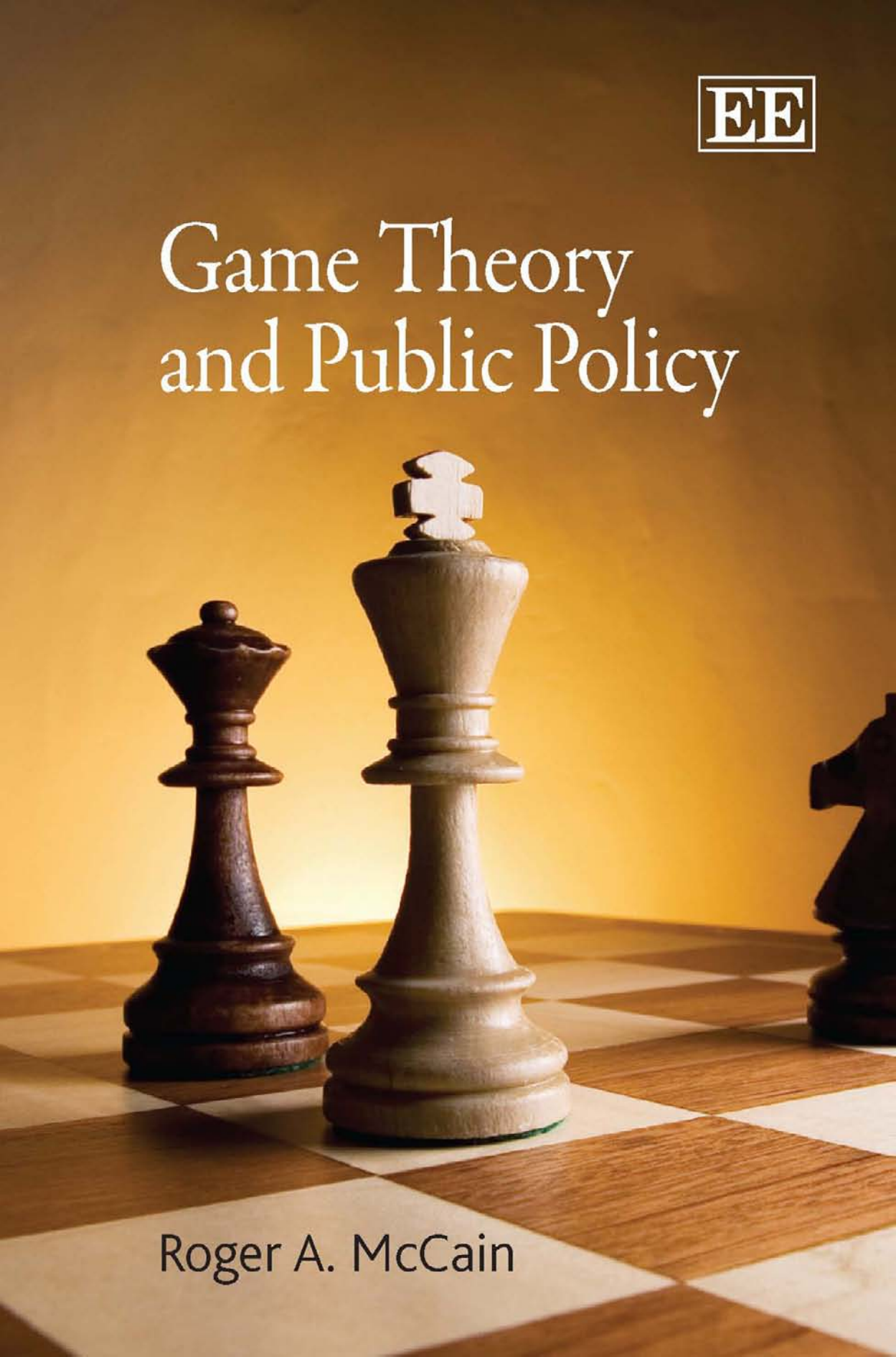
EE
Game Theory and Public Policy
Roger A. McCain

# Game Theory and Public Policy

# Game Theory and Public Policy

Roger A. McCain
*Drexel University, USA*

**Edward Elgar**
Cheltenham, UK • Northampton, MA, USA

Published by
Edward Elgar Publishing Limited
The Lypiatts
15 Lansdown Road
Cheltenham
Glos GL50 2JA
UK

Edward Elgar Publishing, Inc.
William Pratt House
9 Dewey Court
Northampton
Massachusetts 01060
USA

A catalogue record for this book
is available from the British Library

Library of Congress Control Number:


ISBN 978 1 84720 827 9

Printed and bound by MPG Books Group, UK

# Contents

# PART I

# Historical and critical survey

# 1. Objectives and scope of the book

In recent years game theory has become more prominent as an aspect of research and applications in public policy disciplines such as economics, philosophy, management, and political science, and in work within public policy itself. One reason for this growing prominence may be understood from some comments of Thomas Schelling (1960) and Robert Aumann (for example, 2004). They have said that the subject matter of game theory would be better described as interactive decision theory. Schelling and Aumann shared the Nobel Memorial prize in 2005 for their work in game theory, and Aumann was the first president of the world Game Theory Society.

Why then use the term "game theory" for a field that is not really about games? The *game* is to *game theory* as the *experiment* is to *experimental science*. After all, experimental science is not about experiments. It is about the natural world. Nevertheless experiments are a powerful aid to our understanding of the natural world. Similarly, when we conceive interactive decisions as games, we have a powerful aid to understanding them (and among other things, to the design of experiments).

Game theory is, as Aumann says, an interdisciplinary field. "There are very few subjects that have such a broad, interdisciplinary sweep. Let me just put over here some of the ordinary disciplines that are involved in game theory. We have mathematics, computer science, economics, biology, (national) political science, international relations, social psychology, management, business, accounting, law, philosophy, statistics. Even literary criticism . . . We have sports . . . (Aumann, 2003, p.4). Of course, none of these disciplines depends on game theory for its existence. Nevertheless, game theory can be set apart as an attempt to understand collective human activity as the outcome of interactive decisions. On the one hand, this is a remarkably ambitious venture. On the other hand, to the extent that it is successful, it must surely be a crucial foundation for the study of public policy.

The objective of this book is to survey and advance our understanding of game theory as a tool of public policy analysis. The hope is to advance that understanding less by the statement and proof of broad theorems (although the value of such proofs is not to be minimized, and will play some role) as by the clarification and critical assessment of the theorems

we have, and by multiplication of examples and survey and extension of specific cases of application. In practice, the influence of game theory on public policy and the related disciplines has been less a consequence of broad theorems than of insightful examples. Accordingly, it is hoped that a critical reconsideration of some of those examples, and discussion of some less-known ones, will contribute to the study and ultimately the practice of public policy.

Public policy is a pragmatic field. The pragmatic perspective leads to a view of public policy as an outcome of a *process*, and public policy *analysis* is often carried on in terms of the *public policy process*. We might sketch the public policy process roughly as follows: (1) A *problem* is identified which seems to call for public initiative as a solution. (2) Alternative solutions are proposed. (3) Solutions are evaluated, and to the extent possible, the most promising solution specified. At this point the process may be abandoned, if it is found that the best solution does not require public initiatives. We should note, too, that different individuals with different values or interests may regard different proposals as best, and this is the stuff of which politics is made. From this point we suppose that one particular political perspective has been adopted, and the proposal is considered best from that particular perspective. (4) The proposal is advocated and public support for it sought, in the course of which new interest groups and organizations may come into being. (5) The proposal is brought before the legislative or executive branch of government at an appropriate level. (6) The proposal is enacted with or without modification. (7) The proposal is implemented. (8) Experience with the program as implemented leads to feedback from those affected. (9) The cycle begins again with proposals for improvement, replacement, or abandonment of the policy.

How will game theory fit into this outline? It is widely understood today that there are two great branches of game theory, the non-cooperative and the cooperative branch. Of the two, non-cooperative game theory has been the more influential, especially in the last quarter of the twentieth century. This is often treated as an institutional difference: cooperative game theory is applicable when agreements are enforceable, while non-cooperative game theory is applicable otherwise. This book will argue that, on the contrary, the two branches of game theory reflect different conceptions of rationality. Moreover, neither conception is altogether satisfactory. The book argues that non-cooperative game theory is effective as a problem-finding or diagnostic method – non-cooperative behavior is common enough so that a social arrangement that is unstable in the face of non-cooperative behavior will probably fail. However, solutions based on non-cooperative game theory may be unstable in the face of cooperative or collusive behavior, and cooperative behavior is common enough

that such solutions will themselves often fail. Thus non-cooperative game theory is far less effective as a prescriptive tool for public policy.

This should be qualified in the following way, however. There is also some research that combines cooperative and non-cooperative game theory, and one particular branch is sometimes called implementation theory or social mechanism design.[1] If game theory is interactive decision theory, we may think of the outcome of the interaction as being jointly determined by the decisions and the "rules of the game." In social mechanism design, a particular goal for action is specified, and the objective is to find "rules of the game" that will make the goal the non-cooperative solution of the game. In the context of social mechanism design, non-cooperative game theory may also be useful at the second and third stages, proposal and evaluation of new policies. There have been some successes in this way, but also some failures, with both occurring in particular in the design of public auctions of electromagnetic spectrum for telecommunications.

In game theory a state that meets certain conditions, such as stability in some specific sense, may be a candidate *solution* of the game. The word solution is meant in a mathematical rather than a pragmatic sense, here. An array of decisions that is stable in the sense that no one can improve his outcome by changing his strategy unilaterally (while others continue their strategy decisions unchanged) is a Nash equilibrium, and the Nash equilibrium is probably the best known and most widely applied concept of non-cooperative solution.

In cooperative game theory, binding agreements to choose a common decision or a joint strategy are considered to be possible. A group that makes such an agreement is said to form a "coalition." The word "coalition" is best known in its political usage, as a group of political parties in a parliamentary government who join together to form a majority and govern jointly. In cooperative game theory, the word has been generalized to refer to any group of players in a "game" who join together to choose their strategies jointly. Most games with more than two players, applicable to problems of public policy, will provide cases in which individual actors could benefit by forming coalitions with binding agreements to choose a joint strategy. Indeed, as Maskin (2004) points out, we live our lives in coalitions. Thus an account of social life (and especially of public policy) that ignores cooperative game theory must be incomplete.

In any case, the formation of coalitions will be crucial at stages 4–6 of the public policy process as sketched above. Coalitions are likely to be important at other stages as well. Non-cooperative game theory can fail because it assumes that people act non-cooperatively when in fact they can and do form coalitions, such as bidding cartels in auctions. Therefore cooperative game theory may be essential at stages 2 and 3 as well. We

acknowledged that stage 3, in particular, would be dependent on values and interests that might differ. Even when that is so, there may be scope for the differences to be accommodated and the distinct interests and values to be advanced jointly. That, too, is the stuff that politics is made of, and it is also the subject of cooperative game theory. We cannot avoid the conclusion that cooperative game theory is essential for a complete understanding of public policy.

This presents a number of difficulties. First, there are several concepts of solution in cooperative game theory. Which (if any) will be most helpful for our purposes? Second, much of the literature relies on powerful simplifying assumptions. Such simplifying assumptions permit the statement and proof of broad and powerful mathematical theorems, but at the same time they indicate the limits of the applicability of the theorems. Together, these simplifying assumptions mean that most cooperative game theory is not applicable to very many problems of public policy. To be specific,

(1) Expressing the game in the simplified coalition function form means that it cannot be applied to any case in which there are *externalities* and consequent *inefficiencies.*
(2) The common assumption of superadditivity means that if agents are rational, the grand coalition will always form and will efficiently determine the strategies of every agent. This means it is simply not applicable to any case in which excessive centralization may cause problems.
(3) The world we observe, the world relevant to public policy, seems to be one in which many coalitions form and often act independently and indeed competitively with one another. In cooperative game theory such an array of distinct coalitions is called a "coalition structure" (Aumann and Dreze, 1974). We would like a theory that would give us some insight as to just what coalition structures would be likely to form, and why, and game theory based on coalition functions and superadditivity is not helpful with that. This is the problem of endogenous formation of coalitions (Carraro, 2003).

There are approaches to cooperative game theory (as we will discuss in the next chapter) that allow both for externalities and coalition structures, but these approaches are "mathematically intractable." That is, they probably do not have very general solutions, and if they do, the solutions are very hard to find and only tentative progress has been made in this direction. We might nevertheless find solutions for particular cases, and even develop a tool-kit for seeking such special-case solutions.

The objective of the book, then, will be a critical review of some major topics from both cooperative and non-cooperative game theory, including

some less known ideas in non-cooperative game theory, and some constructive proposals for new approaches, to assemble a tool-kit for the analysis of public policy, with the pragmatic purpose of identifying problems and exploring potential solutions. At the same time, we may find resources for a clearer understanding of the public policy enterprise itself.

## NOTE

1. The 2007 Nobel Memorial Prize honored contributions of this sort.

# 2. Representing games

The first step in any application of game theory, whether to public policy or for any other purpose, is to represent the real-world phenomenon of interest as a problem of interactive decision, that is, a "game." This chapter will set out some forms for representation of games that will be important for the remainder of the book. Some will be familiar, even pedestrian, to the reader who is well grounded in game theory. Nevertheless some topics may be important for the game theorist, if only for differences of stress. Contingent strategies are well known, but this book will often make them more explicit and formal than they sometimes are in the game theory literature. Nested games may be a novel topic to the game theorist, as the concept comes from applications in political science, and are crucial to the distinction of a private from a public sector. "Imperfect recall" is very little mentioned in recent game theory, and needs to be discussed in the context of cooperative game theory. Finally, the partition function approach in cooperative games, and its importance for the concept of externality, may be novel to some game theorists. These are important concepts for public policy. Nevertheless, the chapter is expository, with nothing new to the literature except specific examples, some terminology, emphasis, and expression.

## 2.1 GENERAL CONSIDERATIONS

Game theory is a (mathematically) formal study, with deep roots in mathematical set theory. The language of set theory is designed for generality even at the expense of intuition and common sense. For example, in set theory we routinely speak of a set without any members, the "null set," or a set with only one member, or a set consisting of all of the members of some population. These usages may seem strange from the point of view of ordinary English. Generally words come to us with connotations as well as formal definitions, and a word like "set" tends to connote a plural grouping within some larger grouping. Thus the idea of a set without members may seem silly, and the natural impulse of the reader is to make the charitable assumption that the author is not silly, so that something else must be meant. In the present context, this charitable impulse is likely to cause

confusion. Assume instead that I am silly. Because of these conventions, though, some of the most important and productive forms of representation are distant from intuition.

Since a game is an interactive decision problem, our representation must include at least a set of decision-makers, some alternatives among which they must decide, and some objectives to be advanced by the decision. Let us call the set of decision-makers N and denote the members of the set of decision-makers $i = 1, 2, \ldots n$. The set of decision-makers is non-empty; that is, it has at least one member. We will usually refer to the members of this set as "players" or "agents." We will sometimes talk about "games" with only one player, although in that case there is no interaction. The objectives for different players will usually be different, and may be conflicting. For now, we will simply represent those objectives as numbers, and think of the numbers as money payoffs from the "game."

Here is an example of interactive decision theory. (We will call it Game 2.1, the Water Game.) Eastland and Westria share the valley of Southflowing River, which forms the boundary between them. Each country controls some of the northern tributaries of the river, and could divert water from the tributary streams for their own use. However, any diversion from the tributaries of the river will divert water that the citizens of the southern regions of both countries use for irrigation and other purposes, and if both countries divert the water of the tributaries, the flow in the south will be so reduced that silting and problems of navigation will also occur. Reliable cost–benefit studies have provided the following figures: if just one country diverts water from the tributaries, the net benefit to that country will be 3 billion euros, but the other country will lose 4 billion. However, if both countries divert water from the tributaries, each country will suffer a net loss of 2 billion. The two countries do not trust one another and keep their decisions strictly secret from one another as long as possible, so each country can only conjecture as to what the other country will decide and sees no chance of influencing the decision of the other. In this example, the players are the two countries, and the alternatives are different ways of obtaining water for each country's needs. The decisions are to divert water from the tributaries or not. The cost–benefit studies establish that the decisions are interactive: that is, each country's net benefits or losses depend on the other country's decision as well as their own. This example illustrates the simplest class of nontrivial games, two-by-two games; that is, two-player by two-strategy games.

## 2.2 THE GAME IN EXTENSIVE FORM

The most intuitive way to represent a complicated decision problem is as a tree diagram, in which each decision is represented by a branch in the tree. In our example, the two countries make their decisions more or less simultaneously, each in secrecy. This lack of information should also be represented in the tree diagram. That's a complication we shall leave for a little later. First consider the following, somewhat simpler example: Game 2.2, the Entry Game.

