 to –0.01	Adaptive
Tronstad and McNeill (1989)	Sow farrowing	Yes	Yes	–0.0013 to –0.164	Downside
Bar-Shira, Just, and Zilberman (1997)	Crop system	Yes	Yes	N.A.	Adaptive
Coyle (1999)	Crops and livestock system	Yes	No	Not reported	Adaptive

[a] Often a single paper includes a variety of specifications. "Yes" means that some of the risk parameters were significant.

[b] Often a single paper includes a variety of specifications. Thus, the elasticities reported are an attempt to convey approximate risk response.

[c] Adaptive here is used very loosely. It is intended to imply a weighting scheme where the weights sum to one. Some "adaptive" used polynomial lags rather than geometric declining; some use simple fixed weighting schemes.

[d] A = crop acreage or similar spatial measure.

[e] N.A. means "not applicable."

[f] f-m and m-r are margins: respectively, farm to mill and mill to retail.

SOURCE: Author's calculations from selected cited references.

Other Forms of Diversification

Another risk-reducing activity is to diversify family labor. Mishra and Goodwin (1997) find a significant positive relationship between the coefficient of variation of farm income (standard deviation divided by the mean) and off-farm employment. Thus, when farm income is more variable, one risk-reducing strategy is to apply more of one's labor portfolio to safer off-farm income-generating activities. Further, farm operators who receive large payments from government farm programs are less likely to supply off-farm labor. Both of these findings are consistent with farmers balancing risks in a portfolio generated by owner labor and owner capital.

Not only can a farmer self-insure through reducing labor allocations to risky endeavors, but capital can be allocated to safe investments as well. Mishra and Morehart (2001) calculate that off-farm financial assets in 1995 for the United States were 18 percent of total assets for farm families. This is up from 14 percent in 1990. Thus, farms are becoming more diversified outside of agriculture. One way to view these data is that agriculture is more risky so farmers are increasingly diversifying outside of agriculture in order to reduce the risk of total wealth or income. Perhaps recent market events have reversed that trend.

Diversification and Farm Size

Because the foundation of much agricultural policy in the United States historically has hinged on the survival of the family farm, one issue of concern is the relationship of scale and risk. Unlike the portfolio approach, there may be substantial economies of scale in production; that is, as a farm produces more of a particular crop, marginal and average costs of production fall. Economies of scope may also arise, meaning there are cost advantages to diversification. Part of the reason for such economies of scope is that there are inputs that are productive across products. For example, a tractor can be utilized to produce a variety of crops, especially when they don't compete at the same time for services. Economies of scope imply that expected utility with two products is greater than expected utility when specialized. These can come from the diversification motive discussed above or from cost

advantages from public-like inputs. Economies of scale promote specialization while economies of scope promote diversification. Thus, there are three relevant effects: diversification in response to risk with no scale effects, scale economies in enterprises, and economies of scope due to cost advantages in jointly producing two or more products.

Attempting to empirically untangle these three effects is difficult, and there is not a satisfactory conclusion. Econometric studies provide some evidence that larger farms are more diversified, *ceteris paribus*. Further, wealthier farms, *ceteris paribus*, have less diversification (Pope 1976; Pope and Prescott 1980; Dunn and Williams 2000; Zenger and Schurle 1981). Pope (1976), using factor analysis for California farms, found evidence that there is a combination of minimum efficient scale and economies of scope due to spreading the services of fixed inputs across time.

However, looking broadly across this literature, one is struck by the large volume of prescriptive literature on optimal diversification and the relatively small set that positively examines behavior. Though there is little doubt that the principles of diversification are always potentially important, exactly where they are used is still a matter of some debate. For example, perhaps farm diversification across enterprises due to risk aversion is relatively unimportant in explaining farm behavior.

If initially an optimal portfolio of actions or investments is chosen, then policy that reduces the risk in a particular agricultural commodity will see greater supply of that commodity (and at the expense of others). Thus, a well-meaning policy attempting to assist wheat farmers in the northern plains because of variable profitability may have the unintended consequence of increased production and greater demand for help in the future. Further, it is apparent that behavioral and market responses to risk may be diminished in response to a public policy that attempts to reduce risk. For example, diversification may fall if the government provides a safety net for farms.

RISK REDUCTION AND INPUT USE

One can surely view the entire portfolio choice as one of choosing inputs, such as land allocations for crops. However, one aspect of input choice applies to specialized or diversified farms and asks what the distributional consequences are of input choice. To illustrate the issue, suppose that a farm is specialized in the production of corn. Corn has a production function that depends on an input, *h*. Both its mean output and the variance of output depend on *h*. That is, $\mu = \mu(h)$; $\sigma^2 = \sigma^2(h)$. What might a farmer do in choosing how much of this input to apply? Just as in diversification, there is a marginal benefit in that expected output would increase. For example, if fertilizer is applied, we expect it to raise output or to have a positive expected marginal product generally. At some point, it is expected that additional fertilizer will diminish output (negative expected marginal product). If the farmer's only concern were expected profit μ, the farmer would choose fertilizer such that the expected marginal benefit equaled the cost (price) per unit of the fertilizer. In which case, economists say that a farm chooses inputs such that the expected marginal revenue product is equal to marginal factor price. That is, the marginal benefit of input use is equal to its marginal cost.

However, if the farmer is risk-averse, there is concern with how increasing input allocations might alter the variance of profit. If the decision maker is risk-averse, then increasing the variance of profit will reduce utility. The important question is how each input contributes to expected profit and the variance of profit at the margin. We will call inputs that reduce the variance of profits at the margin *risk-reducing*, while inputs that increase the variance of profits are *risk-increasing*. If an input is risk-reducing and the farm is risk-averse, there will be an additional marginal benefit from using more of it than would be implied by maximizing expected profit. This is a self-insuring technique. Firms might have more machinery or labor than would seem advisable based upon average marginal product because using more of it reduces risk. Farms might use more pesticides than would seem profitable on average because of its self-insuring capabilities. A little reflection shows that irrigation may perform that function. Irrigation often virtually lops off the lower tail of the distribution of yields; nec-

essarily, it reduces the variance (and raises the average) of crop yields. However, land is likely risk-increasing under this definition: adding more acres of corn increases the variance of profits (and expected profit). The policy significance of describing agricultural technology is apparent. If agricultural decision makers believe that the environment is more risky, then they may use more of inputs that lead to degradation of the environment when the inputs are risk-reducing.

The motivation for answering the question about how inputs affect the distribution of output comes from two sources. First, it is relevant to prescribing optimal input use to farms. It is particularly relevant in developing countries. If a modern variety of a crop is chosen, there are often very large variations in output if there are modest variations in inputs like fertilizer. This is often in contrast to native varieties, which have a number of resistances to input variations. Secondly, there are many environmental issues regarding the use of modern chemicals.

Roumasset (1976) considered rice production in the Philippines in 1971–1972 and found that the green revolution was not as successful as expected. Farmers often adopted "miracle rice" varieties, but they did not use the recommended amount of nitrogenous fertilizer. It was hypothesized that less than the recommended level of fertilizer was used because of risk aversion. After estimating risk preference functions and the random properties of technology, Roumasset discovered that risk neutrality was more consistent with observed behavior than was risk aversion, contrary to the Officer and Halter (1968) and Lin, Dean, and Moore (1974) conclusion that risk aversion often explained behavior better than risk neutrality.