One of the most important of simple games, both for theory and for applications in economics and public policy, is the game of market entry. Firm A is an established monopolist, and Firm B is a firm considering entry into competition with Firm A. Firm B has two choices: it can enter or not. Firm A then has two choices: it can retaliate against the entrant by means of a price war, or it can accommodate the new firm by maintaining a price that will be profitable to both of them. Either way, Firm A will face lower profits, but the price war results in even lower profits than the strategy of accommodation. Game 2.2 illustrates this case with payoffs on a scale of 5, and the first payoff to the entering firm, Firm B, and the second payoff to Firm A. (The reader may add as many zeros as seems realistic.)

Figure 2.1 shows this game in the form of a tree diagram, reading from left (the root) to right (the branches). The first number at the tip of each branch is the payoff to Firm B, and the second to Firm A.

Figure 2.2 represents the Water Game in the tree diagram form conventional in game theory. The first payoff is to Eastland and the second to Westria. We see that Westria's decision is enclosed in a larger lozenge

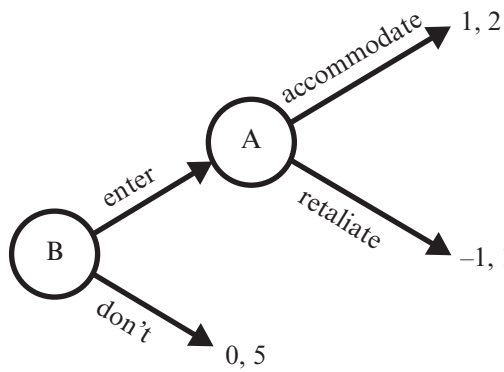


*Figure 2.1 Game 2.2: the Entry Game*

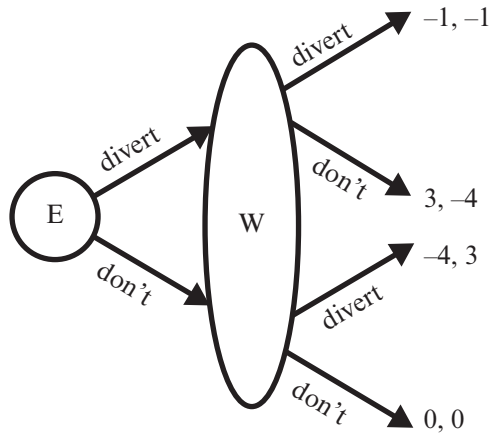


*Figure 2.2 Game 2.1: the Water Game in extensive form*

that includes both of the branches that come from Eastland's decision. This lozenge is called an "information set,"[1] and it encodes the fact that Westria does not know at which branch it is (which decision Eastland has made) when it makes its decision. Conversely, the node labeled "E" in Figure 2.2 is a *full information node*, as are "A" and "B" in Figure 2.1. At a full information node the player is aware of commitments already made by the other players, if any. For a game like the Water Game, in which the decisions are made simultaneously, either decision-maker can be taken first, provided that the information available to each decision-maker is accurately represented.

When games are represented as tree diagrams with information sets to indicate decisions made under ignorance, the game is said to be represented *in extensive form*. The extensive form is highly intuitive, but is not the representation most often used in game theory.

## 2.3 THE GAME IN STRATEGIC NORMAL FORM

Continuing to use the example of the Water Game, we may instead represent it in tabular form. Consider Table 2.1. By choosing to divert or not, Eastland determines whether the outcome will be the payoffs in the first or second of the bottom two rows of the table. Westria's decision whether to divert or not determines whether the outcome will be the payoffs in the last or next to last column. In each cell the first payoff is to Eastland and the second to Westria.[2] Putting all this together, the table tells us (for example)

*Table 2.1 Game 2.1 in strategic normal form*

| Payoff order: Eastland, Westria | | Westria | |
|---|---|---|---|
| | | Divert | Don't |
| Eastland | Divert | −1,−1 | 3,−4 |
| | Don't | −4,3 | 0,0 |

that if Eastland diverts and Westria does not, the net benefits are as shown in the upper right cell, a loss of 4 for Westria and a net gain of 3 for Eastland. This presentation will be familiar to those acquainted with the Prisoner's Dilemma example. When the game is presented in a tabular form such as this, the game is said to be represented in *strategic normal form* or, more briefly, in *normal form* or in *strategic form.* Fairly obvious extensions of the tabular presentation can be used with games in which there are more than two strategies or in which there are three or even four players.

For larger games, a more mathematical presentation is necessary, and in general we may borrow language from set theory. Letting $S_i$ be the set of all strategies available to player $i$ of $n$ in the game, $\Sigma$ be a set with n elements of which element $i$, which we may call $\sigma_i$, is an element of $S_i$ (that is, $\sigma_i$ is the strategy chosen from $S_i$ by agent $i$), and $\mathbf{v} = (v_1, v_2, \ldots, v_i, \ldots v_N)$ be a vector of the $n$ payoffs to the $n$ players in the game. In place of the table we have a function that gives a vector of payoffs $\mathbf{v} = f(\Sigma)$, with one payoff for each of the players $i$, corresponding to each possible set of strategies $\Sigma$. Then the game (in strategic normal form) is said to comprise the set of players $N$, the set of strategies $S_i$ for each player, and the payoff function $f$. In general, any table is just a visual way of presenting a mathematical correspondence, and that is true equally of the payoff table in game theory as of other tables.

Von Neumann and Morgenstern (2004) proved that any game in extensive form can be represented also in strategic normal form, and in particular, the Entry Game can be represented in that way. However, there is a trick to it, and the trick is often neglected, even in game theory research that is in some ways quite advanced. For Firm B, the case is similar to the case for Eastland and Westria: Firm B simply has to choose between two actions, enter or don't. Firm A, however, knows Firm B's decision when it makes its own decision, and Firm A's decision is conditional on that knowledge. This is a *contingent strategy*. For firm A there are four contingent strategies:

Strategy 1: "If Firm B enters then retaliate, otherwise retaliate."
Strategy 2: "If Firm B enters, then accommodate, otherwise retaliate."

*Table 2.2 Game 2.2 revised: the Entry Game in strategic normal form*

| Payoff Order: Firm B, A | | Firm A | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| B | Enter | −1,1 | 1,2 | 1,2 | −1,1 |
| | Don't | 0,5 | 0,5 | 0,5 | 0,5 |

Strategy 3: "If Firm B enters, then accommodate, otherwise accommodate."
Strategy 4: "If Firm B enters, then retaliate, otherwise accommodate."

This list of contingent strategies may seem trivial, redundant, and (in the case of Strategy 2) downright silly, by comparison with a simple enumeration of the choices to retaliate or accommodate, but all are contingent strategies that are available to Firm A in the light of the information it has when it makes the decision. Therefore, all are necessary for a valid presentation of the game in strategic normal form as defined by von Neumann and Morgenstern. Table 2.2 shows Game 2.2 in strategic normal form.

It is common to refer to decision alternatives such as "enter" and "don't enter" and "accommodate" or "retaliate" as "strategies," but this is not consistent with the representation of the game in strategic normal form as understood by von Neumann and Morgenstern. The extensive form was clarified in a key early paper of Kuhn (1997, pp. 46–68). In it Kuhn distinguished between strategies as conceived by von Neumann and Morgenstern and what he called "behavior strategies," that is, local decisions in the different decision nodes of the tree. It has become common not to make that distinction, and in some cases confusion can result. For this book, the decision alternatives such as "enter" and "don't enter" and "accommodate" or "retaliate" will be called *behavior strategies*[3] and strategies such as strategies 1, 2, 3, and 4 above will be called *contingent strategies*. I will make every effort not to use the term "strategy," without modification, unless the meaning will be clear from the context.

Here is an example that will illustrate the importance of distinguishing contingent from behavioral strategies. To give the example a real-world background, consider a case in allocation of intellectual property. Firms A and C have patents on alternative methods of producing widgets. Firm C is an established monopolist but the cost of production with their patented technology is relatively high. The technologies are complementary, so that a company in possession of both technologies could be a low-cost producer in the widget market. Firm B is known to be interested in entering

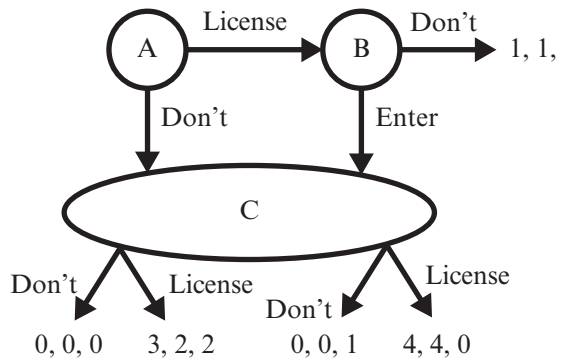


*Figure 2.3 Game 2.3: Selten's "Horse"*

the widget market and has applied to license Firm C's patent. Firm C does not know whether or not Firm A has licensed its patent to Firm B, but Firm B's application to Firm C for a license would make sense only if Firm B intends to enter. If B holds the license for A's patent, and does not enter the widget market, then each firm can continue in its traditional markets, and we represent this outcome by payoffs of 1,1,1 (in the order of Firm A, B, C). If Firm B does enter, and does so with a license only for Firm A's technology, the market will be shared profitably for all firms; we represent this by payoffs 3,3,2. If Firm B enters with neither license, its competition will leave all three firms without profits, which we represent by payoffs 0,0,0. If Firm B enters with only the license for A's patent, then it will be unprofitable and unable to pay license fees to Firm A, so that those firms will be unprofitable, but Firm C will continue with some profits, giving payoffs 0,0,1. Finally, if B enters with both licenses, it will become dominant as the low-cost producer in the widget market, so that Firms A and B will be highly profitable but Firm C unprofitable, for payoffs 4,4,0. Again, the reader may multiply these payoff numbers by a factor large enough to make them "realistic."

Thus we have the game shown in Figure 2.3. With this sequence of decisions and payoff numbers we have replicated an example of Selten (Kuhn 1997, pp. 312–54)[4]. It is called "Selten's Horse," because of a slight resemblance of the diagram for the game in extensive form to a horse. Accordingly, we will follow Selten's notation and denote the behavior strategies as follows:

- Player A. License = R1, don't = L1
- Player B. Enter = L2, Don't = R2
- Player C. Don't = L3, License = R3

*Table 2.3 Game 2.3: a normal form representation of Selten's "Horse"*

| Payoff order Firm A, B, C | Firm A | | | |
|---|---|---|---|---|
| | Don't | | License | |
| | Firm B | | Firm B | |
| | If License then Don't else no action | If License then Enter else no action | If License then Don't else no action | If License then Enter else no action |
| C If Don't or License, Enter then Don't else no action | 0,0,0 | 0,0,0 | 1,1,1 | 0,0,1 |
| If Don't or License, Enter then License else no action | 3,2,2 | 3,2,2 | 1,1,1 | 4,4,0 |

This game is presented in a valuable advanced textbook as an example where some standard methods break down (Montet and Serra, 2003). It illustrates a case in which some decisions must be made with very limited information, which is the case for Firm C. The information available to Firms A, B and C is complex but clear enough from the diagram. Firm B, the second player, knows what Firm A has done, since it would have no opportunity to choose had Firm A not chosen to license its technology. For Firm C, who plays last, there are three possible sequences of choice by the previous two players. They are

1. L1
2. R1, R2
3. R1, L2

Of these, Firm C can rule out only the second – the first and third are equally possibilities. In order accurately to represent this game in normal form, we have to preserve this information. To do this it is necessary to use the contingent strategies.

The normal form representation of the Horse Game is shown as Table 2.3. In order to represent this three-person game, the strategy choice of Firm A determines which side of the table the other two players play in. If firm A chooses "License", then the other two play in the right side of the table; otherwise, they play in the left. Montet and Serra suggest that this game is particularly problematic for analysis in terms of *behavior strategies*, since

(1) Firm C has no rational basis to choose either of the behavior strategies "Don't" or "License," and therefore (2) Firms A and B, unable to anticipate the choice that will be made by Firm C, also do not have any rational basis for their own choices. However, Selten (1975) found a rational-action solution for the game represented in strategic normal form with contingent strategies. This solution will be discussed in Chapter 6. For further discussion of contingent strategies and further examples see Dutta (1999, Ch. 2) and McCain (2004, Ch. 2).

When the game is represented in extensive form, we define a *subgame* as a full-information node (such as A in Figure 2.1) together with all possible moves that follow it in the game tree (as "accommodate" and "retaliate" follow A in Figure 2.1). In a larger, more complex game, subgames could themselves be quite large and complex. For generality, it is conventional to regard the whole game (beginning with B in Figure 2.1) as one of its subgames; smaller subgames are called "proper subgames." Notice that the 'Horse' Game has no proper subgames. These concepts are standard in introductory texts, where many more examples can be found; see, for example, McCain (2004, Ch. 14) and Dutta (1999, Ch. 11).

Following Tsebelis (1990), we may denote any sequence of moves that is part of a game in extensive form as a *nested game*. If the sequence of moves is a subgame, then we will denote it as an *imbedded game.* Examples may be found in Chapter 6 below, and see also McCain (2004, Ch. 15). These concepts will be important for our purposes since the activities of the private sector in a market economy will always be nested in the larger game that includes the determination of public policy. If, in addition, the private sector activities constitute an imbedded game, that is, a subgame of the larger game, then non-cooperative game theory can validly be applied to them. If the private sector activities are nested but not imbedded, then we will need to be more careful (as Tsebelis observes).

The strategic normal form representation has been central to most applications of non-cooperative game theory, and will be extensively discussed in Chapters 4 and 5.

## 2.4 UNCERTAINITY AND CALIBRATION

While most non-cooperative game theory is based on "games" with numerical payoffs, this needs to be qualified in two ways. First, it is recognized that the actual benefits attached to decisions are subjective benefits, and that the numbers ideally should be quantities of subjective satisfaction, that is, in the language of twentieth-century economics (and earlier utilitarianism) the payoffs are quantities of "utility." Second, payoffs may be uncertain from

the point of view of the player, the theorist, or both. In the Water Game, as we have seen, Westria must make their decision in what is called an "information set" but might better be thought of as an ignorance set – Westria does not know what decision Eastland will have made. In this example, as in game theory in general, the interactive decisions have to be made under *uncertainty.* Put otherwise, Westria chooses between two strategies, not on the basis of their payoffs, but on the basis of some probability distribution over the payoffs, which in turn depends on a probability distribution over the strategies of others. The simplest way to deal with this problem is also the one usually used: we assume that the decision-maker wants to choose the strategy that yields the largest mathematical expectation of payoffs. Thus probability and mathematical expectations play a central role in game theory. In much of the research literature it is taken for granted that the objects of choice are probability distributions over strategies and payoffs, and the literature is hardly intelligible if this is not kept in mind.

Then (a frequent question in beginning classes) how do we figure out what the payoff numbers should be? Generically, game theory models are not usually calibrated precisely. In some applications, the application itself may provide evidence that can be used for calibration. The Water Game is an example in which (hypothetically) the calibration is derived from cost–benefit analysis in the specific case. Very commonly the payoffs are treated as algebraic unknowns, with some restrictions on their values. It is fairly common to find that the payoffs can vary over some positive range without affecting the qualitative results, while payoff values outside that range have very different results. In the Water Game, for example, all the payoffs can be multiplied by any constant number, or have any constant number added to them, or both, and the rational decisions of the two players in the game will not change. Thus exact calibration is often not necessary, and may not be very helpful, and the numbers chosen for illustrative purposes can be arbitrary, so long as they are within appropriate ranges.

## 2.5 COOPERATIVE GAMES

The examples we have seen so far are drawn from "non-cooperative" game theory. The Water Game, in which the agents act with deliberate secrecy and distrust one another, illustrates a non-cooperative game particularly well. Cooperative game theory is applicable whenever the players in a game can form "coalitions," that is, groups that choose a common strategy to improve the payoffs to the members of the group. Cooperative game theory relies especially on mathematical set theory for many of its basic ideas. The usual assumption is that any group of agents in the

"game" can form a coalition, and a coalition of agents $a$, $b$, and $c$ would be denoted as $\{a, b, c\}$. The brackets {} are conventional in set theory to indicate the "elements" of a "set," or in alternative ordinary language, the individuals who make up a grouping.

Most studies in cooperative game theory will begin with an enumeration of all possible coalitions. As before, suppose there are $n$ "players in the game." An individual agent can be indicated by $a_i$, or simply by the index $i$, with $i = 1, \ldots n$. Common sense would see that any group of agents with more than one and less than $n$ could form a coalition (with more or less difficulty). That's right, but it is not complete. In addition to those groupings, we also enumerate all *singleton coalitions*, $\{a_i\}$, that is, "coalitions" with just one member, and also the *grand coalition* of all n agents in the game and the null coalition, $\varnothing$, a "coalition" with no members. (By convention in set theory $\varnothing$ means a set with no members.)

When a coalition is formed, the expectation is that by working together and choosing a joint strategy they will be able to improve their results overall. It may be that one member, let us say $c$, bears a special cost for this, or another agent, such as $a$, gets most of the benefit. An example might be the modification of a river course, so that those upstream benefit (from the diverted water) but those downstream lose. Then $c$ is downstream and $a$ is upstream. To enlist $c$ in the coalition it may be necessary for $a$ to pay $c$ some compensation. This could be a problem faced by the governments of Eastland and Westria in their domestic water policies, if we think of a government as in each case a coalition of interest groups within the country. It is common to assume *transferable utility*,[5] which means that a simple transfer of some of $c$'s winnings to $a$ can fully compensate $a$. This could be true, for example, if all payoffs are in money and the players' "utility functions" are proportional to their money payments, including game payoffs and side payments. This simplifying assumption is often called TU.

Assuming TU, all that matters is the total payoff to the group $\{a, b, c\}$. Therefore, it is common in cooperative game theory to ignore all the details and to focus on the total values the various coalitions can obtain, and assign to each coalition a number expressing that value. It is commonly called the value of the coalition. (Of course, the value of $\varnothing$ is zero.) This assignment is called a "characteristic function" in mathematical set theory and is sometimes called the "coalition function" in cooperative game theory.

Rousseau's tale of the Stag Hunt has given rise to a widely used example in the theory of non-cooperative games. Here, though, we will use it as an example of a cooperative game. Rousseau's ideas are an important root of modern political theory and the Stag Hunt example is something of a paradigm of collective action. (For representative treatments see Gardner,

2003, pp. 115–18; Osborne, 2004, pp. 20–21.) Rousseau writes in part 2 of the *Discourse on Inequality* (1754, G. D. H. Cole translation):

> In this manner, men may have insensibly acquired some gross ideas of mutual undertakings, and of the advantages of fulfilling them: that is, just so far as their present and apparent interest was concerned: for they were perfect strangers to foresight, and were so far from troubling themselves about the distant future, that they hardly thought of the morrow. If a deer was to be taken, every one saw that, in order to succeed, he must abide faithfully by his post: but if a hare happened to come within the reach of any one of them, it is not to be doubted that he pursued it without scruple, and, having seized his prey, cared very little, if by so doing he caused his companions to miss theirs. It is easy to understand that such intercourse would not require a language much more refined than that of rooks or monkeys, who associate together for much the same purpose.

It is not clear that Rousseau has anything in mind that corresponds to the rational-action analysis typical of game theory. Rather, he is concerned with lack of foresight among beings not yet human enough to have language! However, the behavior described is not necessarily irrational.

It seems that Rousseau had something like the following hunting technique in mind: the individual hunters act as beaters and spearmen, driving the stag into a narrow defile where it can no longer escape. While they are beating, it may be that one spies a rabbit and, to pursue it, abandons his post, so that the stag escapes through the gap so created. The individual thus forfeits his own share of the stag (while depriving the others of theirs as well) but he may reason that if he does not pursue the rabbit, some of the other beaters will do so, and thus the stag will get away anyway, so that he has nothing to lose by pursuing the rabbit.

It is rather odd, nevertheless, to treat the Stag Hunt as a non-cooperative game. In order to pursue a stag, it is necessary for the two (or more) hunters to form a coalition for the purpose, and if they obtain a stag the coalition acquires a single large source of meat, not separate payoffs to different hunters. The hunters must have some agreement as to how the meat will be divided, and the individual payoffs can be based only on the agreement.

Let us instead treat the Stag Hunt game as a cooperative game, and to better illustrate the cooperative-game concepts, let it be a three-person game. As before we suppose that a stag can be taken only if all (three) hunters collaborate in pursuing the stag, that anyone can catch a rabbit, that a rabbit is worth one day's supply of meat for a family while the stag can supply 10 family-days of meat to be divided among the three hunters. We can then say that any singleton coalition is worth just one (rabbit) family-day of meat, any two-person coalition is worth two (rabbits) family-days of meat, while a three-person grand coalition is worth ten family-days

*Table 2.4 Game 2.4: a cooperative, three-person Stag Hunt*

| Coalition | Value |
|---|---|
| $\{a,b,c\}$ | 10 |
| $\{a,b\}$ | 2 |
| $\{a,c\}$ | 2 |
| $\{b,c\}$ | 2 |
| $\{a\}$ | 1 |
| $\{b\}$ | 1 |
| $\{c\}$ | 1 |

of meat. This gives us the coalition function shown as Table 2.4. The three hunters are indicated as $a$, $b$, and $c$.

A simplifying assumption that is usually made is that the game is "superadditive," that is, a coalition formed by the merger of two or more coalitions will realize a value at least as great as the sum of the values of the coalitions merged. We see this in the Stag Hunt game.

Now let us consider a game of production of a public good,[6] Game 2.5. Once again we will think of a three-person game, since it allows for the formation of some range of coalitions while remaining comparatively simple. For this game, the three agents, again imaginatively called $a$, $b$, and $c$, each begin the game with wealth amounting to 5 units, supposing that the public good is not produced. For simplicity, we suppose that the public good is imperfectly divisible. Specifically, each agent can provide a fixed lump of 1 unit of the public good at a cost of 3, and the strategies will be to provide or not to provide this indivisible lump of public good. Payoffs will be the sum of the initial wealth and double the number of units of the public good provided, minus the individual's cost of providing the public good if he determines to provide it. We now consider three cases:

(a) *A singleton coalition.* If the public good is produced, the wealth of the singleton coalition (individual agent) that produces it would be $2k + 5 - 3 + 2 = 2k + 4$, where $k$ is the quantity of the public good produced by other agents. The payoff from not producing is $2k + 5$ so that the singleton coalition would be better off not producing it.

(b) *A two-person coalition.* A two-person coalition can produce one or two units of the public good, and none will be produced by the singleton residual coalition, as we have already seen. With one unit the value of the coalition is $10 - 3 + 4 = 11$. With two it is $10 - 6 + 8 = 12$. Thus the two-person coalition would be better off producing two units of the public good. For this example we assume that they do.[7]

(c) *A three-person coalition.* A three-person coalition can produce up to three units of the public good, and similar reasoning leads to the conclusion that it will produce 3 for a value of $15 - 9 + 3 * 6 = 24$ if the public good is produced, and 15 if not, so the grand coalition would choose to produce three units of the public good.

Notice, though, that the payoff for a singleton coalition also depends on whether the public good is produced. Suppose, on the one hand, that the group is "organized" into three singleton coalitions. Then, as we have just seen, the worth of each is 5. Suppose, on the other hand, that $a$ and $b$ form a two-person coalition and produce two units of the public good. Then the value of $\{c\}$ is $5 + 4 = 9$. Evidently, the value of a singleton coalition in this example depends on whether the other two agents in the game form a coalition among themselves, and if so what strategy they jointly choose.

These changes in the value of a coalition, as a result of the formation of another coalition, are called *externalities*[8] in some recent work in game theory (for example, Carraro, 2003). When the coalition function is taken as the only important information about the game, as so much cooperative game theory has done, this amounts to the unstated assumption that externalities are unimportant.[9] This, too, seems quite problematic for economic applications. As early as 1963 Thrall and Lucas proposed a more complex way of assigning values to coalitions, the *partition function*, which allows for externalities in this broad sense.

A partition is a common construct of set theory. A *partition* $\mathcal{P}$ of set N is a set of subsets of N with two properties: the subsets do not overlap and every member of N is in one or another of the subsets that are members of $\mathcal{P}$. A second partition $\mathcal{Q}$ is a *refinement* of $\mathcal{P}$ if it breaks down the subsets that are members of $\mathcal{P}$ into smaller subsets. These can be applied to sets of any kind, from brands and models of automobiles to classes of probability distributions over the unit interval. Thus, for example, the set of all automobiles is partitioned when we identify them according to brand, and since each model (such as the Chevrolet Impala or the Subaru Outback) is produced by only one brand, the identification of automobiles by model is a refinement of the identification by brands. The *partition function* then is a function from coalitions and the partitions of which the coalitions are members to the value of the coalition, where the coalition can take different values in different partitions. An *imbedded coalition* is a pair consisting of a partition $\mathcal{P}$ and a coalition $C$ that is a member of the partition. Thus the partition function assigns values to imbedded coalitions. (For more detail see Chapter 11. These terms are given mathematically formal definitions in Chapter 13, which may be read independently of the other chapters.)

*Table 2.5 Game 2.5: a partition function for the three-person public goods game*

| | Partitions | Values |
|---|---|---|
| 1 | $\{a, b, c\}$ | 24 |
| 2 | $\{a, b\}, \{c\}$ | 12, 9 |
| 3 | $\{a, c\}, \{b\}$ | 12, 9 |
| 4 | $\{a\}, \{b, c\}$ | 9, 12 |
| 5 | $\{a\}, \{b\}, \{c\}$ | 5, 5, 5 |

A partition function for the public goods game, as discussed above, is shown in Table 2.5. Once again, this game is superadditive. Suppose, nevertheless, that we focus on one of the lines other than the grand coalition, and ask for a "cooperative solution," that is, an assignment of values to the individual agents in the coalitions in that partition. An example would be the middle line, $\{a, c\}, \{b\}$, where the problem would be to determine the allocation of the coalition value of 11 between $a$ and $c$, consistently with the assumptions of cooperative game theory. The partition $\{a, c\}$, $\{b\}$ would be called a *coalition structure*, and the solution would be for that coalition structure.

As we observed, a partition is a general concept of set theory. It will sometimes be helpful to think in terms of partitions of other kinds of sets. For example, any decision can be thought of as creating a partition among the different possible outcomes of a game. In Selten's Horse Game, for example, Firm A's decision creates a partition of the four possible outcomes into two sets of two: {(0,0,0), (3,2,2)} and {(0,1,1),(4,4,0)}. Similarly, chance events partition all possibilities into different subsets that may follow from different realizations of the chance event.

## 2.6 "IMPERFECT RECALL"

In his classic paper that founded the literature on games in extensive form, many of Kuhn's (1997, 46–68) results were limited to the case of "perfect recall." Kuhn's example of "imperfect recall" was the game of bridge. In bridge, two partnerships play against one another, and each partnership receives a single score. Thus, Kuhn argued, the partnership is a single player, but the player has multiple "agents" each of whom may be unaware of the strategic commitments of the other. Many of the conventions of bridge play are designed to address this difficulty.

Nontrivial coalitions are compounds of multiple agents, and when we observe coalitions interacting competitively against one another, they are compound players such as Kuhn described. Nevertheless, there seems to be no literature whatever on cooperative solutions for games of "imperfect recall." The popular view is that cooperative game theory has been less influential in economics since the early period because noncooperative game theory is better able to deal with informational problems (Weintraub, 1992, p. 7). To address these problems in cooperative game theory would require a discussion of games of "imperfect recall." Even for superadditive games in coalition function form, such as exchange games, "imperfect recall" can have crucial implications, points that will be discussed in Chapter 8.

## 2.7 NON-NUMERICAL OBJECTIVES

So far we have assumed that the consequence of choosing a particular strategy, while others choose their strategies, is a numerical payoff to each player. In many cases the numbers we have chosen have been based on very rough reasoning, a measure of conjecture, and no more. Numerical payoffs lend themselves to transferable utility games, in which side payments take a central role; and side payments are very common in the real world. Salaries, wages, interest payments, dividends, and taxes are all instances of side payments. While side payments can be part of a game without transferable utility, in TU games the determination of side payments can be treated explicitly rather than being left in the background. Numerical payoffs are almost always assumed in the theory of noncooperative games. Moreover, von Neumann and Morgenstern thought very carefully about how the numbers should be established, and their von Neumann-Morgenstern utility theory has not been improved upon. We have not gone into that, because it is not central to our purpose. However, it will be worthwhile to sketch how game theory might be done without numerical payoffs.

We may, then, define a game in a more general way. Here is a concrete example drawn from World War II. The successful Allied invasion of Normandy followed a massive and remarkably successful attempt to deceive the Germans into believing that the invasion would be elsewhere. By this time, Germany could not have achieved a victory and hoped by defeating an Allied invasion to bring the war to a stalemate that would have allowed the Nazi regime to survive. The Germans were led to believe that the invasion would be at Calais rather than Normandy, and this continued after the invasion itself, leading the Germans to believe that this

*Table 2.6 Game 2.6: the Normandy invasion game*

| | | Germany | |
|---|---|---|---|
| | | Defend Normandy | Defend Calais |
| Allies | Attack Normandy | Stalemate | Victory |
| | Attack Calais | Victory | Stalemate |

invasion was a feint and the real, larger invasion was yet to come at Calais. With some simplification – well, with a lot of simplification – we can see this as a two-by-two game, as shown in Table 2.6. Each country had two alternatives to choose from (by this stage in the war) and there were just two outcomes: Allied victory or stalemate. These are shown in place of payoffs in Table 2.6.

How are we to attach numbers to these outcomes? If we assign a value of zero for both sides to a stalemate, and one for the Allies and minus one for the Germans in case of Allied victory, this seems to represent the case adequately, within the limits of the information that we are given. We can do no better than that.

The *outcome* of a game, then, is a complicated object that may have many numerical dimensions or none, and for some dimensions may be Boolean (that is, either-or states as with the Normandy invasion) or real numbers, or other number forms, perhaps in combination. All that we insist on is that there is a set of outcomes, comprising no less than two, and that the players are allowed to disagree as to which one they would wish to have realized. For games of exchange, an outcome is an allocation of goods, services and resources among the agents in the market, and has dimensions of all goods and services for each agent. For another example, think of a political coalition in a parliamentary government as in many European countries. The formation of a governing coalition is a complex game comprising various elections and negotiation among parties following the election. The outcome of such a game is a government program, which may (depending on the strategies of the contending parties) include more or less subsidy to agriculture, an aggressive or defensive foreign policy, support or nonsupport for an established church and its clergy. Even if we cannot assign utility numbers to outcomes, we suppose that players have preferences over them and can evaluate them in relative terms.

This more general approach to representing games has been used primarily in cooperative game theory, especially where (realistically) there may be difficulties making side payments. In that case, we think in terms of non-transferable utility (NTU) games. This approach is little used

in non-cooperative game theory, but can be useful in thinking through practical problems even when the approach is non-cooperative. As we begin to frame a policy problem as a game theory model, it will often be more natural to begin by asking what different outcomes are possible, and thinking of the strategies as determining outcomes, and only then to translate the outcomes into payoff numbers. Recall the Water Game, in which (hypothetically) the payoff numbers were derived from cost–benefit analysis. But the economists who conducted the cost–benefit analysis will have begun with some qualitative listing of the outcomes of the water policies of the two countries: ample supplies of drinking water in one country or another, more or less depletion of water for irrigation, silting and loss or costly maintenance of navigable channels.

## 2.8 SUMMARY

Since game theory is interactive decision theory, representing a social arrangement as a "game" means representing it as an interactive decision. There are two major ways of doing this for non-cooperative game theory. The most commonly used is the game in strategic normal form, essentially a table or function from the strategy choices of all players to their payoffs. For this purpose, it is important to represent the choice of behavior strategies as contingent on any information the agent may have at the time a decision is made. The second way of representing games for non-cooperative game theory is the game in extensive form, that is, a decision tree in which the different "players in the game" govern different decision nodes. With information sets, this allows a much more explicit representation of the information available in the game, and is often used when information is important. For cooperative games, the game is usually represented by a coalition or characteristic function, assigning a total value to each coalition that might form. This is usually linked to the assumptions of transferable utility and superadditivity. For practical purposes, however, the coalition function and superadditivity are often not helpful as they assume away the very problems we want to analyze. Accordingly, we may instead represent the game as a partition function (in which the value of each coalition depends on the other coalitions that form) and attempt to determine which partition, or coalition structure, is likely to be observed. Despite the long history of this approach, however, the theory remains unsettled. As a rule, we will think in terms of numerical payoffs and transferable utility, recognizing that this is at best an approximation to reality, but usually, we may hope, a good one.