As mentioned earlier, most attention was focused on chemical inputs. The empirical results were often mixed, but there is no reason for inputs to behave similarly across soil qualities, climatic conditions, and crops. Secondly, results vary because of many methodological issues associated with functional forms and estimation of higher order moments. Regev, Gotsch, and Rieder (1997) found significant evidence that fungicides are risk-increasing at low levels of rainfall, but found no conclusive evidence of nitrogen being risk-reducing or risk-increasing. Horowitz and Lichtenberg (1993) found evidence that fertilizer and pesticides may be risk-reducing. Mixed results on pesticides are found in Carlson (1979), Horowitz and Lichtenberg (1994), and Hurd (1994). There seems to be a growing consensus that there is no

evidence that pesticides are risk-reducing. Thus far, agricultural economists are only beginning to build a consistent body of findings upon which to infer a coherent set of stylized facts about risk-reducing/increasing inputs (Antle 1983; Griffiths and Anderson 1982; Hall and Moffit 1985; Just and Pope 1979). The most clear-cut evidence seems to come from experimental plots commonly studied by agricultural experiment stations throughout the world, but there are questions about how these data apply to actual farming experience under less controlled situations.

CROP INSURANCE

As economists have thought about the new economics of uncertainty, one of the early insights was that insurance markets rationally could not exist unless coercion was involved or unless there was free choice with significant risk aversion. Excluding coercion, a risk-neutral person will maximize expected wealth and therefore will pay at most the expected loss due to acts of nature. That is, if there is a 0.001 probability that fire will destroy a $200,000 building in a given year, the largest insurance premium a risk-neutral individual would pay is the expected loss, $E(L)$, which equals $200. Insurance provision involves marketing, adjusting, and other monitoring costs denoted by c. Let this total cost of insurance provision be $C = c + E(L)$. No insurance market could exist unless people are willing to pay at least C for insurance. The amount an individual is willing to pay beyond $E(L)$ is called the risk premium, ρ. The risk premium is zero for risk-neutral individuals and positive for those who are risk-averse. Thus, a risk-averse individual is willing and able to purchase insurance if the provision costs are less than the premium, or $c \leq \rho$. The left side of the inequality is the supply price, and the right side is the demand price. No insurance market can exist (for $c > 0$) without compulsion unless market participants are risk-averse such that they are willing to pay for the costs, c. Any risk-averse individual would surely purchase "fair insurance" where the insurance premium is equal to the expected loss. This is a simple initial insight into a necessary condition for the existence of an insurance market.

A second insight comes from the notion of insurability. Insurers generally are thought to have little exposure to risk if they have a large number of independent contracts. In this case, the payouts (indemnities) will thus be remarkably predictable (low variance). This is evident from the law of large numbers in probability. Using this indemnity data, the insurance product can be readily priced and most of the competitive assumptions can ensue. These conditions for insurability hold for life insurance and fire insurance. These conditions rarely hold for acts of nature to agriculture (hail insurance is an exception), which can often be catastrophic. Insurers of acts of nature in the Midwest would have a highly correlated portfolio if the insured losses were due to drought. Thus, the liabilities could be large one year and low the next, implying a high variance of the return. This may mean that the probability of ruin for an insurer would be substantial, leading to risk-responsive behavior by insuring firms (Duncan and Myers 2000). However, there are reinsurance markets and other means to trade away some of the risk in a risky undertaking.

It appears that no multiple-peril private crop insurance markets have emerged (e.g., see Glauber and Collins 2002). Due either to issues of insurability or just plain old rent seeking, policy has focused in recent decades on the provision of federally organized and provided crop insurance. To illustrate the essence of the program, a farm might select the 0.75 option. When yields are 75 percent of approved program yield, this triggers a payment from the government.

In 1980, the "Crop Insurance Improvement Act" was passed in the United States, allowing the private sector to sell multiple-peril federal crop insurance (MPCI) with a subsidized premium. Since that time, five additional acts have been passed to extend and reform the federal program. Federal subsidies have risen to around $1.4 billion. Liabilities have grown sevenfold since 1980 to around $35 billion, also showing the tremendous growth in the program. During the 1980s and much of the 1990s, the ratio of indemnities/total premiums (ignoring the government subsidy) or loss ratio was greater than 1, indicating that the program was actuarially unsound. Because of enormous policy interest in the program, significant amounts of intellectual effort, computer time, and ink were spent studying crop insurance.

Crop insurance has also been a focus of international attention as countries around the world study the viability of similar programs

(both Canada and Japan have programs similar to the United States; see the bibliography of Coble and Knight 2002). Because of the federal subsidy, this insurance is more than "fair" to some farmers. Thus, even a risk-neutral farmer may strictly prefer the insurance. Thus, risk tools are relevant, but insurance purchase isn't *prima facie* evidence of risk aversion, as it would be in a laissez-faire market for insurance.

There are many possible reasons for the excess losses, including that government may wish to transfer wealth to agriculture. However, using the best available actuarial methods, there are good reasons to expect the program to fail. At least part of the answer is well known to economists. When farmers have more information than those setting the rates (asymmetric information), moral hazard and adverse selection may occur. These will be explored conceptually first and then empirically.

Adverse Selection

Suppose that rate makers have access to average actuarial data and set what is known as a pooled rate, that rate where the average loss ratio is 1. Suppose also that there is heterogeneity; that is, some farms have a high probability of loss below the insured level, while other farms have a low probability of loss. Farms that are good risks (low probability of a payout) in that the probability of yields falling below the threshold value is low, will find the price of insurance too high because it is based upon the average farm's probability. They will not purchase the insurance. Farms that are poor risks will find the average rate attractive and will self-select into the insurance program. Thus, risks that are adverse to the long-run viability of the program select in and low risks select out. This implies that the government will lose money and may wish to raise rates. If rates are raised, some of the good-risk farms will exit the program. Again, the program will lose money.

The incentives to purchase insurance are now threefold under risk-aversion: 1) the incentive to participate based upon an increase in expected profit due to the subsidy, 2) an incentive due to risk aversion (reduced risk), and 3) an incentive due adverse selection. One way to calculate the three effects is as follows: the first calculates the increase in expected profit from being insured, the second calculates the difference in the risk premium due to being insured, and the third follows from the increase in expected indemnity due to adverse selection. Each

of these provides incentive to participate. The greater the subsidy, the larger the first incentive. Greater risk aversion implies greater incentive to purchase insurance. Finally, for the adverse selection effect, higher-risk firms will benefit because of larger expected indemnities than typical. The third incentive will imply that the expected loss to insurers increases with participation in insurance by high-risk producers.

Adverse selection need not be a problem if the insurance provider can monitor or know the nature of the heterogeneous firm. Experience rating is an example of trying to adjust premiums for the type of firm demanding insurance.