# NOTES

1. My students feel that a better term would be "ignorance set," since it is expressive of ignorance rather than information. Nevertheless, of course, we will use the conventional terminology.
2. It is common, as a matter of convention, to record first the payoff to the player to the left, but in this book the order will routinely be indicated in the upper left space. I am indebted to my colleague Richard Hamilton, MD, for that presentation.
3. This follows Kuhn's terminology. We should observe, though, that "behavior strategies" in this sense have nothing to do with the perspective of behavioral science or "behavioral game theory," in that the behavioral strategies may be supposed to be chosen with rationality that is flawless, regardless of experimental evidence of the limited cognitive capacity of real human beings to make such choices.
4. This specific example was included because of its importance in the history of game theory. If the example of patent licensing seems a little stretched, it is because the original example had no such application in mind – and indeed no application of this example has ever been offered before, to the best of my knowledge! The representation of the patent licensing problem could probably be improved; but the general principle – that a decision-maker may not know the sequence of decisions by others that has brought his own decision about – probably is not uncommon in reality.
5. This phrase recognizes that the ultimate benefits of economic activity are subjective, that is, in the economist's language, "utility," but assumes that utility is nearly enough proportional to money that money side payments will assure that everybody in a coalition benefits on net from the coalition's activity.
6. When written in this way, with the indefinite article – "*a* public good" – this term makes reference to a concept systematically discussed by Paul Samuelson (1954). With the definite article – "*the* public good" – the meaning is of course much broader and the usage much older. Unfortunately, the two meanings are often confused. A (Samuelsonian) public good is defined by its technical conditions of production and offering: the incremental cost of adding one user is always zero, and there is no practical way to make payment a condition of use. Thus provision at a zero price is (on the one hand) unavoidable and (on the other hand) efficient, since the marginal cost is also zero. In particular, free provision from general tax revenues or on a philanthropic basis is efficient.
7. Most cooperative game theory follows von Neumann and Morgenstern (2004) in assuming (on the contrary) that coalition values are the worst that an opposed coalition can bring about. We return to this issue in Chapter 10.
8. The term "externalities" is used more narrowly in economic theory. When a cartel is formed and thus imposes costs on customers, this would have been called a "pecuniary externality" in, for example, Scitovsky (1954); but modern economic theory does not regard "pecuniary externalities" as externalities.
9. There is a modest literature on externalities in cooperative games in coalition function form. A key paper, Shapley and Shubik (1969), will be reviewed and criticized in Chapter 10.

# 3. A brief interpretive history of game theory

Game theory has, of course, a prehistory; but much as we can say that economics (as a distinct field of study) began with Adam Smith's 1776 *Wealth of Nations* (Smith, 1994), so we can say that game theory began with von Neumann and Morgenstern's 1944 book, *The Theory of Games and Economic Behavior* (von Neumann and Morgenstern, 2004). Accordingly we shall pause only briefly over the prehistory. In 1913, Zermelo had initiated the mathematical literature on analysis of games. Borel had written important papers that seem to have influenced von Neumann (see Poundstone, 1992, pp. 41–2). With a presentation in 1926 and publication of the corresponding paper in 1928 (von Neumann, 1959), von Neumann had set out many of the themes that would recur in his book with Morgenstern, to which we will return. Aumann and Maschler (1985) find a cooperative solution concept in the Babylonian Talmud. A Korean scholar suggests to me that Sun Tzu should be considered a game theorist[1]. Indeed it is likely that insights of game theory have often occurred to thoughtful people engaged in their own conflicts throughout much of history. See Paul Allen's website for a schematic history of game theory, including several other "prehistoric" contributions (Allen, 1998).

Morgenstern had used his example of Sherlock Holmes and Moriarty to illustrate problems of interactive decision, relating them to economics and forecasting, in a 1928 book and a 1935 paper (TGEB, pp. 712–14[2]). By the mid-1930s the convergence of his ideas with those of von Neumann had been pointed out to Morgenstern, but Morgenstern was unable to pursue that direction until he had been dismissed from his position in Vienna by the new Nazi regime as "politically unbearable" (TGEB, p. 715).

## 3.1 THE FOUNDING BOOK

In his 1928 paper von Neumann assumed that games would have numerical payoffs. In the founding book, von Neumann and Morgenstern first take up the nature of the payoff numbers. Assuming that the ultimate benefits of any economic activity are subjective, in the nature of utility,

Chapter 1, Section 3 develops a numerical utility concept. This is the von Neumann-Morgenstern utility index. Given that agents have consistent subjective preferences over risky prospects such as lottery tickets, the index allows utility to be expressed in terms that are consistent with decisions that maximize expected values of the utility index. These decisions will then be rational in terms of the given preferences.

Von Neumann and Morgenstern then take up the character of a solution in game theory. Ideally, the solution is an imputation, that is, it tells us the quantity of "utility" that each participant in the game can expect on the basis of "rational behavior." For each individual, the solution would also constitute a set of rules for rational behavior in any conceivable circumstances. However, this may not always be possible, and in general the solution may constitute just a set of imputations.

In the second chapter von Neumann and Morgenstern address the representation of a game (interactive decision problem) for mathematical analysis. Von Neumann had in 1928 given both of the representations now common for non-cooperative games, the extensive and normal form.[3] For a given game in extensive form (sequence of decisions at the successive stages of the game) each *strategy* for a given player is the sequence of decisions made in the successive stages of the game. But a key point is that these decisions are contingent decisions, as discussed in Chapter 2 above. Von Neumann writes (1959, p. 18) "For each possible combination of results of the 'draws' and 'steps' [known to player $S_m$] . . . it must be specified what $S_m$'s decision . . . is going to be."

In the book, three representations are given. The first is the intuitively appealing idea of a game as a sequence of decisions at the successive stages of the game, some by individual participants in the game, and some by a random mechanism. This can be visualized as a tree diagram (TGEB, p. 65) and corresponds to the game in extensive form as it is discussed in recent game theory. The second, however, is a rather more complex object. Here von Neumann and Morgenstern represent the game as a sequence of partitions. Von Neumann and Morgenstern begin with a set of all outcomes of the game, and decisions and chance events are treated as partitions of this underlying set of outcomes, and later decisions or chance events produce refinements of the partitions generated by earlier decisions and events. (For this discussion, an "outcome" is simply a list of the payoffs to all players.)

Von Neumann and Morgenstern then discuss the meaning of a *strategy* for a game specified in this way: "Imagine now that each player . . . instead of making each decision as the necessity for it arises, makes up his mind in advance for all possible contingencies . . . We call such a plan a *strategy*" (TGEB, p. 79). Once strategies have been defined in this way,

as contingency plans, it is possible to return to the "very simplest description," (TGEB, p. 79), that is, the game in strategic normal form. This is the third and crucial representation in von Neumann and Morgenstern's Chapter 2.

The significance of the formalization in terms of partitions is that it clearly establishes the link between the game in extensive form and the game in strategic normal form. The treatment of information in the game expressed as a sequence of partitions is cumbersome, but it disappears in the game in strategic normal form. This powerful and brilliant simplification undoubtedly accounts for much of the impact of game theory on economics and on other fields in which strategy or interactive decisions are important.

In the most famous and definitive section of the book, Chapter 3, von Neumann and Morgenstern address two-person, zero-sum games, deriving the minimax solution in mixed strategies that von Neumann had already discussed in the 1928 paper. A brief digression on mixed strategies is in order. The set of strategies (sequences of contingent choices) already discussed are designated as "pure strategies." In a mixed strategy equilibrium the players may "randomize" their choices of strategies; that is, they may choose among the finite set of "pure" strategies according to a probability distribution. It is the probability distribution that is adapted so that the player balances the advantage of choosing a particular strategy against the danger of having her strategy "found out" by the opponent (to use a phrase that recurs in *The Theory of Games and Economic Behavior*).

The remainder of *The Theory of Games and Economic Behavior* was devoted to generalization of that model to the more general case of N-person, nonconstant-sum "games." To do this von Neumann and Morgenstern adopted a research strategy that has been highly productive in mathematics, literally for thousands of years. The strategy is to make the solution to the simple case, with some appropriate transformation or extension, the solution to the more complex case.

In his 1928 paper, von Neumann had made the zero-sum restriction part of the definition of the game. The introduction of nonconstant-sum games was an important contribution of *The Theory of Games and Economic Behavior*, and it was clearly essential in order that the theory be applicable to "economic behavior" and to the social sciences in general. Nevertheless, in 1928 von Neumann sketched some essentials of a theory of n-person zero-sum games, with some of the difficulties to be encountered. Even for three-person games, von Neumann admitted no solution that would not allow for coalitions (von Neumann, 1959, p. 33). Coalitions are also central to the discussion subsequent to Chapter 3 of *The Theory of Games and Economic Behavior*. This is seen as being all the more crucial for

applications to "economic behavior;" as von Neumann and Morgenstern write (TGEB, p. 15) ". . . the great number of participants [in potentially competitive markets] may not become effective; the decisive exchanges may take place between large 'coalitions,' few in number . . ." In a footnote, they elaborate: these coalitions may include "trade unions, consumers' cooperatives, industry cartels, and conceivably some organizations more in the political sphere."

Thus, in the remainder of *The Theory of Games and Economic Behavior*, they develop theories of n-person games, first with and then without the zero-sum restriction. In each case, they proceed step by step, with detailed analyses of three-person games and some other small-n cases both as preliminary case studies and in reconsideration. For this purpose they adopt the transferable utility assumption and define the characteristic (or coalition) function as a fourth representation of the game. This step, representation of a coalition by the total value it can realize, reflects the fact that side payments may be necessary to form some coalitions and the "transferable utility" assumption implies that they can be made costlessly. For n-person zero-sum games, they argue that the game will always be resolved to a confrontation between two coalitions with absolutely opposed interest. "Since we have an exact concept of 'value' (of a play) for the zero-sum two-person game, we can also attribute a 'value' to any given group of players, provided that it is opposed by coalition of all the other players" (TGEB, p. 238).

For games represented in coalition function form, von Neumann and Morgenstern again proceed largely as von Neumann had done in his 1928 paper. A candidate solution is an "imputation," that is, an assignment to each player of the amount he can expect to receive, with the total of the amounts for each coalition limited by the value of that coalition. They then define a dominance relation on imputations, as follows: one imputation dominates another if there is a set of players who can form a coalition and force the second imputation, and increase their payoffs as a result. Unfortunately (as von Neumann had noted in the 1928 paper) this relation is not transitive. The solution then consists of all imputations such that (1) an imputation in the solution is not dominated by any other imputation in the solution, and (2) every imputation outside the solution is dominated by at least one imputation within the solution. Dominance cycles are a possibility, since an imputation in the solution can be dominated by an imputation outside the solution. There may be many imputations in the solution, and moreover there may be many solutions, and this multiplicity is recognized as a shortcoming of their solution concept (TGEB, pp. 264–6).

The final step is to extend the solutions to n-person nonconstant-sum games. To take this step, von Neumann and Morgenstern construct an

n + 1-person game $\overline{\Gamma}$ corresponding to the n-person game $\Gamma$. In $\overline{\Gamma}$, the n + 1st player is simply a fictitious player whose payoff is the negative of the sum of the payoffs of all the others (TGEB, pp. 505–6). The solution to the zero-sum game $\overline{\Gamma}$will then be the solution to $\Gamma$. However, this requires some modifications, in that the fictitious player controls no strategies and can join no coalitions and make no side payments. Once these elements are included in the "rules of the game," the analysis of n-person constant-sum games is recapitulated. Since the game is superadditive, one may presume that the grand coalition of all actual players will form to exploit the fictitious player (nature?) most effectively, but it remains to determine how the overall gain will be distributed. To set limits on this, once again, the game is expressed in coalition function terms. In the absence of the grand coalition, the situation will again resolve itself to an opposition between two coalitions of actual players. (Just two, since any partition into three or more coalitions will be unstable. With superadditivity, the opposition will be better able to defend themselves by merging into a single oppositional coalition.) Von Neumann and Morgenstern then again apply the minimax theorem to assign a value to each coalition. To do this, they have to assume that a coalition S will face a counter-coalition of the remainder of the actual players in $\Gamma$, called $-S$, who will inflict maximum harm on S even at cost to themselves. This procedure is called "the assurance principle" and the value obtained is "the assurance value," as it is the largest value the coalition can assure themselves of in all circumstances. Von Neumann and Morgenstern concede that ". . . the desire of the coalition $-S$ to harm its opponent, the coalition S, is by no means obvious. Indeed the natural wish of the coalition $-S$ should be not so much to decrease the expectation . . . of the coalition S as to increase its own expectation . . ." (TGEB, p. 540). However, the assurance principle is nevertheless assumed, as inflicting harm is seen as a threat strategy by which the group in $-S$ would hope to influence the others and increase their imputation in the grand coalition that will ultimately form (TGEB, pp. 541, 559). With these qualifications, the dominance relation is again used and a solution set of imputations, perhaps quite a large set, is derived.

As with constant-sum games, solutions of general games could include multiple imputations and there could be many solutions; moreover, it was not known whether every game had a solution. (Lucas later demonstrated that it is not true that every game has a solution, 1968). Further, Von Neumann and Morgenstern concede that their analysis of n-person nonconstant-sum (general) games depends very crucially on the assurance principle (TGEB, p. 559) and, after all, "$\overline{\Gamma}$ is merely a 'working hypothesis'" (TGEB, p. 540) based not on "a purely mathematical analysis [but] more in the nature of plausibility arguments" (TGEB, p. 506) and to be

vindicated, if at all, by its success in applications (TGEB, p. 542). The relative lack of applications of the von Neumann-Morgenstern solution sets in recent game theory suggests that the working hypothesis was inadequate. This is hardly surprising in the founding work: at that time there existed no theory to expand the implications of the possibility that ". . . the natural wish of the coalition –S should be . . . to increase its own expectation" in a nonconstant-sum game (TGEB, p. 540). Nash's equilibrium theory would address that, but was of course not available to von Neumann and Morgenstern; and despite the emergence of the Nash equilibrium theory, writing on cooperative game theory tends to assign values via the assurance principle even today (Telser, 1978; Peleg and Sudhölter, 2003; Forgo et al., 1999). In any case, the founding book of game theory had founded not one but two important streams of research: with the theory of two-person, zero-sum games it founded non-cooperative game theory, and with the theory of n-person nonconstant-sum games it founded cooperative game theory, providing the concepts and solutions in particular cases without which neither would have grown.

## 3.2 THE DICHOTOMY OF COOPERATIVE AND NON-COOPERATIVE GAMES

The appearance of *The Theory of Games and Economic Behavior* caused a great deal of interest, of course, especially at Princeton. Game theory was promptly taken up for defense research by the new RAND corporation, and some of the book reviews published were themselves important contributions. Nevertheless, such a path-breaking work required a few years for absorption, and the next important advance occurred in 1950 as John Nash reported (1950a) an equilibrium concept for n-person nonconstant-sum games. Nash would expand this (1951) into an explicit theory of non-cooperative games. It is important that Nash's equilibrium solution is identical to that of von Neumann and Morgenstern in the case of two-person, zero-sum games. Thus it is again an instance of the classical research strategy of mathematics, making the solution for a simple case, with some appropriate transformation or extension, the solution to the more complex case. Nash's solution differs for all games with three or more players and all nonconstant-sum games because it does not allow for coalitions based on enforceable agreements.

But Nash also (1950b; 1953) made an important contribution to the theory of cooperative games, in the form of an axiomatic theory of bargaining. (Nash's bargaining theory supports the same conclusion as the earlier theory of Zeuthen, 1930, but Nash seems to have been unaware of

this.) In addition, in the 1953 paper, Nash advances on the 1950b paper and on Zeuthen's bargaining model by allowing the bargainers to choose their threats before bargaining begins: thus his is a variable-threat, rather than a fixed-threat bargaining model. Finally, and most important, Nash provides a model for the development of the theory of cooperative games in general. In Nash (1951; CGT, 1997, p. 26[4]) he writes:

> One proceeds by constructing a model of pre-play negotiation so that the steps of the negotiation become moves in a larger non-cooperative game . . . describing the total situation. . . . if values are obtained they are taken as the values of the cooperative game. Thus the problem of analyzing a cooperative game becomes the problem of obtaining a suitable, and convincing, non-cooperative model of the negotiation.

This reduction of cooperative game theory to non-cooperative game theory is the *Nash program* (Serrano, 2003).

In this series of papers, Nash not only extended both non-cooperative and cooperative game theory, but in addition originated the distinction between the two. No such distinction exists in von Neumann and Morgenstern. The idea that the same game might have alternative solutions, cooperative and non-cooperative, with the first applicable only in case enforceable agreements can be made, originates with Nash and is one of the most influential ideas of game theory.

The first experimental study in game theory probably took place in 1949 at the RAND corporation. Merrill Flood involved two secretaries in an experiment that roughly anticipated the Ultimatum Game and the Dictator Game, with surprising results. This was followed in January 1950 with a formal experiment in non-cooperative games. Melvin Dresher collaborated and the subjects were John Williams and Armen Alchian, respectively a mathematician and an economist. While the experiment is of considerable interest in itself, its greatest impact was probably indirect. Alfred Tucker observed the experiment and, in May 1950, addressing a group at Stanford University, originated the Prisoner's Dilemma example. The Prisoner's Dilemma is a symmetrical modification of the game in the Flood-Dresher experiment. (This account follows Poundstone, 1992, Ch. 6.) The Prisoner's Dilemma outcome is a particular case of Nash equilibrium, but a simple and compelling instance in which individual self-regarding action makes both parties worse off than they might otherwise be. As such, it was to have enormous impact and this has been one very important reason for the predominance of non-cooperative approaches in game theory in the later twentieth century.

The game theory research of the 1940s was reflected in 1950 by the first volume of *Contributions to the Theory of Games*, edited by Kuhn

*Table 3.1 Game 3.1: McKinsey's game in strategic normal form*

| Payoffs: Player 1, Player 2 | Player 2 | |
|---|---|---|
| | A | B |
| Player 1 | 0,−1000 | 10,0 |

and Tucker, as number 24 of the *Annals of Mathematics Studies*. Many of these studies are extensions of and computational approaches to the minimax theorem; games with an infinite number of strategies are seen as the research frontier (Kuhn and Tucker, 1950, p. x). Some interesting new developments are found in the collaboration of Nash and Shapley on a simplified three-person poker game. At p. 109 we see what seems to be the first elimination of dominated strategies in the solution of a non-cooperative game. (While the other papers in this volume were also important contributions, brevity will require selectivity from this point on.)

McKinsey published in 1952 what seems to have been the first textbook of game theory. One important novelty in this book (Tucker and Luce, 1959, p. 2; Luce and Raiffa, 1957, p. 190) is the beginning of serious criticism of the representation of cooperative games in coalition function form. McKinsey writes (1952, p. 351) "von Neumann's whole theory of games is based on the notion of the characteristic function. This implies that if two games have identical characteristic functions, then they will have the same solutions. It is, to say the least, debatable, however, whether this is satisfactory from the point of view of intuition." He considers (1952, pp. 351–2) a two-person game in which Player 1 has only one strategy (no alternatives) and Player 2 has two. The game is shown by Table 3.1. As we can see it is highly asymmetrical, and intuition suggests that Player 1, despite his lack of alternatives, is in the better position. Player B can avoid a very large loss only by choosing strategy B, which grants a payoff of 10 to Player 1. Nevertheless, when the game is expressed in coalition function form, it is symmetrical. For notice that the least payment Player 1 can assure himself of is 0 (for although Player 1 can take no action his payoff cannot be less than zero in any case). The least payoff of which Player 2 can assure himself, by choosing strategy B, is zero. Therefore the game in coalition function form is as shown in Table 3.2.

Thus, the coalition (characteristic) function form is symmetrical, failing to capture the most important aspect of the game in strategic normal form. Moreover, when we consider $v\{1\} = 0$ as reflecting a threat by Player 2, it is not very plausible. To reduce Player 1 to the payoff of 0, Player 2 must

*Table 3.2 Game 3.1: McKinsey's game in coalition function form*

| | |
|---|---|
| v{1} | 0 |
| v{2} | 0 |
| v{1,2} | 10 |

take a loss of 1000. According to the assurance principle, this is what Player 2 would do. Is it likely that Player 2 would make such a threat or (more importantly) that Player 1 would find it credible if he did?

In 1952, Shapley and Shubik presented an analysis of cooperative games without the assumption of transferable and linear utility (TU) that von Neumann and Morgenstern had made. This was in a conference of the Econometric Society at East Lansing, Michigan, and the paper is apparently available only in the form of the abstract published in *Econometrica*. Nevertheless it deserves mention here. Shapley and Shubik assume that preferences can be indicated by a numerical index that would not be transferable nor interpersonally comparable, but which attaches higher numbers to more preferable alternatives. Corresponding to any outcome or probability mixture of outcomes would be a vector of utility indices for the N players in the game. A coalition S is then "effective" for utility index vector **x** if there is a joint strategy or mixture for S that will assure them of at least the utility indices in **x** regardless of the strategies chosen by the players not in S. This definition of effectiveness is equivalent to von Neumann and Morgenstern's assignment of coalition values via the maximin operation, that is, the assurance principle. They then define dominance and solution in terms of effectiveness, otherwise following the example of von Neumann and Morgenstern.

The second volume of *Contributions to the Theory of Games* (Kuhn and Tucker, 1953) contained two very important new contributions that would be republished in the collection *Classics in Game Theory* (Kuhn, 1997). The first of these was Kuhn's "Extensive Games and the Problem of Information" (CGT, pp. 46–68). Here Kuhn returned to the representation of a game as a series of partitions of the set of all outcomes, but defined the partition in a different and more general way that allowed for a treatment of the information available to a player at a particular play in a way that is at once more compact and general. The sets that make up Kuhn's "information partition" are the "information sets." Adopting Nash's equilibrium concept as a generalization of the minimax solution, Kuhn proves that all games of perfect information have equilibria in pure strategies, an extension of the theorem of Zermelo and Von Neumann. Kuhn also supplied a geometric visualization of extensive games and their information

conditions that has become standard (note CGT, p. 64). Kuhn defines subgames in the way that has also become standard (CGT, p. 56).

Kuhn's formalization, unlike that of von Neumann and Morgenstern, extends to games in which a player may not be aware of the number of plays that have already taken place (such as Selten's "Horse;" also note CGT, p. 52). It also includes games in which a player is represented in different plays by different "agents" some of whom may be unaware of previous moves made by other "agents" of the same player, that is, games of "imperfect recall" (CGT, p. 65). Kuhn advocates "behavior strategies," that is, local randomization at each step of decision, making use of the information available at that point, rather than the contingent pure strategies of von Neumann and Morgenstern. Kuhn points out considerable computational advantages of behavior strategies, in that rational decisions need not be computed for decision points that will never be reached.

As we noted, Kuhn took Nash's equilibrium as his concept of solution, extending that concept by assuming that at each decision point, the behavior strategies chosen would be *local* best responses given the information available at that point. As a rule (he stressed) these would be randomized strategies. On that basis, he proved (1) that every sequence of behavior strategies chosen in this way would correspond to at least one contingent strategy, (2) every contingent strategy leading to the same payoff outcomes would be identical to the equilibrial sequence of behavior strategies on the information sets actually reached, although there might be many such contingent strategies with different decisions on "irrelevant" information sets not actually reached in equilibrial play, and (3) if the game has "perfect recall," then every such sequence of equilibrial behavior strategies corresponds to a Nash equilibrium of the original game. This is sometimes expressed by the phrase "behavior strategies suffice" and probably accounts for the neglect of the distinction of behavior and contingent strategies in much subsequent work in non-cooperative game theory. This will be reconsidered in Chapter 10.

The other contribution from volume 2 of *Contributions to the Theory of Games* that must be mentioned here is Shapley's value theory, a solution concept for n-person cooperative games (CGT, pp. 69–79). The solution is a value function which, for a given game, assigns a value to each player that is the player's expected payoff from participating in the game. It will be discussed in Chapter 8, at Section 8.2.2. Shapley and Shubik (1954) applied the value theory as an index of power in political organizations.

Although it was not to be published until volume 4 of *Contributions to the Theory of Games* (CGT4, pp. 47–85).[5] Gillies had by this time (1953) developed the concept of the core of a cooperative game and presented it in his doctoral dissertation. The core, like the von Neumann and Morgenstern

solution set but unlike the Shapley value, may include many imputations or may be null; that is to say, there may be no imputations that meet the conditions for membership in the core. Thus, with Shapley's value theory, there were three distinct concepts of solution of cooperative games, an *embarras de richesses* that was only to become more pronounced.

## 3.3 GAME THEORY AS DECISION THEORY

In 1957 Luce and Raiffa published *Games and Decisions*, the first book-length work of research on game theory after von Neumann and Morgenstern. Much of the ground covered was that of von Neumann and Morgenstern, as Luce and Raiffa regarded that work as still the canon for game theory. However, Luce and Raiffa aimed at a more accessible, less mathematical presentation, and they incorporated a number of advances made over 1944–57, including Nash equilibrium in non-cooperative games (GD, Chapter 5[6]), linear programming (GD, p. 17, appendix 5), and Kuhn's formulation of games in extensive form with information sets (GD, p. 42), although they continued to characterize pure strategies and games in normal form in a way consistent with von Neumann and Morgenstern, as a series of contingent decisions with a decision chosen in advance at each information set (GD, p. 51). In cooperative game theory they discuss McKinsey's criticism of the characteristic function (GD, p. 190), the core solution concept (GD, pp. 192–6), Vickrey's then unpublished attempt to introduce farsightedness into von Neumann-Morgenstern solution theory, and also incorporate Nash's bargaining theory and the Shapley value, with some criticisms of them. An original point is that Luce and Raiffa treat these as alternative arbitration schemes (GD, Chapter 6, parts 4–10). Among the advances in this book were a very early discussion[7] of repeated play in the Prisoner's Dilemma (GD, p. 99), including the concept of the unraveling of a cooperative agreement from the last period forward, what seems to have been the first discussion of correlated equilibria in non-cooperative games (GD, pp. 116–9), and a discussion of cooperative arbitration schemes in case interpersonal comparisons of utility may be made (GD, Chapter 6, parts 10–11). They discuss cooperative games in strategic normal form (GD, Chapter 7), present their own solution concept for cooperative games, $\psi$−stability (GD, Chapter 10), and incorporate the work of Savage and its sequelae (GD, Chapter 13) on decisions under uncertainty and of Arrow on collective decisions, along with some discussion of elections, into their game-theoretic framework. And this is not a comprehensive summary!

Luce and Raiffa's $\psi$−stability deserves some further comment. As

they observe, (GD, p. 191) the characteristic function provides very little information about the game. They propose a more informative beginning point: the characteristic function along with a "boundary condition" in the form of a function, $\psi$, from partitions into sets. If $\tau$ is the current partition and S is an element of $\psi(\tau)$, then S is a group of players, not a coalition in $\tau$, that can form and enforce a new partition if its members should choose to do so. If S is not in $\psi(\tau)$, then it is not capable of upsetting the existing partition however much it might have to gain. At the same time, the $\psi-$stable solution is a partition of the population into coalitions – a coalition structure – as well as a set of imputations. This seems to be the first attempt to construct a theory that would explain a stable "coalition structure" other than the grand coalition.

During the 1950s, the theory of differential games (games in which strategies evolve over continuous time) was developing rapidly, and several papers in *Contributions to the Theory of Games*, volume 3 (Dresher et al., 1957) and in *Advances in Game Theory* (Dresher et al., 1964) focused on this theory. However, these will not be important in the chapters that follow.

Volume 4 of *Contributions to the Theory of Games* (Tucker and Luce, 1959) was focused on n-person games (CTG4, p. 1) and reflects especially the search for alternatives to the von Neumann and Morgenstern solution set, especially the definition of the characteristic (coalition) function in terms of the assurance principle (CTG4, p. 2). Papers by Shapley (CTG4, pp. 145–62) and by Shubik (CTG4, pp. 267–78) applied cooperative game theory to market exchange. Shubik had throughout the 1950s published a number of contributions relating game theory to economics (and to some extent to management and political science). His paper introduced the idea that the core of a market game would correspond to Edgeworth's market theory, an idea that was to dominate applications of cooperative game theory (and, arguably, of game theory in general) to economics in the decade to follow. Vickrey (CTG4, pp. 213–46) proposed to modify the von Neumann-Morgenstern theory by taking into account that some dominance relations among imputations might be shortsighted (though he did not use that terminology explicitly). We will see this concern recur in Chapter 11. Aumann (CTG4, pp. 287–324) introduces the supergame as follows. Consider an infinite sequence of repetitive plays of the non-cooperative game $\Gamma$. Suppose that the players adopt rules to determine their choice of strategies in each play of $\Gamma$, where the rules may be conditioned on the other agent's past play. The supergame is the *non-cooperative* game of choosing rules by which $\Gamma$ will be played. Aumann then identifies the cooperative solution of $\Gamma$ with a non-cooperative (strong Nash) equilibrium of the supergame. Many of the powerful developments in the theory of repeated play over the following fifty years are suggested in this

paper. Harsanyi (CTG4, pp. 325–56) proposed a generalization of Nash's bargaining theory to n-person cooperative games. Kemeny (CTG4, pp. 397–406) sounds the call for more informative priors: "While I agree that with the information usually given for n-person games no more can be said [than the von Neumann Morgenstern solutions], it seems to me that we must ask for more information" (CTG4, p. 398). But the editors respond, "The difficult question, then, is what more to assume." (CTG4, p. 11). Kemeny adds an index of the bargaining power of each agent and builds his (relatively informal) solution concept around that.

Thomas Schelling's *The Strategy of Conflict* appeared in 1960, though some chapters had been published earlier. A major motivating factor in the book is the game-theoretic analysis of the repressed conflict between the United States and the Soviet Union, which was probably then near its peak of intensity. The book is usually remembered for the focal equilibrium concept (to which we will return) but more systematically the book explores the insight that in non-cooperative games "the power to constrain an adversary may depend on the power to bind oneself . . ." (SC, p. 22[8]) with a voluntary sacrifice of freedom of action. This idea is inherent in non-cooperative games. Nash had written (Nash, 1953, p. 130) "Supposing A and B to be rational beings, it is essential for the success of the threat that A be *compelled to carry out his threat* T if B fails to comply. Otherwise it will have little meaning. For, in general, to execute the threat will not be something A would want to do, just of itself" (italics added). This element, that compels the agent to carry out threats as well as promises, distinguishes a cooperative game for Nash; its absence distinguishes a non-cooperative game. Schelling's book is firmly non-cooperative in its approach, but more consciously so than much previous work, and does not simply take the non-cooperative approach as given but argues for it (for example, SC, pp. 23–5, 115–18, 123–50). He also points out other implications that seem to have been overlooked before: signaling theory (SC, p. 24) deserves mention in particular.

Coordination games and coordination problems are also a major concern in the book. Coordination problems arise in games with two or more Nash equilibria. Luce and Raiffa had given examples including the famous Battle of the Sexes Game (GD, pp. 90–94; SC, p. 286n.) The game is shown as Table 3.3. Clearly the Nash equilibria are at the upper left and the lower right. The interest of both is in avoiding confusion that might leave them in i, II or ii, I. "What the players need is some signal to coordinate strategies; if they cannot find it in the mathematical configuration of the payoffs, they can look for it anywhere else" (SC, p. 294).

There is also a mixed-strategy Nash equilibrium, but although symmetrical, it is inferior as it imposes a 50 percent chance that both lose 1.

*Table 3.3 Game 3.2: the Battle of the Sexes*

| Payoff order: A, B | | B | |
|---|---|---|---|
| | | I | II |
| A | i | 2,1 | −1,−1 |
| | ii | −1,−1 | 1,2 |

Luce and Raiffa note, however, that if the two agents can communicate, they can arrive at a correlated strategy solution, flipping a coin and assigning a 50 percent probability to i, I and to ii, II, and zero to i, II and ii, I. Schelling's question is, however, whether they cannot coordinate their strategies even without communication, although perhaps sacrificing symmetry. A major theme of the book is that people are actually quite good at doing this. This is a focal-point solution with "some characteristic that distinguishes it from the surrounding alternatives . . ." (SC, p. 111). The most famous example Schelling gives is the example (from his classroom experiments) of two Yale students who have to rendezvous in New York but have not agreed on the place. Most chose the information booth at Grand Central Station (SC, p. 54n.)

Schelling's contributions cannot be contained within game theory, and the focal point idea pre-dated game theory. According to Schelling's account,[9] it arose in a student cross-country road trip about 1940 when the travelers were briefly separated. They began to think through how separated travelers might get together in a big American city in general, at that time, and decided that they could go to the general delivery window of the main post office – at 12 noon, of course.

Schelling's thinking was also a reflection of his experience in international relations and bargaining, and the concern for practical applications in these areas recur throughout the book. Even more than Luce and Raiffa, Schelling's book is mathematically informal, and he expresses some doubt that mathematical analyses are always useful rather than confusing (SC, pp. 10, 113–15, 164). Conversely, some game theorists who value mathematics highly dismiss Schelling's work as unimportant. A major objective of Schelling's book, expressed in the title of Part II, was "A Reorientation of Game Theory" by which game theory would be thought of as "A Theory of Interdependent Decision" (SC, p. 81). Aumann (who shared the 2005 Nobel Memorial Prize in Economics with Schelling) used that phrase to denote the subject matter of game theory in his address to the Game Theory Society as outgoing (founding) president of the society, (see also Aumann and Dreze, 2005), so this reorientation must be seen as successful.