Moral Hazard

Moral hazard implies another type of asymmetric information. Here, knowledge of the insured's actions is hidden from the insurer when comparing pre- and post-insurance behavior. The most extreme form of hidden action is arson, but more subtle behaviors involve taking inappropriate care or effort. Antitheft devices might not be purchased if a car is fully insured. Regarding health insurance, an insured person might see a doctor more often than if uninsured. For MCPI, the opportunities to change behavior if insured are many. Thus, a fourth incentive to purchase insurance relates to moral hazard: fewer inputs may be applied when insured. This will save costs and will increase the probability of collecting indemnity payments. Again, if the provider can monitor behavior and pay indemnities according to deviations from best practice, moral hazard need not be an issue. Monitoring is expensive and difficult to do, except for obvious behaviors.

Many policy proposals have tried to deal with the moral hazard problem. One such program makes payments based upon area yields rather than individual yields. In this case, adverse selection and moral hazard are virtually eliminated. However, the amount of insurance that an individual receives is dependent on how the farm outcomes are correlated with the area outcomes. If a farm risk is largely idiosyncratic, then the areawide insurance will provide little benefit to the farm.

Empirical Results

Consider first the demand for insurance. Empirical work on crop insurance demand has used simulation methods assuming particular characteristics and risk preferences of the farm (e.g., Kramer and Pope 1982; Mapp and Jeter 1988) or econometric techniques (e.g., Gardner and Kramer 1986; Goodwin 1993; Barnett and Skees 1995; Richards 2000; Vandeveer and Loehman 1994; Coble et al. 1996). A central question is how does the demand for insurance respond to various characteristics of the farm and the contract and insurance premiums? These studies find that the demand for crop insurance is very price (premium) inelastic despite wide variation in crops, regions, subsidies, and in the nature of the program (contract). The 1998 ad hoc disaster relief bill provided for an additional 30 percent of subsidies for premium subsidy. Studying this change, Coble and Barnett (1999) find the price elasticity of demand to be approximately 0.65 in terms of acres insured. That is, a 1 percent decrease in premiums would increase acres insured by 0.65 percent.

Empirical work on moral hazard and adverse selection is much more difficult than measuring insurance demand elasticities. A number of studies find substantial scope for or direct evidence of adverse selection (Goodwin 1994; Ker and McGowan 2000; Luo, Skees, and Marchant 1994; Just, Calvin, and Quiggin 1999). Adverse selection is a large problem in the program for at least three reasons. First, farmers can choose to participate knowing early spring soil moisture and weather forecasts. For example, soil moisture at enrollment and long-run weather forecasts can be beneficial. Using El Niño/La Niña weather patterns can exacerbate the adverse selection problem for insurers (Ker and McGowan 2000). This implies that farmers often have more information than rate makers. Second, there is great heterogeneity, and farmers may choose to insure particular parcels of their land. Third, the U.S. program is marked by procedures that imply large difficulties. For example, a farm without an approved yield history could use the county average. If a farm's yields were substantially lower than this average, there would be a large indemnity paid and relatively small premium received, leading to program losses.

Regarding moral hazard, Horowitz and Lichtenberg (1993) estimated chemical use for Midwest corn producers. They estimated that

insurance participation in MPCI led to increased use of nitrogen, herbicides, and pesticides. Smith and Goodwin (1996) examined Kansas dryland wheat production and obtained opposite results. Firms purchasing insurance significantly reduced total chemical input. Babcock and Hennessy (1996) argued that using reasonable measures of risk aversion and estimates of technology, insurance implied very modest reductions in fertilizer usage. Coble et al. (1996) found evidence of increased yield shortfalls for those insured. Taken as a whole, these results suggest that moral hazard is a potentially serious problem.

There is also research that substantiates that the uninsured behave differently from the insured, but that does not attribute this to a particular explanation. Quiggin, Karagiannis, and Stanton (1993) examined typical revenue and input share equations and noted that revenue was statistically less for insured farms. No corresponding significant results were found for inputs. The impacts on crops grown (likely moral hazard) are substantial and likely clearer. Glauber (1999) estimates that a revenue insurance program for North Dakota durum wheat producers led to a 25 percent increase in production. Wu (1999) estimates that crop insurance for corn causes corn acreage to increase. Keaton, Skees, and Long (1999) estimate that a 10 percent increase in crop insurance participation increased an increased planted area of 6 major crops of 5.9 million acres. This is an unusually large response and likely overestimates the response to crop insurance alone (Glauber and Collins 2002). Goodwin and Vandeveer (2000) estimate a 2.2–3.3 percent increase in corn and soybean acreage planted. Orden (2001) estimates that that would increase production by 0.28–4.1 percent. Finally, if production increases, price must fall. Babcock and Hart (2000) conclude that the elimination of crop insurance subsidies for corn would increase price by $0.02–$0.16 per bushel.

To summarize, it seems that there is every expectation to believe that adverse selection and moral hazard will be a problem in the MPCI. The dates allowed for enrollment, the fact that separate fields can be enrolled, and the difficulties of monitoring complex behavior all contribute to these possibilities. Though the empirical research is not as broad and uniform as desired, the available evidence suggests that the two economic problems identified with provision of insurance under asymmetric information are alive and well in MPCI (Coble and Knight 2002).

As this section concludes, a question arises: Should the government be insuring yields in the first place? First, if the elasticity of demand is unity, yields may vary considerably but total revenue (price times quantity) is fixed. This suggests that if policy wants to provide some safety net for farms rather than transfer wealth to them, then revenue insurance may be a preferred policy to yield insurance. Second, it is far from clear that there is a strong demand for agricultural crop insurance. This is to be distinguished from a strong demand for a subsidy or transfer to farmers.

HEDGING/FORWARD MARKETS

As discussed in virtually every textbook on economic theory or practice, hedging can reduce exposure to risk. Examples abound of markets for risk. Many commodities are listed on the Chicago Board of Trade, and a number of instruments are relevant. Though using futures markets is available, there is no reason to suspect that this is the efficient mechanism to trade risk. Often the efficient mechanism for a farm to shed risk is a forward contract. A forward contract is merely a contract at a negotiated price today for delivery in the future. A futures market is an organized forward market specifying delivery at a particular date, quantity, and grade of the commodity at a specified place (e.g., Chicago). However, the basic advantages and risks of farm hedging can be told equivalently with either a forward or a futures market. I shall use the latter because it is commonly discussed in most texts in microeconomics.

A farmer plants corn in the spring and knows that the futures price is $3.00 for September corn. This is the current price for future delivery of corn in September. If the futures price converges to the actual price of corn on the spot market, then when fall comes, both prices will be equal. These prices might be equal to $4.00 or $2.00. They are random when viewed from the point of view of the farmer in the spring.

Now consider the following three transactions for a bushel of corn assuming the fall price is $\$x$:

	Spot	Futures
Sell (spring)		\$3.00
Sell (fall)	$\$x$	
Buy (fall)		$-\$x$

Summing yields +\$3.00. The farmer, by placing the hedge, has received a certain \$3.00 for corn rather than the random price $\$x$. If risk-averse and if the futures market is fair or unbiased (expected spot is the futures price), farmers would surely prefer to use the futures market. The farmer could forget risk aversion and use the certain \$3.00 price signal to decide how much acreage to plant in corn. Summarizing and generalizing the above example, farm profit using the futures market can be described equivalently as:

(6.6) profit = total revenue in the spot market – costs of production + (futures price – spot price) × (quantity of output hedged),

or equivalently,

(6.7) profit = (unhedged output) × (spot price) – costs of production + (hedged output) × (futures price),

when the futures price converges to the spot price at any point in time.