## 3.4 TWO THEORIES, COOPERATIVE AND NON-COOPERATIVE

In 1964, Nutter proposed a non-cooperative analysis of duopoly price competition in the Bertrand-Edgeworth tradition. Drawing on the growing interest in the Prisoner's Dilemma, Nutter argues that even in a duopoly, price competition is a dominant strategy equilibrium. This established a link between the Prisoner's Dilemma example and price competition that was to influence thinking in economics and industrial organization for decades to come.

In 1962 Shubik proposed that the Shapley value could be used for cost accounting in a case of shared joint costs, the first of a small stream of applications of cooperative game theory to cost assignment. This will be discussed in Chapter 8 at Section 8.2.3.

In 1963, Thrall and Lucas proposed a generalization of the game in characteristic function form, the partition function form. This form assigns a value to each coalition in a way that depends on the other coalitions that are formed, but in such a way that the value of a coalition can be different depending on the other coalitions that form. This innovation had little impact on cooperative game theory.[10] Thrall and Lucas did not present it as an alternative that could resolve the questions that had been raised about the characteristic function. Instead they followed von Neumann and Morgenstern in resolving the partition function to a characteristic function by using the assurance value. Making their theory a direct generalization of that of von Neumann and Morgenstern was a reasonable research strategy in 1963, but reintroduced the very points that had been raised against that theory. Moreover, Thrall and Lucas suggest no method by which the value of a coalition imbedded in a partition could be assigned, for example, from a representation of the game in normal form. Perhaps for these reasons, there were only a handful of extensions and applications of their theory before the 1990s. All of this justifies the neglect of partition functions by the authors of textbooks and many other cooperative game theorists, but the partition function will be crucial for Chapters 11–16 of this book.

Developments in cooperative game theory during the 1960s were important at the time and remain important for our purposes. In Aumann (CGT4, pp. 287–324), already referenced, and elsewhere, Aumann, Davis, Maschler, and Schmeidler developed a number of new solution concepts for cooperative games: bargaining sets (Aumann and Maschler, 1964), the kernel (Davis and Maschler, 1965), and the nucleolus (Schmeidler, 1969). Some of these topics will recur in Chapters 11–16.

In a number of papers Shubik, often in collaboration with Shapley,

developed the theory of Edgeworth market games that he had described in the 1959 paper referenced above. Debreu and Scarf (1963) and Scarf (1967) also made important contributions, clarifying the relation of the core of a market game in characteristic function form to competitive equilibrium.

Exchange games are market games in which there is no production, but each agent begins with an endowment of two or more kinds of goods and coalitions may be formed for exchange. There are no externalities.[11] Following Shapley and Shubik (1952) and Shubik (CGT4, 1959) these analyses mostly adopted the nontransferable utility approach. A typical result was that the competitive equilibrium is always within the core of an exchange game, and on some assumptions (for example, Scarf, 1967) the core shrinks in such a way as to approach the competitive equilibrium as the number of participants increases. Difficulties arose in introducing production, variable returns to scale, and externalities into the model. With production, the core tended to be empty, that is, there would be no allocations that could satisfy the criteria of the core. However, Telser (1978) made that a basis for a theory that we will return to in Chapter 8. On the whole, there has been less evident interest in the core analysis of market games since the 1970s. Shapley and Shubik would extend their model to the case of externalities in 1969. This contribution will be reconsidered in Chapter 10.

*Advances in the Theory of Games* (Dresher et al., 1964) contained several papers to which reference has already been made. One other that deserves mention at this point is Selten's (ibid., pp. 577–626) exploration of cooperative solutions for games in extensive form. This paper includes a discussion of the principle of backward induction, although Selten finds it inconsistent with other, more imperative properties for cooperative solutions (pp. 582, 596). He also points out – in a note added in print – that to be complete, cooperative game theory needs the assumption that all agents can commit themselves to particular (contingent) strategies. This is a response to Schelling (1960) and is by contrast with non-cooperative games in Schelling's treatment.

Experimental work on non-cooperative game theory had been undertaken in the 1950s and before, some of which we have noted. Rapoport and Chammah (1965) reported a very large study of experiments centered on the *Prisoner's Dilemma*, the title of the book, a study that was to be influential. Their conception of rationality was relatively open (p. 13), and their experimental protocols required a long (but finite) series of repeated plays by the same experimental subjects, who were students. While the theory of perfect equilibrium in repeated play had not been developed, they recognized the argument that collusive agreements would "unravel" back

from the last to the first repetition (Rapoport and Chammah, 1965, p. 29). Accordingly, their observations that the non-cooperative strategies were not played in any large majority of trials was seen as being inconsistent with the non-cooperative equilibrium theory. In one provocative finding they discovered that female subjects cooperated less than males (Rapoport and Chammah, 1965, p. 191). In a discussion of further experiments that might be tried, they speculated about the role a tit-for-tat strategy rule might play (Rapoport and Chammah, 1965, p. 207). On the whole, experimental studies of the period similarly indicated that non-cooperative game theory was not a strong predictor of empirical results.

In the late 1960s, Harsanyi (1967–68) introduced Bayesian reasoning into game theory in a series of three papers in *Management Science*. In 1972, Aumann and Maschler discussed some examples that raised doubts about the extensive use of behavior strategies in place of contingent strategies, which had already become common. In 1972, the *International Journal of Game Theory* was founded. In the same year biologist Maynard Smith (1972) introduced the concept of evolutionarily stable strategies.

In 1973 and 1975, respectively, Gibbard and Satterthwaite published highly influential papers on the manipulation of voting schemes, in which they relied on non-cooperative game theory. These will be discussed in detail in Chapter 7, at Section 7.3. A large literature of studies both of voting systems and of implementation of cooperative and normative objectives in terms of non-cooperative equilibria has arisen from these contributions.

The 1970s were a particularly productive period for Robert Aumann, whose contributions in this period bear comparison with those of Nash around 1950. In 1973, he pointed up some difficulties with the theory of monopoly in cooperative games. In 1974, Aumann and Dreze extended and consolidated the analysis of cooperative solutions for games in coalition function form with arbitrary coalition structures (that is, partitions into distinct coalitions). In motivation for their study, Aumann and Dreze raised questions about the superadditivity assumption, which we will revisit in Chapter 9, at Section 9.1. This paper is the source of most of the subsequent literature on coalition structures (Greenberg, 1994) but much of the subsequent literature on coalition structures does not follow Aumann and Dreze in allowing for non-superadditive games.

In 1974, Aumann also addressed subjective probabilities and correlated strategies. The theme of this paper is that it makes a difference if different players have different subjective estimates of probabilities of events, in the spirit of the saying "that's what makes a horse race." Aumann finds that even simple examples in non-cooperative game theory must be modified to allow for these possibilities. In 1976, however, he was to argue that if two individuals have "common knowledge" of any event, which must include

common prior probabilities of the event, then their posterior probabilities could not disagree – that is, "agreeing to disagree" could make no sense.

Nevertheless, Aumann's 1974 paper has been influential in another way. This paper is sometimes credited with originating correlated strategies in non-cooperative games, but Aumann makes no such claim, writing (p. 70) "it has been in the folklore of game theory for years. I believe the first to notice this phenomenon (at least in print) were Harsanyi and Selten (1972)." As we have seen, the phenomenon was in fact reported in Luce and Raiffa (1957), a book which is the source of a great deal of game theory folklore. However, Aumann extends the concept with an example in which correlated equilibria can support a non-cooperative equilibrium that dominates any linear (probability) combination of Nash equilibria, and this insight is the source of a stream of subsequent work on correlated equilibria. This will be discussed in detail in Chapter 5.

## 3.5 THE TURN TOWARDS NON-COOPERATIVE GAME THEORY

In 1975, Selten (CGT, pp. 317–54) re-examined the concept of perfect equilibrium that he had introduced in a paper first published in German in 1965. This paper focuses on a refinement of Nash equilibrium called the trembling hand equilibrium. Selten had introduced subgame perfect equilibrium in his 1965 paper, but this publication brought it to the English language audience, and subgame perfection has probably been the more influential concept. This paper is regarded as the beginning of the literature on "refinements" of Nash equilibrium. Selten's model will be discussed in Chapter 6 at Section 6.1.

In 1976 Myerson defined an extension of the Shapley value to games in partition function form. Myerson gives an example with negative externalities (p. 26) but that is weakly superadditive. He assumes that the grand coalition will ultimately form, with supperadditivity presumed. This paper seems to be the origin of the modest stream of subsequent research on superadditive games in partition function form.

In the late 1970s, work by Myerson and Maskin established implementation theory, also known as mechanism design theory, following a program proposed by Hurwicz (1973). This work was honored by a Nobel Memorial Prize in 2007. See especially Maskin (1999) and Myerson (1979; 1986). The objective of implementation theory is to design a game so that its *non-cooperative* equilibria correspond to the *cooperative* or other normative outcome that is desired. Implementation theory will be discussed in Chapter 7.

The development of Selten's conception of perfect equilibrium made possible some important progress on what has become known as the "folk theorem" in game theory. The "folk theorem" is the idea that, for games such as the Prisoner's Dilemma (with very bad non-cooperative results in one-off play) repeated play might lead to a cooperative outcome in some circumstances. As early as 1981, however, in a working paper of the UCLA department of economics, Fudenberg and Levine (1981, p. 19) sketched an analysis of repeated play of the Prisoner's Dilemma in terms of perfect equilibria. A few years later Fudenberg and Maskin (1986) gave the general analysis that has now become standard. A quite different but related approach to repeated play in non-cooperative games emerged with Axelrod's (1981; 1984) computational studies. Coding simple rules for the selection of behavior strategies in repeated Prisoner's Dilemmas, Axelrod played the rules one against another in a tournament, and found that tit-for-tat[12] (a trigger strategy in which one plays cooperatively until the first defection by the other player, but responds with a single round of non-cooperative play) did relatively well against a wide array of challengers. As much as the folk theorem work, this study contributed to the emergence of tit-for-tat and other trigger strategies as standard tools for understanding repeated play of non-cooperative games.

In 1984 Bernheim introduced the concept of rationalizable strategies; a simultaneous paper by Pearce (1984) shared the innovation. One important departure in this paper is that Bernheim allows players to condition their decision rules on conjectures about the conjectures that others may make about them. This leads, in some cases, to a much larger set of equilibrium strategies. In his Nobel Address, Aumann (2005) was to admit conjectures as to the *rules other players might use in selecting behavior strategies* among the conditions of a choice of strategies, with a further extension of the range of possible non-cooperative equilibria in repeated games.

When we combine rationalizable strategies and correlated equilibrium, the case for Nash equilibria as predictors of behavior is very much reduced. If the game is played one-off, then players are not likely to have enough information to exclude non-Nash rationalizable equilibria, and the same will be true in the first plays of a repeated game. For later plays of a repeated game, though, correlated equilibria may emerge, and these, too, may be non-Nash. In an evolutionary model, where players are randomly matched to play one-off but can learn from the experience of one another, evolutionarily stable (Nash) equilibria seem a reasonable prediction. Even here, though, boundedly rational learning might result in correlated equilibria. More generally, where the Nash equilibrium in pure strategies is unique (including, but not limited to, the family of social dilemmas) correlated strategy equilibria can be excluded; and if in addition the Nash

equilibrium is subject to some stringent stability conditions (Bernheim, 1984, p. 1020) then Nash equilibria are the only rationalizable strategies. All in all, Nash equilibria can no longer be treated as "solutions" to non-cooperative games, but only as candidate solutions and as tools that may be useful in finding other (for example, correlated equilibrium) solutions.

In 1988, Harsanyi and Selten offered a framework to resolve the growing family of refinements of Nash equilibrium, suggesting a hierarchy of criteria for choosing among Nash equilibria. They rank the equilibria in terms of relative stability, so that, for example, Pareto-dominant equilibria are considered more stable than those that are not Pareto-dominant but are risk-dominant.

In 1989, the journal *Games and Economic Behavior* was founded.

## 3.6 BEHAVIORAL GAME THEORY

In the 1990s and 2000s, advances continued to be made in the topics of non-cooperative and cooperative game theory that had come to be traditional, and some research pursued new directions that will be useful for this book. The period was, of course, dominated by the Nobel Memorial Prizes of 1994, 2005, and 2007, which kept the traditional topics in view. In 1990, Greenberg proposed a "theory of social situations" as an alternative both to cooperative and non-cooperative game theory. Some progress was also made on the incorporation of externalities in games in partition function form (for example, Zhao, 1992; Chwe, 1994; Ray and Vohra, 1999; Carraro, 2003; Koczy, 2007). Returning to the long-neglected topic of coalitions in non-cooperative games, Bernheim et al. (1987) proposed a property of coalition-proofness as a refinement of Nash equilibrium. An outstanding development in this period is the maturation of behavioral game theory.

Traditional game theory proceeds from strong assumptions about human rationality to strong conclusions about the nature of equilibrium. One can ask whether either the assumptions or the conclusions are empirically valid. If we find evidence that they are not, and attempt to rebuild game theory with more "realistic" assumptions about rationality, we are entering the sphere of *behavioral game theory*. The more traditional studies based on those strong assumptions will be called *classical game theory.*

Indeed there is a long history of experimental studies in game theory, some of which have been mentioned in context. Social psychologists and others quite early provided evidence that people facing a Prisoner's Dilemma-like game do not always act as neoclassical maximizers. (Lave, 1965; Rapoport and Chammah, 1965; Morehouse, 1967; Kreps et. al.,

1982, among many others). For an argument that game theory ought nevertheless be based on strict rationality, see Morgenstern and Schwödiauer (1976). Some scholars suggest that even the successful trials are attributable to training effects (Marwell and Ames, 1981; Carter and Irons, 1991). But game theoretic equilibria also gain some experimental support, especially in their evolutionary interpretation (for example, Cooper et al., 1990; Van Huyck et. al., 1990). Mailath (1998) surveys evolutionary game theory, to determine the extent to which it may support the predictions of Nash equilibrium in particular, indicating that "Evolutionary game theory has provided a qualified answer . . . In a range of settings, agents do (eventually) play Nash" (p. 1348). However, he also indicates the limits of this range.

Some of the early experiments on the Prisoner's Dilemma were interpreted as evidence that altruism is an element in human behavior. Unfortunately, altruism is not always well-defined. Altruism was inferred, however, from a tendency to choose the cooperative strategy even when it is not a best-response strategy, for example, in Prisoner's Dilemma games. More recent studies have often focused instead on reciprocity. Berg et al. (1995, p. 139) say their ". . . results suggest that both positive and negative forms of reciprocity exist and must be taken into account . . . [and] provide strong support for current research efforts to . . . integrate reciprocity into standard game theory…" Positive reciprocity means that players respond to generous behavior generously, even at a sacrifice to themselves; negative reciprocity means that they retaliate against aggressors even when it makes them worse off to do so.

One game that has been extensively studied in the experimental literature is the "Ultimatum Game" (for example, Henrich et al., 2005). The Ultimatum Game is a two-person game along the following lines: the two agents may be able to share a fixed amount, such as $100. The first agent, the proposer, suggests a payment to go to the second agent, the responder. If the responder accepts the payment, he receives it, and the balance is paid to the proposer. However, if the responder rejects the payment, neither agent gets anything. The non-cooperative equilibrium is one in which the proposer makes the smallest possible positive offer and the responder accepts it. However, experimental evidence disagrees with this prediction. If the proposer makes a very small offer, the responder is sometimes observed to reject the proposal despite sacrificing the small positive payment. Moreover, offers are often more than the minimum needed to avoid a rejection, and 50-50 offers are fairly common.[13] This and other experimental games have given rise to their distinct specialized literatures.

In all, the experimental evidence does not support many of the predictions of non-cooperative game theory. This is not to say that it supports

any of the cooperative solutions in any systematic way, either. We must suppose that real human beings are both more complex and less accurate calculating machines than classical game theory supposes. Solution concepts based on strict rationality can define hypotheses as to attractors and stable points in dynamic models with boundedly rational learning. On that score the more recent experimental evidence is not merely negative. Models based on strict rationality also define the base from which deviations from rationality are predicted.

## 3.7 BRIEF SUMMARY

In 45 years from 1944 to 1989, game theory became a cross-disciplinary study of great importance for the mathematical social sciences. It also became a compound field – not one study of interdependent decisions, but largely separate studies of non-cooperative and cooperative game theory, a situation decried by Aumann in his inaugural presidential address to the Game Theory Society (2003).