The example and concepts discussed above bring about four important issues. First, the separation result of production from hedging does not extend to the amount hedged; it depends on the magnitude of risk aversion. However, when production itself is uncertain, the farmer does not know how much of her crop is hedged by a given quantity sold forward in the spring. Second, there is basis risk where basis is the difference between the spot price and the futures price at any point in time and in the place where production takes place. Third, how should a hedge change over time in reaction to new information and what are the time series properties of prices? Dynamic or rolling hedges are an important issue. Finally, there may be substantial trans-

actions costs in using the futures market (fees/margin calls etc.). These issues are reviewed in sequence.

The Hedge

When production is certain, when there is no basis risk, and when firms are risk-averse, one of the first observations involving the optimal hedge is that it will be less (greater) than output produced as the expected spot price is greater (less) than the futures price. This states that in order for a risk-averse firm to be rewarded for risk taking by selling more futures contracts than it has output (speculating), it must be true that the futures price is greater than expected spot price. When the futures or forward price predicts unbiasedly the spot price, output will be completely hedged because there is no incentive to speculate in either the cash or futures markets. Thus, a key question is whether futures or forward prices unbiasedly predict spot prices. The available evidence is mixed. However, across many commodities and countries, my reading of the evidence suggests that when spot prices are not longer than 3–6 months out, futures prices are unbiased estimates of future spot prices.

Production Uncertainty

When production is uncertain, the correlation between production and price uncertainty is crucial to any analysis. For a farmer producing in the corn-belt, this correlation is likely significantly negative. To illustrate why this covariance matters, consider a common description of technology where production shocks enter production multiplicatively. When expected production is expanded, the marginal benefit in terms of expected profit is expected price plus the covariance between the production shock and price. Because we presume this covariance is negative, firms will produce less output because the more output produced, the greater the reduction in profit on average. Further, increasing the scale of production will increase the variance because the variance of profit is proportional to the scale squared. Now we ask how the possibility of a forward or futures contract affects hedging and production choice.

Many authors have used mean-variance notions to calculate the optimal hedge when both production and the price of output are uncertain. Because price and production tend to be negatively correlated, the optimal hedge under risk aversion is generally found to be less than expected farm production.

Dynamics

The optimal dynamic hedge depends crucially on the evolution of prices (time-series properties) and whether they are unbiased. If prices are unbiased and production is certain, then reasonably simple dynamics are implied in the optimal hedge (or ratio of hedge to production) in most cases (Myers and Hanson 1996). When production is uncertain, then strong assumptions are required in order to make much headway on solving the problem (see references in Myers and Hanson).

Use of Futures and Forward Markets

Moschini and Hennessy (2001), citing a report from the U.S. General Accounting Office (1999), state that the available evidence is that farmers use futures markets some but use forward markets frequently. For farmers with sales exceeding $100,000, forward contracts were used by 55 percent of farms and futures contracts or options were used by 32 percent of farms. Patrick, Musser, and Eckman (1998) surveyed large, well-educated, progressive Indiana farmers over a three-year period on their use of forward and futures markets. Those who used some form of forward contracts exceeded 75 percent. Use of futures markets to hedge was limited to less than 25 percent for corn and soybeans and usually was less than 15 percent. My interpretation of the general tone of much agricultural extension work seems to be: “futures are a risk reduction tool that has been under-exploited.” However, it is a very costly and imperfect mechanism for trading risk compared to forward contracts—particularly where a large purchasing entity can use the futures market to “lock in” price and then extend forward contracts to farmers.

FOOD SAFETY

Consumer confidence about food safety has fallen precipitously in recent times (Kramer 1990). This is likely due to highly publicized occurrences in the 1980s and 1990s. The Alar and Chilean grapefruit scares are examples of concern about chemical residues on produce. In 1993, an *Escherichia coli* outbreak in several fast-food restaurants sickened hundreds of people and resulted in four deaths. In the summer of 1997, there was a much-publicized case where 25 million pounds of hamburger were produced with suspected *E. coli* contamination. The Centers for Disease Control and Prevention (CDC) estimate that between 6.5 and 33 million people in the United States become ill each year from food-borne pathogens, and that up to 9,000 die (Buzby et al. 1998). Of these cases of illness, more than 4,000 deaths may be associated with meat and poultry products. In addition, chemical residues from fertilizer, herbicides, and pesticides may pose long-term risks to the public.

Safety policy is concerned with the delivery of existing foods within some level of confidence that it is safe. It also extends to new foodstuffs such as genetically modified organisms. We expect that the usual marginal benefit–marginal cost calculations inform decision making: absent externalities, the optimal level of care or safety is where the marginal private benefit equals marginal private costs. The marginal benefit could be modeled with expected utility or a mean variance utility and the willingness to pay for each additional unit increase in safety. Apparently, however, there are significant externalities to other firms and consumers if a firm chooses a low level of safety. Thus, because the optimal level of safety is where marginal social benefit equals marginal social cost, private incentives as embodied in supply and demand may not lead to the social optimum. Contrary to the rhetoric often heard, this optimum will most often allow for some contamination/risk.

Though measurement of each of these entities is not easy to do well, there have been numerous attempts to shed light on the costs and benefits of a policy proposal. The costs are relatively easy to conceptualize and calculate. These are the additional costs to firms when safety is efficiently increased. For a recent policy change by the Food Safety

and Inspection Service (FSIS) called Hazard Analysis and Critical Control Points (HACCP), the costs are estimated to be at least $100 million annually. Antle (2000), studying meat-processing plants, argues that these estimated costs are much too small due to the loss in productive efficiency involved in complying with the HACCP regulations. Any attempt to measure the costs of a regulation must count both the direct costs of the program and the indirect costs due to the loss in productive efficiency.

Many estimates of the benefits have been much larger: $3.7 billion–$19.1 billion, depending on quantity and type of pathogens ameliorated and assumptions about the value of life. Cutting some corners, the conceptual notion of willingness to pay (WTP) for food safety can be illustrated using Equation (6.2) for a consumer. Let a consumer be given a choice between two probability distributions. The current distribution possessed has a mean of $100 and a variance of a 500. The second distribution has a mean of $96 and a variance of 400. The WTP is the value that equates the following utilities:

$$U(100,500) = U(96 - WTP, 400).$$

(A more realistic depiction would embody not two means or variances but two probability distributions of contamination.) It is the purchase or demand price for the second probability distribution given that the individual possesses the first one. In general, it can be positive or negative. The Food Safety and Inspection Service estimates of the yearly public benefits using the cost of illness method (discussed below) are $990 million–$3.7 billion. This wide range of numbers immediately suggests the difficulty of measuring consumer benefits for the United States.

The four methods used to estimate benefits are: 1) ask people in a survey (contingent valuation, or CV method) how they would value an increase in safety, 2) use experimental auctions to try to evaluate consumer's willingness to pay for improved safety (experimental method), 3) use cost of illness or liability as measure of consumer benefits, and 4) direct econometric estimation of the shift in demand functions controlling for other factors (Caswell 1998; Buzby et al. 1998).

A few introductory comments will serve as background. Attempts to measure econometrically the effects of food safety on consumer

demand are fraught with measurement problems and often cannot apply to a prospective program. If one had measures of food characteristics, including safety attributes, then a regression of price on food characteristics (hedonic regression) could yield the WTP for safety changes. For example, as safety varied, the economist could measure the effect on price. This marginal effect on price could be used to infer WTP. However, one seldom has such data. Yet, it may be possible to measure the impact of information or safety on demand. Further, some economists make a distinction between safety claims by a manufacturer and scientific supportable claims. That is, if a manufacturer labels eggs with a particularly low probability of *Salmonella,* and charges $*x* more for them, is that the correct measure of the social value of improved safety irrespective of scientific evidence of efficacy?

Applications of the Methods

In the first method, surveys elicit a response to a hypothetical environment. For example, one might propose a baseline probability of food poisoning and severity and ask the respondent what they would be willing to pay for a particular scenario of risk/severity reduction. This is conceptually the most direct and appealing method, although there may not be sufficient incentives and context for respondents to be truthful. The second method need not rely on hypothetical scenarios, but the experiment may not be representative of actual decision making by the population at large (a sampling problem), or the experiment itself may not represent the complexity of the environment and choice. The third method often is not necessarily linked to WTP or social value. For example, the cost of illness may not include pain and suffering and may miss the long-term consequences of illness on growth and development. Liability may be a better measure, but it is not very helpful for a prospective evaluation of a policy. Either the cost of illness or producer liability likely underreports the WTP for improved food safety.

The empirical findings are interesting but often do not yet yield a precise and consistent pattern (Shogren et al. 1999). Buzby et al. (1998) discuss the following CV experiment. Store A is a conventional U.S. grocery outlet, but store B eliminates or reduces, through testing, the amount of pesticide residues on fresh produce. Store A is called

pesticide-free and store B is set to government residue standards. Demographic variables and a risk index that the respondent estimates are included in the regressions. The only demographic variable that was statistically significant was gender: women are more likely to shop at store B and have a higher WTP. As expected, those who estimated the risks from residues as being high were also more likely to shop at store B and have a higher WTP. The median weekly WTPs for a government standard store and a pesticide-free store were $5.31 per week and $5.88, respectively. Buzby, Ready, and Skees (1995) used CV to measure the costs and benefits of eliminating a post-harvest chemical sodium *ortho*-phenylphenate (SOPP) from use on Florida grapefruit designated for the fresh markets. Sodium ortho-phenylphenate is a fungicide that reduces molds and rots but is perceived by consumers to have health risks. After calculating the costs (lost fruit) to the industry from the ban, CV is used to calculate the WTP. Average WTP was between $0.19 and $0.28, depending on what one assumed about the WTP to nonrespondents. On average, respondents are willing to pay about 38 percent more for SOPP-free grapefruit. Regression analysis found no significant evidence that household size, race, or gender affected WTP. More affluent and older people were found to have a lower WTP. van Ravenswaay and Hoehn (1991) found that consumers were willing to pay about 17 percent of the current purchase price to avoid Alar in fresh apples.

In a typical experimental market, participants are given a choice between a chicken sandwich with the usual chance of contamination by *Salmonella* (probability of contamination may not be specified) if purchased at a local outlet and a sandwich that is screened and is reported to have 1/1,000,000 chance of contamination. Bids are in increments over the price of the sandwich with the usual risk of contamination. Similar experiments have been done in Arkansas, Massachusetts, Iowa, and California. Incentives are put in place to obtain relevant bids. In Arkansas and Massachusetts, average bids often exceeded $1, but in Iowa, California, and Kansas the average was approximately $0.55 or less for a given run of the experiment. It is unclear how one extrapolates this to a countrywide cost/benefit calculation which includes nonstudent participants.

The essence of the methodological difficulty involving eliciting WTP is found in the excellent experimental study of Shogren et al.

(1999). They designed an experiment where subjects chose irradiated or nonirradiated chicken breasts. The price of nonirradiated chicken breasts was held constant at $2.88/lb. The price of the irradiated chicken breasts varied from a 10 percent discount ($2.59/lb.) to a 20 percent premium ($3.45/lb.). The first experiment involved actual retail market trials with clear labeling and prominent display of USDA summary data on food irradiation. In the second experiment, an experimental auction was conducted. A budget of $30 was offered, and the participants were asked to spend approximately $5.00 and keep the rest. Briefly, after providing each participant with the USDA summary data on irradiation, each participant responded with their preferred choice. The final experiment was a random sample of 400 households where the survey requested information on purchase behavior given the same choices as in the retail and experimental markets. In the latter case, a much more rich set of attitudinal, experience, and demographic data were available to the researchers.

There was general agreement among all three approaches in that the demand for irradiated chicken is downward sloping. However, informing market participants with the best available scientific information (which is generally supportive to higher health and safety with irradiation) led a significant percentage of customers to demand a 10 percent discount on irradiated chicken. Further, in this category (requiring a 10 percent discount), there was a reasonably large (greater than 33 percent) difference among the three methods in the percentage that would purchase the treated chicken. The nature and explanation of these anomalies are part of an ongoing debate (e.g., Bockstael 1999). When it comes to the value of human life and safety, there are many methodological and policy issues (Hooker, Nayga, and Siebert 1999). The concluding question arising from these experiments is: "Based upon available information, is consumer sovereignty to be respected even if tastes and preferences conflict with accepted scientific evidence?"

The last method is based upon secondary data. Henneberry, Piewthongngam, and Qiang (1999) tried to measure a risk information variable and placed it in a system of demands for 14 major fresh produce categories. The risk information variable was seldom statistically significant but suggested an average percentage elasticity of 0.05–0.07 percent due to a marginal decrease in risk information. For example, a

1 percent increase in the risk information index reduced crucifers, carrots, and foliage consumption by an average of 0.07 percent. However, if high-frequency data are used with a specific risk, it appears that one can establish through event studies the impact of contamination on prices. For example, for specific USDA *E. coli* O157:H7 recalls, McKenzie and M.R. Thomsen (2001) established that prices (using daily prices) for boneless beef react significantly to the recalls. This is the most likely category affected by the bacteria. However, no such relationship can be established for the more aggregative categories of live cattle and boxed beef prices.

CONCLUSION

It is a daunting task to try to summarize the content of risk research in agriculture. Large areas of agricultural economic research have been neglected: adoption of technology, storage, grading and standards, contracting, environmental risks, finance, and others. Risk research pervades agricultural economic research because risk is pervasive in agriculture. Biological and physical processes (such as weather) are so complex that risk is often treated as endemic. This is not the only way to view research. Perhaps more investment should be made to understand these biological processes so that deterministic methods can be coherently employed. My conclusion is that risk research in agricultural economics has been a very fruitful intellectual endeavor. However, as is likely apparent throughout this chapter, I am not sure that the profession has invested sufficient attention to carefully measuring behavior. Normative prescriptions to government or individuals are likely to mislead if there is no firm grounding in behavioral social science knowledge. To be sure, there are some risk-related stylized facts such as the econometric response of enterprise choices to changing risk. However, there is much more work to be done in order to understand whether many current interpretations of research results based on aggregate data rests on firm micro-foundations.