What game theory offers is a tool-kit applicable to decision problems in which the consequences of one decision may depend on the decisions of others, previous decisions creating the conditions for current decisions, simultaneous, and subsequent decisions, with or without mutual knowledge, with or without some degree of honest mutual commitment to a common strategy. If we choose our tools to fit the job and disregard the dogmas and dichotomies of cooperative and non-cooperative, superadditive valuations and perfect or ideal rationality, we will find that the tools contribute to the solution of problems of real-world public policy.

## NOTES

1. Kyu Uck Lee, personal communication by e-mail, 22 June 2007.
2. In what follows, page citations indicated by TGEB will refer to von Neumann and Morgenstern (2004).
3. This paper seems to have been available only in German prior to the publication of the translation in Tucker and Luce (1959).
4. In what follows, citations to CGT will refer to Kuhn, *Classics in Game Theory* (1997).
5. In what follows, page references indicated by CTG4 will refer to volume 4 of *Contributions to the Theory of Games*, edited by Tucker and Luce (1959).
6. In what follows, references denoted GD will refer to Luce and Raiffa (1957).
7. Von Neumann and Morgenstern had alluded to this in TGEB, p. 224.
8. In what follows, page references to SC refer to Schelling, *The Strategy of Conflict* (1960).
9. This paragraph relies on an address given by Schelling at Trinity University, San Antonio, Texas on 18 April 2007, and is from memory.

10. One evidence of this is that it is not mentioned in some recent advanced texts with coverage of cooperative game theory. See, for example, Peleg and Sudhölter, *Introduction to the Theory of Cooperative Games* (2003), and Forgo et al., *Introduction to the Theory of Games* (1999).
11. This is an assumption of fact, implicit in the standard neoclassical model of preferences. In that model, each person has preferences over his own consumption of goods and services and not over the consumption of others. Suppose instead that Agent 1 prefers imputations in which other agents' consumption of wine is zero to those in which it is positive. Exchanges that increase the number of agents who consume wine would then impose a negative externality on the singleton coalition comprising Agent 1. Such "busybody" preferences create difficulties for neoclassical economics (Sen, 1970).
12. According to the *Oxford English Dictionary*, the phrase tit-for-tat is traceable to the sixteenth century phrase "tip for tap," meaning, roughly, push for shove.
13. For example, Guth et al. (1982), Henrich et al. (2005), and note also Roth and Erev (1995), Stanley and Tran (1998), Roth et al. (1991), Andreoni and Blanchard (2006), Oosterbeek et al. (2004).

# 4. Nash equilibrium and public policy

The best-known ideas in game theory are within non-cooperative game theory, and probably the single best-known example in game theory is the Prisoner's Dilemma, a non-cooperative example. This example shows how interactive self-interested decisions may lead to results that are less favorable to all participants than some other outcome would be. The Prisoner's Dilemma example can be generalized to a class of non-cooperative normal form games known as "social dilemmas" (Dawes, 1980) that share similar broad qualities. From the pragmatic point of view, non-cooperative game theory provides powerful tools for the identification and specification of problems, as the social dilemmas exemplify. On the whole, moreover, non-cooperative game theory is a relatively settled, mature study. Social dilemmas are a class of Nash equilibrium models, and Nash equilibria are well understood and the foundation of most applications of non-cooperative game theory. However, there are some unsettled issues and some other proposed approaches to the solution of non-cooperative games. This chapter will review a number of Nash equilibrium models with a view to their applicability to public policy studies.

## 4.1 SOCIAL DILEMMAS

While the Prisoner's Dilemma is the best-known example in game theory, it is also one of the simplest, and its simplicity does place some limits on its application.

### 4.1.1 Symmetrical Dilemmas

The Prisoner's Dilemma begins with a story of interrogation. For this discussion, we may instead recall the Water Game from Chapter 2, where it is shown in normal form as Table 2.1.

Eastland knows that it cannot influence Westria's strategy choice, and conversely. Instead, each one chooses his best response to the strategy choice made by the other. This defines *Nash equilibrium*. Moreover, in this case, the best response is "Divert" regardless of the other agent's strategy choice. That means "Divert" is a *dominant strategy*: by definition, if a

strategy is the best response to any strategy choice made by the other agent or agents, it is called a *dominant strategy*. Thus, when both agents choose the dominant strategy "Divert," we have a *dominant strategy equilibrium*, which is a particularly simple instance of a Nash equilibrium. A dominant strategy equilibrium can be defined as a Nash equilibrium in which each agent has a dominant strategy.

Nevertheless, if both agents were to choose "Don't" in Game 2.1, both would be better off, with net payoffs of 0 rather than −1. We may borrow terminology from welfare economics and say that the strategy pair "Don't, Don't" *Pareto-dominates* the pair "Divert, Divert." A strategy vector $\Sigma_1$ Pareto-dominates strategy vector $\Sigma_2$ if no agent is worse off with $\Sigma_1$ than with $\Sigma_2$ and at least one agent is better off with $\Sigma_1$ than with $\Sigma_2$. Together, these observations define Game 2.1 as a *social dilemma* (Dawes, 1980). Generally, a social dilemma is a game in which (1) there is a dominant strategy equilibrium indicated by strategies $\Sigma_2$, and (2) there is vector of strategies $\Sigma_1$, such that each component of $\Sigma_1$ differs from the corresponding component of $\Sigma_2$ and $\Sigma_1$ Pareto-dominates $\Sigma_2$. Social dilemmas are usually also treated as being symmetrical (so that interchanging any two agents would leave the payoff table unchanged).

A social dilemma model such as Game 2.1 predicts that, in the absence of some public intervention, the dominant strategy equilibrium, $\Sigma_2$, will occur. Since it is Pareto-dominated by a different set of decisions, $\Sigma_1$, this outcome is *inefficient*. Decisions $\Sigma_1$ are said to constitute a *cooperative solution* and, in a symmetrical game such as this, the payoffs correspond to the Shapley value and nucleolus in particular. (These will be discussed in more detail in Chapters 8, 12 and 13.) Public policies may then be advocated that move individual decisions toward the efficient set $\Sigma_1$. In this way, social dilemmas capture the principles that seem to underlie a number of major problems of modern societies and public policy, but they may not be very good descriptions of the real world. The symmetry that is usually assumed in social dilemma models is one example. In most real-world applications, there is likely to be some lack of symmetry among the agents. Since the problem (inefficiency) arises in symmetrical models, however, we can be assured that it does not arise from the lack of symmetry in the real world; thus asymmetry is a complication but not an underlying cause of the problem. This is a valuable point that might be missed if the simplified, symmetrical model were not considered. All the same, for some practical applications, it may be necessary to reintroduce some asymmetry in a model with heterogeneous agents.

Social dilemmas can be generalized to a large number of players following Schelling (1978) and Moulin (1982, pp. 92 *et seq.*). Think of a large number of people living in the watershed of a lake, each of whom may act

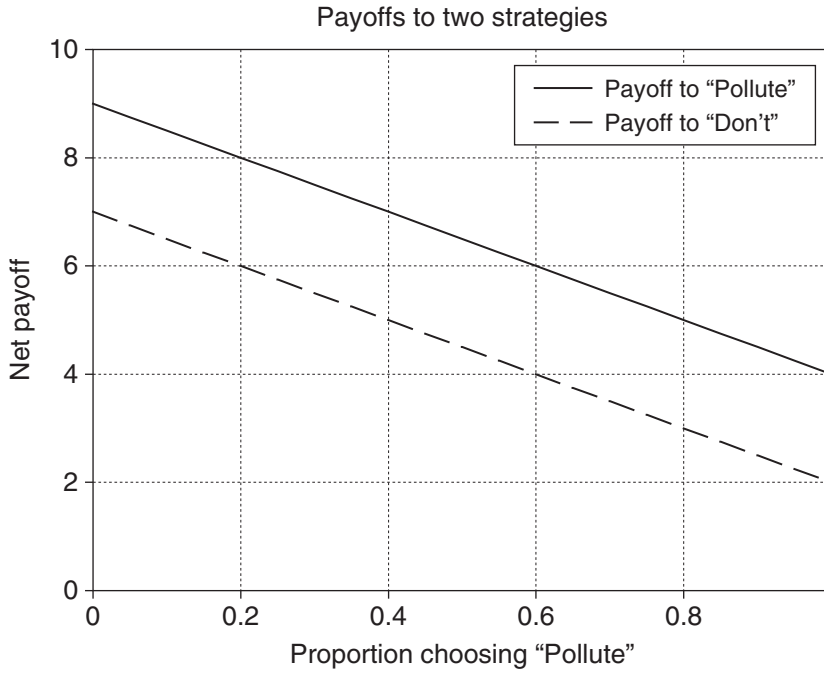


*Figure 4.1 A social dilemma with large N*

so as to pollute the lake or, at some cost, refrain from pollution. Suppose there are N agents, N very large, each of whom must make the same absolute choice of strategies "Don't" or "Pollute." The overall amount of pollution will depend on the proportion of the population that choose "pollute;" so that the payoffs to both strategies will depend on the same proportion. Borrowing terminology from the theory of differential games, we can describe the proportion of agents who choose "Pollute" as a *state variable* for the game. In this usage, a state variable is a variable that is sufficient to determine the payoffs of the different strategies without any other information (such as information on the specific strategy choices of individual agents, for example).

This model is illustrated by Figure 4.1. We see that the payoff to "Pollute" lies above the payoff to "Don't," regardless of the proportion of the group who choose "pollute" as their strategy. The diagram illustrates visually that this is an N-person social dilemma. If any group of players chooses "Don't," they are not choosing their best response to the strategies chosen by the others. The dominant strategy equilibrium corresponds to the rightward extreme of the diagram, the case in which every agent chooses "Pollute."

The N-person social dilemma model can also be interpreted to be consistent with imperfectly rational behavior, if it is interpreted in an evolutionary sense (see, for example, Aumann, 1997). Suppose that individuals usually act with inertia, simply choosing the same strategy over and over, but from time to time, at random, they experiment with reversing their strategies. If the reversal leads to an increase in the net payoff they persist, and if not they return to the previous strategy. This trial-and-error learning process is one of random *variation* and directed *selection* of strategies, a simple evolutionary process. Biologists, having borrowed the concept of Nash equilibrium from game theory, define an evolutionarily stable strategy (ESS) as a Nash equilibrium that is stable under an evolutionary dynamics. The dominant strategy equilibrium for this model is an ESS. Thus the conclusion does not depend on the assumption of perfect rationality.[1]

An appropriately generalized social dilemma model can account for many instances in which inefficiencies persist in the presence of human decisions that successfully seek self-regarding benefits, whether through perfect rationality or through trial-and-error learning. As this example and the Water Game suggest, environmental economics is largely built of social dilemma models. The production of a public good is another instance of a social dilemma. All in all, social dilemma models are powerful diagnostic and explanatory tools for problems of social inefficiency.

### 4.1.2 The Special Case of Price Competition

This may seem a bit strange to a person whose knowledge of game theory is derived from a textbook of microeconomic principles. In microeconomics price competition leads to efficiency, and price competition is non-cooperative behavior; cooperative behavior (collusion) is inefficient and is a danger to be avoided. But price competition is very much a special case, which mixes cooperative and non-cooperative elements right from the start.

A discussion of price competition begins by assuming that a relatively small number (but greater than one) of coalitions called "business firms" make offers to sell some product. Their strategies may be the offer prices (the case known as Bertrand-Edgeworth competition) or the quantities offered (Cournot competition). The buyers are not usually modeled as individual agents but treated as an undifferentiated mass of demand. Implicitly, however, the buyers are treated as acting as singleton coalitions, and the strategy of each buyer is the quantity she or he chooses to buy. If, then, the sellers' behavior is characterized by non-cooperative

Bertrand-Edgeworth competition and the singleton buyers act non-cooperatively, and if there are no externalities, the outcome is efficient. (If seller competition is Cournot, then the outcome approaches efficiency as the number of sellers increases and their sizes approach equality.) On the other hand, if the sellers form a coalition and act cooperatively among themselves, restricting output and raising the price, the outcome is inefficient.

Nevertheless, price competition is not really a non-cooperative game, since the business firms are coalitions based on legally enforceable agreements; and because exchange is itself a cooperative activity. To use Brandenburger and Nalebuff's (1997) expression, this is "co-opetition." Price competition can best be understood as a hybrid solution (Zhao, 1992) for a game with a coalition structure (Aumann and Dreze, 1974). In principle, the grand coalition of all producers and consumers could obtain the efficient outcome, but free-market economists believe (plausibly) that this degree of centralization is nonfeasible. If so, then the hybrid solution of price competition may be our best hope for efficient markets.

The key lessons to be drawn are that there is no real contradiction between the recommendation of price competition in microeconomics and the recommendation of cooperative solutions in game theory in general, and that the results of efficient market theory cannot be generalized to other cases without strong independent justification for doing do so. The neoliberal attempt, in the later twentieth century, to organize all social action on the basis of price competition may ultimately be seen as no less utopian than the attempt, earlier in the twentieth century, to organize social action on the basis of a communistic grand coalition.

### 4.1.3 Other Dilemmas, Nash Equilibria and Public Policy

Social dilemmas seem to be at the basis of many environmental and other problems, but there is a tendency to overuse the social dilemma model. This is confusing because there are many models based on the Nash equilibrium that are not social dilemmas (because some players do not have dominant strategies or because asymmetry is important, for example) but in which the Nash or other non-cooperative equilibrium is inefficient.

Let us consider another example to do with riparian rights and water supplies. For this example, the two players are two municipalities in the same river valley. The Town of Upstream is, as the name suggests, upstream from the Boro of Downstream. Upstream can obtain its water supply by damming Modest River at a point well above Downstream, but to do so would impair Downstream's ability to meet its own water needs from the Modest. Geography is such that Downstream could build either one or two dams (or none). Although two dams would be costly, together

*Table 4.1 Game 4.2: water works*

| Payoff order: Downstream, Upstream | | Upstream | |
|---|---|---|---|
| | | One dam | No dam |
| Downstream | One dam | 3,3 | 5,1 |
| | Two dams | 2,3 | 4,4 |
| | No dam | 1,5 | 1,1 |

they could supply both towns, provided the Upstream dam is not built. Game 4.2 is a payoff table based on those ideas, with the benefits from water provision rated on a scale of 1 to 5 for each town (see Table 4.1).

For Game 4.2, the Nash equilibrium is for each town to build one dam. We may observe that for the Boro of Downstream, building one dam is a dominant strategy: regardless of the strategy chosen by Upstream, building one dam gives higher payoffs than any other strategy. On the other hand, Upstream does not have a dominant strategy: either one strategy or the other may pay best, depending on the strategy chosen by Downstream. However, Upstream can make some reasonable judgments about the strategy Downstream will choose. Upstream knows (after all) that "two dams" and "no dams" are dominated strategies, and also that Downstream is a rational decision-maker. Thus, Upstream can draw the reasonable conclusion that Downstream will never choose those dominated strategies. They can be left out of consideration, since they will not affect the game. Once they are dropped out, we have an equivalent game in which Downstream has only undominated strategies, and the only one is "one dam." In that smaller game, the strategy of building one dam is dominant for Upstream. We then eliminate the dominated strategy for Upstream, and are left with only one strategy combination: "one dam," "one dam." This is an instance of the *iterated elimination of dominated strategies.* A strategy is said to be *strictly dominated* for Player i if it yields a payoff to i that is greater than the payoff to any other strategy, regardless of the strategies that other players choose, and is said to be *weakly dominated* if it yields a payoff no less than that from any other strategy i may choose. When we eliminate the strictly dominated strategies for a player, creating a reduced game, and repeat that so long as it is possible, we are applying *iterative elimination of (strictly) dominated strategies*, and if that procedure yields a unique strategy for each player, those strategies correspond to a unique Nash equilibrium.

As in a social dilemma, payoffs to both players for the Nash equilibrium are dominated by the payoffs that result if Downstream builds two dams and supplies Upstream from its surplus water. Nevertheless this is not a

social dilemma. As we have seen, Upstream does not have a dominant strategy. For Upstream the best response is to build a dam if Downstream builds less than two. And this makes a difference. In Game 4.2, we have assumed that both decisions are made simultaneously, with neither decision-maker aware of the decision the other has made or will make. Suppose instead that one of the decision-makers can commit himself by making the first move. In that case the second mover can choose contingent strategies, using his knowledge of the first-mover's decision. In the social dilemma this makes no difference, since each decision-maker's best strategy is the same no matter what the first mover may do. In Game 4.2, by contrast, if Downstream is first mover we will see a different outcome. Downstream can anticipate that, if they build no dam or one dam, Upstream will build its dam, leaving the Downstream with a payoff of 3 at most; while if Downstream builds two dams and supplies both towns, Upstream will not build its dam, giving Downstream a payoff of 4. For Downstream as first mover, the strategy of building two dams becomes a best response to the contingent strategy "if one or fewer Downstream dams, then build, else do not" on the part of Upstream.