References

Allen, D.W., and D.L. Lueck. 1998. "The Nature of the Farm." *Journal of Law and Economics* 41(2): 343–386.

Andrews D.W.K. 2001. "Testing When a Parameter is on the Boundary of the Maintained Hypothesis." *Econometrica* 69(3): 683–734.

Antle, J.M. 1983. "Testing the Stochastic Structure of Production: A Flexible Moment–based Approach." *Journal of Business and Economic Statistics* 1(3): 192–201.

———. 1987. "Econometric Estimation of Producers' Risk Attitudes." *American Journal of Agricultural Economics* 69(3): 509–522.

———. 2000. "No Such Thing as a Free Safe Lunch: The Cost of Food Safety Regulation in the Meat Industry." *American Journal of Agricultural Economics* 82(2): 310–322.

Aradhyula, S., and M. Holt. 1989. "Risk Behavior and Rational Expectations in the U.S. Broiler Market." *American Journal of Agricultural Economics* 71(4): 892–902.

Babcock, B.A., and D.A. Hennessy. 1996. "Input Demand Under Yield and Revenue Insurance." *American Journal of Agricultural Economics* 78(2): 416–427.

Babcock, B., and C. Hart. 2000. "A Second Look at Subsidies and Supply." *Iowa Agricultural Review* 6(4): 3.

Barnett, B.J., and J.R. Skees. 1995. "Region and Crop Specific Models of the Demand for Federal Multiple Peril Crop Insurance." *Journal of Insurance Issues* 19(2): 47–65.

Bar-Shira, Z., Just, R.E., and D. Zilberman. 1997. "Estimation of Framer's Risk Attitude: An Econometric Approach." *Agricultural Economics* 17: (2–3): 211–222.

Behrman, J.R. 1968. *Supply Response in Underdeveloped Agriculture: A Case Study of Four Major Crops in Thailand, 1937–63*. Amsterdam: North-Holland Publishing Co.

Bockstael, N. 1999. "Valuing the Benefits of Microbial Food Safety Risk Reduction: Discussion." *American Journal of Agricultural Economics* 81(5): 1200–1204.

Brorsen, W. B., Chavas, J-P, and W. R. Grant. 1985. "Marketing Margins and Price Uncertainty: The Case of the U. S. Wheat Market." *American Journal of Agricultural Economics* 67(3): 521–528.

Buzby, J.C., R.C. Ready, and J. Skees. 1995. "Contingent Valuation in Food Policy Analysis: A Case Study of a Pesticide-Residue Risk Reduction." *Journal of Agricultural and Applied Economics* 27(2): 613–625.

Buzby, J.C., J.A. Fox, R.C. Ready, and S.R. Crutchfield. 1998. "Measuring Consumer Benefits of Food Safety Risk Reduction." *Journal of Agricultural and Applied Economics* 30(1): 69–82.

Carlson, G.A. 1979. "Insurance, Information, and Organizational Options in Pest Management." *Ann. Rev. Phytopath.* 17: 149–161.

Castle, E. 1954. "Flexibility and Diversification as a Measure of Meeting Price and Yield Uncertainty in Western Kansas." *JFE* 36: 273–284.

Caswell, J.A., ed. 1995. *Valuing Food Safety and Nutrition.* Boulder: Westview Press.

———. 1998. "Valuing the Benefits and Costs of Improved Food Safety and Nutrition." *The Australian Journal of Agricultural and Resource Economics* 42(4): 409–424.

Chavas, J.P., and M.T. Holt. 1996. "Economic Behavior under Uncertainty: A Joint Analysis of Risk Preferences and Technology." *Review of Economics and Statistics* 78(2): 329–335.

Coble, K.H., and B.J. Barnett. 1999. "Preliminary Evidence of Crop Insurance Demand Elasticity from the 1999 Response to Additional Subsidies." Department of Agricultural Economics, Mississippi State University. Available at: <http://www.agecon.msstate.edu/risk/documents>.

Coble, K.H., and T.O. Knight. 2002. "Crop Insurance as a Tool for Price and Yield Risk Management." In *A Comprehensive Assessment of the Role of Risk in U.S. Agriculture,* R.E. Just and R.D. Pope, eds. Boston: Kluwer Academic Publisher, pp. 445–468.

Coble, K.H., T.O. Knight, R.D. Pope, and J. Williams. 1996. "Modeling Farm-Level Crop Insurance Demand with Panel Data." *American Journal of Agricultural Economics* 78(2): 439–447.

Coyle, B. 1999. "Risk Aversion and Yield Uncertainty in Duality Models of Production: A Mean-Variance Approach." *American Journal of Agricultural Economics* 81(3): 553–567.

Davis, S.B., T.J. Price, M. Wetzstein, and M.W. Rieger. 1997. "Reducing Yield Variation in Peach Orchards by Geographic Scattering." *American Journal of Agricultural Economics* 79(4): 1119–1126.

Duffy, P. A., Shalishali, K. and H. W. Kinnucan. 1994. "Acreage Response under Farm Programs for Major Southeastern Field Crops." *Journal of Agricultural and Applied Economics* 26(2): 367–378.

Duncan, J. and R.J. Myers. 2000. "Crop Insurance under Catastrophic Risk." *American Journal of Agricultural Economics* 82(4): 842–855.

Dunn J.W., and J.R. Williams. 2000. "Farm Characteristics that Influence Net Farm Income Variability and Losses" Paper presented at the Western Agricultural Economics Association Annual Meetings, Vancouver, British Columbia, June 29–July1.

Estes, E.A., Blakeslee, L.L., and R.C. Mittelhammer. 1981. "On Variances of Conditional Linear Leasto-Squares Search Parameter Estimates." *American Journal of Agricultural Economics* (February): 141–145.

Gardner, B.L., and R.A. Kramer. 1986. "Experience with Crop Insurance Programs in the United States." In *Crop Insurance for Agricultural Development: Issues and Experience*, P. Hazell, C. Pomareda and A. Valdes, eds. Baltimore: John Hopkins Press, pp. 195–222.

Glauber, J. W. 1999. "Declaration in the U.S. District Court for the District of North Dakota, Southeastern Division." Testimony given in Paul Wiley et al. Daniel Glickman, Divil No. A3-99-32 (May 18).

Glauber J.W., and K.J. Collins. 2002. "Risk Management and the Role of the Federal Government." In *A Comprehensive Assessment of the Role of Risk in U.S. Agriculture*, R.E. Just and R.D. Pope, eds. Boston: Kluwer Academic Publisher, pp. 469–488.

Goodwin, B.K. 1993. "An Empirical Analysis of the Demand for Crop Insurance." *American Journal of Agricultural Economics* 75(2): 425–434.

———. 1994. "Premium Rate Determination in the Federal Crop Insurance Program: What Do Averages Have to Say About Risk?" *Journal of Agricultural and Resource Economics* 19(2): 382–395.

Goodwin, B.K., and M. Vandeveer. 2000. "An Empirical Analysis of Acreage Distortions and Participation in the Federal Crop Insurance Program." Paper presented at the Economic Research Service workshop, "Crop Insurance, Land Use, and the Environment," held in Washington, DC, September 12–20.

Griffiths W.E., and J.R. Anderson. 1982. "Using Time-Series and Cross-Section Data to Estimate a Production Function with Positive and Negative Marginal Risks." *Journal of the American Statistical Association* 77(September): 529–536.

Hall, D.C., and L.J. Moffitt. 1985. "Application of the Economic Threshold for Interseasonal Pest Control." *Western Journal of Agricultural Economics* 10(December): 223–229.

Heady, E.O. 1952. "Diversification in Resource Allocation and Minimization of Income Variability." *Journal of Farm Economics* 34: 482–496.

Henneberry, S.R., K. Piewthongngam, and H. Qiang. 1999. "Consumer Food Safety Concerns and Fresh Produce Consumption." *Journal of Agricultural and Resource Economics* 24: 98–113.

Holt, M.T. 1994. "Price-Band Stabilization Programs and Risk: An Application to the U.S. Corn Market." *Journal of Agricultural and Resource Economics* 19(2): 239–254.

Holt, M.T., and S.V. Aradhyula. 1990. "Price Risk in Supply Equations: An Application of GARCH Time-Series Models to the U.S. Broiler Market." *Southern Economic Journal* 57(1): 230–242.

———. 1998. "Endogenous Risk in Rational-Expectations Commodity Models: A Multivariate Generalized ARCH-M Approach." *Journal of Empirical Finance* 5: 99–129.

Holt, M.T., and G. Moschini. 1992. "Alternative Measures of Risk in Commodity Supply Models: An Analysis of Sow Farrowing Decisions in the United States." *Journal of Agricultural and Resource Economics* 17(1): 1–12.

Hooker, N.H., R.M. Nayga, Jr., and J. Siebert. 1999. "Preserving and Communicating Food Safety Gains." *American Journal of Agricultural Economics* 81(5): 1102–1106.

Horowitz, J.K., and E. Lichtenberg. 1993. "Insurance, Moral Hazard, and Chemical Use in Agriculture." *American Journal of Agricultural Economics* 75(5): 926–935.

———. 1994. "Risk-Reducing and Risk-Increasing Effects of Pesticides." *Journal of Agricultural Economics* 45(1): 82–89.

Hurd, B.H. 1994. "Yield Response and Production Risk: An Analysis of Integrated Pest Management in Cotton." *Journal of Agricultural and Resource Economics* 19(2): 313–326.

Hurt, C., and P. Carcia. 1982. "The Impact of Price Risk on Sow Farrowing, 1967–78." *American Journal of Agriculture Economics* 64(3): 565–568.

Jensen, C.W. 1961. "Spatial Diversification of the Dryland Farm Unit." *Management Strategies in Great Plains Farming*, Nebraska Agricultural Experiment Station, MP7:91–95.

Johnson, S.R. 1967. "Reexamination of the Farm Diversification Problem." *Journal of Farm Economics* 49: 610–621.

Just, R.E. 1974. "An Investigation of the Importance of Risk in Farmers' Decisions." *American Journal of Agricultural Economics* 56: 14–25.

Just, R.E., L. Calvin, and J. Quiggin. 1999. "Adverse Selection in U.S. Crop Insurance: Actuarial and Asymmetric Information Incentives." *American Journal of Agricultural Economics* 80(4): 834–849.

Just, R.E., and R.D. Pope. 1979. "Production Function Estimation and Related Risk Considerations." *American Journal of Agricultural Economics* 61: 276–283.

———. 2001. "The Agricultural Producer: Theory and Statistical Measurement." In *Handbook of Agricultural Economics*, B.L. Gardner and G.C. Rausser, eds. Amsterdam: Elsevier, pp. 629–741.

Just, R.E., and R.D. Pope, eds. 2002. *A Comprehensive Assessment of the Role of Risk in U. S. Agriculture.* Boston: Kluwer Academic Publishers.

Keaton, K., J. Skees, and J. Long. 1999. "The Potential Influence of Risk Management Programs on Cropping Decisions." Paper presented at the annual meeting of The American Agricultural Economics Association, held in Nashville, TN, August 8–11.

Ker, A.P., and P. McGowan. 2000. "Weather-Based Adverse Selection and the U.S. Crop Insurance Program: The Private Insurance Company Perspective." *Journal of Agricultural and Resource Economics* 25(2): 386–410.

Kramer, C.S. 1990. "Food Safety: The Consumer Side of the Environmental Issue." *Southern Journal of Agricultural Economics* 22(1): 33–40.

Kramer, R., and R.D. Pope. 1982. "Crop Insurance for Managing Risks." *The Journal of the American Society of Farm Managers and Rural Appraisers* 46(1): 34–40.

Krause, M.A., and W.W. Koo. 1996. "Acreage Responses to Expected Revenues and Price Risk for Minor Oilseeds and Program Crops in the Northern Plains." *Journal of Agricultural and Resource Economics* 21(2): 309–324.

Krause, M.A., Lee, J-H, and W.W. Koo. 1995. "Program and Nonprogram Wheat Acreage Responses to Prices and Risk." *Journal of Agricultural and Resource Economics* 20(1): 96–107.

Lin, W. 1977. "Measuring Aggregate Supply Response Under Instability." *American Journal of Agricultural Economics* (December): 903–907.

Lin, W., G.W. Dean, and C.V. Moore. 1974. "An Empirical Test of Utility vs. Profit Maximum in Agricultural Production." *American Journal of Agricultural Economics* 56: 497–508.

Love, A., and S. Buccola. 1991. "Joint Risk Preference-Technology Estimation in a Primal System." *American Journal of Agricultural Economics* 73: 765–774.

Luo, H., J.R. Skees, and M.A. Marchant. 1994. "Weather Information and the Potential for Inter-temporal Adverse Selection." *Review of Agricultural Economics* 16: 441–451.

Mapp, H.P., and K.L. Jeter. 1988. "Potential Impact of Participation in Commodity Programs and Multiple Peril Crop Insurance on a Southwest Oklahoma Farm." In *Multiple Peril Crop Insurance: AS Collection of Empirical Studies.* Southern Cooperative Series Bulletin No. 34, Washington, DC.

Marshall, A. [1890] 1920. *Principles of Economics*. 8th ed. London: Macmillan and Co., Ltd.

McCloskey, D. 1976. "The Open Fields as Behavior Towards Risk." *Research in Economic History: An Annual Compilation of Research*. Vol. 1. Paul Uselding, ed. Greenwich: JAI Press.

McKenzie, A.M., and M.R. Thomsen. 2001. "The Effect of E. Coli O157:H7 on Beef Prices." *Journal of Agricultural and Resource Economics* 26(2): 431–444.

Meyer, J. 1987. "Two-Moment Decision Models and Expected Utility Maximization." *American Economic Review* 77(3): 421–430.

Mishra, A.K., and B.K. Goodwin. 1997. "Farm Income Variability and the Supply of Off-Farm Labor." *American Journal of Agricultural Economics* 79(3): 880–887.

Mishra, A.K., and M.J. Morehart. 2001. "Off-Farm Investment of Farm Households: A Logit Analysis." *Agricultural Finance Review* 61: 87–101.

Moschini G., and D.A. Hennessy. 2001. "Uncertainty, Risk Aversion, and Risk Management for Agricultural Producers." In *Handbook of Agricultural Economics, Vol. 1A*, B.L. Gardner and G.C. Rausser, eds. Amsterdam: North-Holland, pp. 87–153.

Myers, R.J. 1989. "Econometric Testing for Risk Averse Behaviour in Agriculture." *Applied Economics* 21(4): 541–552.

Myers, R.J. and S.D. Hanson. 1996. "Optimal Dynamic Hedging in Unbiased Futures Markets." *American Journal of Agricultural Economics* 78(1): 13–20.

Nelson, R.G., and D.A. Bessler. 1989. "Subjective Probabilities Elicited Under Proper and Improper Scoring Rules: A Laboratory Test of Predicted Responses." *American Journal of Agricultural Economics* 71(2): 363–369.

Officer, R.R., and A.N. Halter. 1968. "Utility Analysis in a Practical Setting." *American Journal of Agricultural Economics* 50: 257–277.

Orden, D. 2001. "Should There Be a Federal Income Safety Net?" Paper presented at the Agricultural Outlook Forum 2001, held in Washington, DC, February 22.

Patrick, G.F., W.N. Musser, and D.T. Eckman. 1998. "Forward Marketing Practices and attitudes of Large-Scale Midwestern Grain Producers." *Review of Agricultural Economics* 20: 38–53.

Pope, R.D. 1976. "An Analysis of Factors Affecting Farm Diversification: With Special Reference to San Joaquin Valley Crop Farms." Unpublished Ph.D. Dissertation. Berkeley: Department of Agricultural Economics, University of California.

Pope, R., and R. Prescott. 1981. "Supply Response and the Dispersion of Expectations." *American Journal of Agricultural Economics* 63(February): 161–163.

———. 1980. "Diversification, in Relation to Farm Size and Other Socioeconomic Characteristics." *American Journal of Agricultural Economics* 62(August): 554–559.

Pope, R., and R. Just. 1991. "On Testing the Structure of Risk Preferences in Agricultural Supply Analysis." *American Journal of Agricultural Economics* 73(3): 743–748.

Quiggin, J., G. Karagiannis and J. Stanton. 1993. "Crop Insurance and Crop Production: An Empirical Study of Moral Hazard and Adverse Selection." *Australian Journal of Agricultural Economics* 37(2): 95–113.

Regev, U., N. Gotsch, and P. Rieder. 1997. "Are Fungicides, Nitrogen and Plant Growth Regulators Risk-Reducing? Empirical Evidence From Swiss Wheat Production." *Journal of Agricultural Economics* 48(2): 167–178.

Richards, T.J. 2000. "A Two-Stage Model for the Demand for Specialty Crop Insurance." *Journal of Agricultural and Resource Economics* 25(1): 177–194.

Robison, L.J., and P.J. Barry. 1987. *The Competitive Firm's Response to Risk.* New York: Macmillan.

Roumasset, J.A. 1976. *Rice and Risk.* Amsterdam: North-Holland Publishing Company.

Saha, A., C.R. Shumway, and H. Talpaz. 1994. "Joint Estimation of Risk Preference Structure and Technology Using the Expo-Power Utility." *American Journal of Agricultural Economics* 76(2): 173–184.

Samuelson, P.A. 1967. "A General Proof that Diversification Pays." *Journal of Financial and Quantitative Analysis* 2(1): 1–13.

Schultz, T.W. 1980. "Nobel Lecture: The Economics of Being Poor." *Journal of Political Economy* 88(4): 639–651.

Shogren, J.F., J.A. Fox, D.J. Hayes, and J. Roosen. 1999. "Observed Choices for Food Safety in Retail, Survey, and Auction Markets." *American Journal of Agricultural Economics* 81(5): 1192–1199.

Smith, V.H., and B.K. Goodwin. 1996. "Crop Insurance, Moral Hazard, and Agricultural Chemical Use." *American Journal of Agricultural Economics* 78(2): 428–438.

Thompson, G.D., and P.N. Wilson. 1994. "Common Property as an Institutional Response to Environmental Variablility." *Contempory Economic Policy* 12(3): 10–21.

Tronstad, R., and T.J. McNeill. 1989. "Asymmetric Price Risk: An Econometric Analysis of Aggregate Sow Farrowings." *American Journal of Agricultural Economics* 71(3): 630–637.

U.S. General Accounting Office (USGAO). 1999. Agriculture in Transition—Farmers' Use of Risk Management Strategies. GAO/RCED-99-90. Washington, DC: U.S. General Accounting Office.

van Ravenswaay, E.O., and J.P. Hoehn. 1991. "The Impact of Health Risk on Food Demand: A Case Study of Alar and Apples." In *Economics of Food Safety*, J.A. Caswell, ed. New York: Elsevier Science Publishing Co., Inc., pp. 155–174.

Vandeveer, M.L., and E.T. Loehman. 1994. "Farmer Response to Modified Crop Insurance: A Case of Corn in Indiana." *American Journal of Agricultural Economics* 76(1): 128–140.

von Massow, M., and A. Weersink. 1993. "Acreage Response to Government Stabilization Programs in Ontario." *Canadian Journal of Agriculture Economics* 41: 13–26.

Wu, J. 1999. "Crop Insurance, Acreage Decisions, and Nonpoint-Source Pollution." *American Journal of Agricultural Economics* 81(2): 305–320.

Young D., W. Lin, R.D. Pope, L. Robison, and R. Selley. 1979. "Risk Preferences of Agricultural Producers: Their Measurement and Use." Paper presented at seminar of Western Regional Research Project, held in San Francisco, January 25–26.

Zenger, S., and B.W. Schurle. 1981. "The Impact of Diversification on Risk at the Farm Level." Abstract. *American Journal of Agricultural Economics* 63: 1043.

Index

The italic letters *f*, *n*, and *t* following a page number indicate that the subject information of the heading is within a figure, note, or table, respectively, on that page.

About the Institute

The W.E. Upjohn Institute for Employment Research is a nonprofit research organization devoted to finding and promoting solutions to employment-related problems at the national, state, and local levels. It is an activity of the W.E. Upjohn Unemployment Trustee Corporation, which was established in 1932 to administer a fund set aside by the late Dr. W.E. Upjohn, founder of The Upjohn Company, to seek ways to counteract the loss of employment income during economic downturns.

The Institute is funded largely by income from the W.E. Upjohn Unemployment Trust, supplemented by outside grants, contracts, and sales of publications. Activities of the Institute comprise the following elements: 1) a research program conducted by a resident staff of professional social scientists; 2) a competitive grant program, which expands and complements the internal research program by providing financial support to researchers outside the Institute; 3) a publications program, which provides the major vehicle for disseminating the research of staff and grantees, as well as other selected works in the field; and 4) an Employment Management Services division, which manages most of the publicly funded employment and training programs in the local area.

The broad objectives of the Institute's research, grant, and publication programs are to 1) promote scholarship and experimentation on issues of public and private employment and unemployment policy, and 2) make knowledge and scholarship relevant and useful to policymakers in their pursuit of solutions to employment and unemployment problems.

Current areas of concentration for these programs include causes, consequences, and measures to alleviate unemployment; social insurance and income maintenance programs; compensation; workforce quality; work arrangements; family labor issues; labor-management relations; and regional economic development and local labor markets.