It is important, then, not to generalize the social dilemma too hastily. Not all social problems are social dilemmas, and that may make a difference in the attainable solutions.

## 4.2 RANDOMIZATION OF STRATEGIES

In the examples in this chapter so far, each decision-maker has to choose among a finite number of "strategies," where each strategy is a description of a course of action, with or without a description of the contingencies in which that action will be taken; that is, either behavior strategies or contingent strategies. These "strategies" are often called "pure strategies," but in this book the word "strategies" will mean pure strategies unless it is indicated otherwise. Indeed, thus far in the chapter, the number of strategies is quite small and decisions are simultaneous so that contingencies will not need to be specified. In each case so far, there is at least one Nash equilibrium in pure strategies. However, even in games with small numbers of strategies and simultaneous decisions, there may not always be a Nash equilibrium if we limit ourselves to pure strategies.

We can illustrate that with an example that was historically important, in that it captured Oskar Morgenstern's attention and led to his collaboration with von Neumann. The example is from the Sherlock Holmes story, "The Final Problem." Holmes is attempting to escape from Moriarty and from England to the continent, and Moriarty is attempting to capture

*Table 4.2 Game 4.3: "The Final Problem" in strategic normal form*

| Payoff order: Moriarty, Sherlock | | Sherlock | |
|---|---|---|---|
| | | Canterbury | Dover |
| Moriarty | Canterbury | 1, −1 | −1, 1 |
| | Dover | −1, 1 | 1,−1 |

and murder Sherlock. Sherlock is on the train to Dover and must decide whether to stay on the train and cross to Europe via Dover or to get off at Canterbury and go to Europe by a different route. Moriarty must decide whether to continue to Dover and try to intercept Holmes on the coast or to leave the train at Canterbury in the hope that he will find Sherlock there. If Moriarty can choose the same stopping point as Holmes, then Holmes is caught; otherwise Holmes can escape to the continent. Assigning a payoff to Moriarty of 1 if he catches Sherlock and −1 otherwise, and to Sherlock of 1 if he escapes and −1 otherwise, we have the game in normal form shown as Game 4.3 in Table 4.2.

This game has no Nash equilibria in the pure strategies "Canterbury" and "Dover." If Moriarty knows that Sherlock will detrain at Dover, then he also will detrain at Dover, but in that case Sherlock's best response is Canterbury, though if Sherlock gets off at Canterbury, Moriarty's best response is to do that too – and so on! In short, there are no two strategies each of which is a best response to the other: no Nash equilibrium in pure strategies. More generally, if either Sherlock or Moriarty acts predictably, he is likely to lose out, as the other person can use that predictability to defeat him. Indeed, the best that either one can do is to choose between the two strategies at random, assigning a probability of one-half to each strategy. If (for example) Sherlock deviates from that fifty-fifty probability, choosing Dover with a probability greater than one half, he loses out on the average, since Moriarty can get off at Dover with a better than even chance of catching Sherlock.

When a player chooses between two strategies at random, according to probabilities that prevent the other player from exploiting his predictability, this is called a *mixed strategy.* There are infinitely many mixed strategies in a nontrivial game, and Nash showed that every game in normal form has at least one Nash equilibrium, which may be a mixed strategy equilibrium. Nash equilibria in mixed strategies can be computed by linear programming methods in general, or, in simple cases, by the algebra of simultaneous equations. Details of computation will be beyond the scope of this chapter.

*Table 4.3 Game 4.4: Terrorist vs. Defender*

| Payoff order: Defender, Terrorist | | Terrorist | |
|---|---|---|---|
| | | Target 1 | Target 2 |
| Defender | Target 1 | 3,0 | 0,3 |
| | Target 2 | 0,3 | 3,0 |
| | Both | 2,1 | 2,1 |

In situations of conflict, it is often in the interest of each party to act unpredictably. The Normandy Invasion, Game 2.6, provides another good example. It illustrates an important point: it is not necessary literally to flip a coin to decide between two strategies. The probabilities that matter are subjective probabilities. As a player in the game, my objective is to manipulate the subjective probability estimates my enemy assigns to my own actions, so as to prevent my enemy from exploiting the predictability of my decisions. The web of costly deception and strategic maneuvers surrounding the Normandy invasion, as described in Brown (1975), *A Bodyguard of Lies*, illustrates this.

Let us consider one further example in that vein. The players in Game 4.4 will be a terrorist and a defender.[2] The terrorist has the capacity to attack target 1 or target 2, but not both and no other target. The defender can prevent an attack on target 1 or target 2 by appropriate defensive measures, or, at somewhat greater cost, can protect both. Complete success for either contestant is recorded as a payoff of 3. If the defender defends both targets, the defender's payoff is reduced to 2, because of the cost of doing so, and this partial success for the terrorist is recorded as a payoff of 1. Notice that the sum of the payoffs is 3 in every cell of the table (Table 4.3). Although this is not a zero-sum game, it is a constant-sum game, and just as with a zero-sum game, this means the objectives of the players are totally opposed.

Suppose, then, that the terrorist chooses a mixed strategy, assigning probability ½ to each target. Then the defender's expected value for defending a single target is

$$\frac{1}{2} \cdot 3 + \frac{1}{2} \cdot 0 = 1\frac{1}{2}.$$

Then the defender has no reason to defend one target in preference to the other: to that extent the mixed strategy has done its job. By defending both targets the defender can have an expected value of 2. Accordingly, the defender assigns probability 1 to defending both targets. By randomizing

their strategy the terrorists have forced the defender to the expense of defending both targets, gaining a minor victory by doing so.

Mixed strategy equilibria in peacetime public policy problems may be uncommon, but the possibility should not be overlooked. Whenever we see decision-makers creating doubt about their strategies, we will need to explore the possibility of a mixed strategy equilibrium. In relations between local governments and large businesses or major athletic teams, for example, it may well be that the uncertainty the firms and teams create about their locational decisions represent a mixed strategy.

## 4.3 COORDINATION AND ANTICOORDINATION GAMES

We may think of the game as a mathematical problem and the Nash equilibrium as a solution. In that perspective the Nash equilibrium in pure strategies has two kinds of shortcomings. First, as we have seen, some games may not have Nash equilibria in pure strategies. When we allow mixed as well as pure strategies, however, that problem disappears: as a matter of mathematical fact, every game in strategic normal form has at least one Nash equilibrium. (As a practical matter, though, randomized strategies may have to be excluded in some special case applications, and in that case the difficulty returns.) The other shortcoming is that there may be more than one Nash equilibrium. The right number of solutions, from a mathematical point of view, is exactly one. Indeed some critics of game theory have made that shortcoming the basis of a claim that the rational action model of human behavior (as expressed in Nash equilibrium) is simply a failure. But that is a hasty conclusion. On the one hand, the plurality of solutions may reflect the conditions of the real world, rather than a failure of mathematics. In that case we would not want a solution that, however perfect mathematically, assumes away the facts of the real world. On the other hand, we may treat the multiplicity of solutions as a problem to be solved, not by the theorist but by the "players in the game," and inquire how in fact people have contrived to solve it. This proves to be a rich field of inquiry, one we will undertake in the next chapter.

A class of games that illustrates both these points is the so-called coordination games. Examples can be drawn from highway traffic. As usual, we may begin with a simple two-person game in which the two agents must each choose between two strategies. The two agents will be motorists approaching one another on a road, and the strategies they must choose between are "drive on the left" and "drive on the right." If they make the same choice, then they pass one another safely. If they make opposite

*Table 4.4 Game 4.5: the fender-bender game*

| Payoff order: First, Second | | Second car | |
|---|---|---|---|
| | | Left | Right |
| First car | Left | 1,1 | −1,−1 |
| | Right | −1,−1 | 1,1 |

choices, then they both lose in the resulting fender-bender. The game is shown in strategic normal form as Game 4.5 (Table 4.4). The payoff numbers are arbitrary: 1 for a safe passage and −1 for a collision.

In this game there are two Nash equilibria in pure strategies: "left," "left" and "right," "right." In addition there is a third Nash equilibrium, a mixed strategy equilibrium in which each motorist chooses randomly with probability one-half for each strategy, and the expected value of the payoff for each motorist is zero. Clearly, the pure strategy equilibria are superior to the mixed strategy equilibrium or to any non-equilibrium state. (They are superior in the sense that the pure strategy equilibria are Pareto-dominant over the others.) Now suppose that each motorist knows *nothing* that has not already been given as part of the game. Then each may reason as follows: "Using the principle of insufficient reason, I must assign equal probabilities to the other person's strategy choice. Therefore I have nothing to lose by also choosing my strategy at random." This points to the inferior mixed strategy equilibrium as the most likely one, and that is indeed an ugly dilemma (although not a "social dilemma").

When we assumed that each motorist knows nothing that has not already been given as part of the game, we assumed a great deal. In particular we assumed that the drivers do not know whether they are driving in North America, England, India, or Europe. In all countries it is *customary* to drive on one side or the other, and, knowing the custom, each agent can make a rational judgment that the other will (with very high probability) drive on the customary side. The result is that fender-benders in this situation are actually quite rare. The familiar fact that the custom can be quite different in different countries – left in England, right in the USA – reflects the fact that both are pure-strategy Nash equilibria. What seems a shortcoming from the mathematical viewpoint proves to be a principle with explanatory power in this application.

Of course, there are also legal standards in most countries *requiring* people to drive on the customary side. But the customs are mostly self-enforcing, and the function of the laws is mainly to assign responsibility when the custom fails and collisions do happen. Simple as this example is, it suggests

reasons why customs, much as they may vary from country to country, can be very stable where they exist, and why the law may be most effective when it reinforces custom and may be largely futile when it goes against custom.

It will be useful as well to view this example in an evolutionary light. Suppose that we have a large population of motorists who are randomly matched to pass one another on the road in a large series of matches. Suppose also that motorists choose their strategies according to some boundedly rational rule of thumb. One possible rule of thumb is conformism: "Do what you see others doing." Another possible rule of thumb is the stick-or-switch rule: "If the strategy yields a positive payoff, stick with it; otherwise switch." Either of these rules will lead to a rapid convergence to the unanimous choice either of left or right, though we cannot predict which.

Games of this kind are called *coordination games*, since the problem faced by the two players is to coordinate their strategies and thus avoid a bad outcome. Following local custom is one instance of the kind of solution suggested by Thomas Schelling (1960) and is sometimes called a "Schelling point" or a "focal equilibrium."

Now let us consider another two-by-two game, again involving two motorists. In this case, however, the two motorists are approaching an intersection, and their strategies are to stop and let the other go through, or to go ahead. There are four possible outcomes: a fender-bender if both go, a waste of time if both stop, and two outcomes in which one is slightly delayed and the other passes without delay when they choose opposite strategies. This is shown as Game 4.6 (Table 4.5), with the payoff numbers assigned arbitrarily (as usual) to represent better and worse outcomes from the different points of view of the two motorists.

In some ways this game is very much like the coordination game: it has two Nash equilibria in pure strategies and a third, mixed strategy equilibrium, in which each driver chooses between the two strategies with probability ½. The expected value of the mixed strategy is $-½$, so the mixed strategy equilibrium is Pareto-dominated by the pure strategy equilibria. Once again, though, with no information but what is contained within the

*Table 4.5 Game 4.6: the intersection game*

| Payoff order: First, Second | | Second car | |
|---|---|---|---|
| | | Go | Stop |
| First car | Go | −2,−2 | 1,0 |
| | Stop | 0,1 | −1,−1 |

game, the two motorists face a problem in choosing strategies that will avoid the bad outcomes. In this case, though, they have to choose *opposite* strategies, and for this reason games of this kind have come to be known as *anticoordination* games.

In some ways, though, this is a more difficult problem. In a coordination game, coordination can be accomplished when both motorists receive the same signal, as with a custom that all motorists drive on the right. For this case, though, the two motorists would have to receive different signals, or a signal complex enough that they would interpret it in opposite ways. For the same reason, a simple evolutionary process may not lead to the pure strategy Nash equilibria in this case. Suppose that a large population of motorists are randomly matched to play the two-person game, again and again. Clearly, a rule of conformism will not lead them to make opposite decisions. The mixed strategy equilibrium proves to be evolutionarily stable, since any tendency for more than half of the motorists to choose one strategy rather than the other just makes coordination of the strategy choices less likely! Anticoordination games are a problem to which we will return in the next chapter.

## 4.4 CONSISTENT CONJECTURES

In a non-cooperative game, the problem for an individual decision-maker is that the consequences of his own decision depend also on the decisions of others. One way to express this is to say that, in order to estimate the consequences of his own alternative decisions, an individual must first make a conjecture as to what decisions the others will make. Each individual's decisions will depend (among other things) on his conjectures as to the decisions of the others. Now suppose that, for each decision-maker, the conjectures he makes lead him to act in just such a way as the others had conjectured that he would act. Thus, all of the conjectures prove to be correct! In that case, the decisions of the group, taken together, have the property of *consistent conjectures*.

The concept of consistent conjectures does not originate in game theory but in the theory of industrial organization (Bresnahan, 1981). In industrial organization theory, the decisions of concern are (principally) the pricing strategies of firms in oligopolies. But the concept can be applied to non-cooperative games more generally (Fudenberg and Levine, 1999). To choose his best strategy, each "player in the game" must make a conjecture as to the strategies chosen by the others, and select his best response to the strategies conjectured. If each chooses the strategy the others had conjectured, then we might call that set of conjectures and strategies a

consistent-conjectures equilibrium, CCE. Now, it will be evident that a Nash equilibrium in pure strategies is a CCE.

In coordination and anticoordination games, consistent conjectures are sufficient for efficient action and a cooperative outcome in the game. The problem is to increase the probability that conjectures will be consistent. In a coordination game, a common custom achieves this. The common belief that "in North America, it is safest to drive on the right hand side of the road" proves to be a true belief – not because it is a truth in any metaphysical sense, but because the actions that people choose, on the basis of the belief, make it a true belief. In a social dilemma, by contrast, consistent conjectures are not at all sufficient for efficient and cooperative action, and in fact correspond to an inefficient outcome. This is an important distinction for public policy, since in the first case, the case of coordination games, the efficient arrangement is self-enforcing, while in the second case, the social dilemma, enforcement must play the main role in achieving an efficient cooperative outcome.

## 4.5 COALITIONS IN NON-COOPERATIVE GAMES

In non-cooperative games, there are no enforceable agreements. Nevertheless, when there are two or more Nash equilibria, coalitions may form and may make a difference. Consider Game 4.7, of conflict among three nations (Table 4.6), with the following assumptions: Country *a* is the strongest of the three, and capable of projecting both land and sea power; Country *b* is landlocked, and thus unable to influence the balance of power at sea; while country *c* has fine harbors but indefensible land borders, and so can influence the balance of power at sea but not on land. Each of the three countries must decide between a forward and a defensive military posture. The defensive position is cheaper and less likely to lead to war. The payoffs to the three countries are calibrated so that each receives

*Table 4.6 Game 4.7: conflict among three nations*

| Payoff order: Country *a*, *b*, *c* | | *c* | | | |
|---|---|---|---|---|---|
| | | Forward | | Defensive | |
| | | *b* | | *b* | |
| | | Forward | Defensive | Forward | Defensive |
| *a* | Forward | 1,1,1 | 7,3,1 | 7,1,3 | 8,3,3 |
| | Defensive | 2,6,6 | 4,4,6 | 4,6,4 | 5,5,5 |

a payoff of 5 when all choose cheap defensive postures, and each receives a payoff of 1 when all choose costly and risky forward strategies. However, when some choose forward strategies and others defensive, the country with the forward strategy can benefit. This can favor countries *b* and *c*, however, only if both choose forward strategies simultaneously: if only one of the countries does so, country *a* can concentrate its forces against that country and deprive them of any benefit from their enterprise.

This game has two Nash equilibria, one at the upper right where country *a* has a forward posture and *b* and *c* are defensive, and one at the lower left where country *a* maintains a defensive posture against forward postures by countries *b* and *c*. To say that the upper right cell in the table – strategy combination "forward," "defensive," "defensive" – is a Nash equilibrium is to say that no country can benefit by unilaterally deviating from it. Thus a shift by *a* would reduce its payoffs from 8 to 5; a shift by *b* would reduce its payoffs from 3 to 1, and similarly a shift by *c*. But if *b* and *c* were to make a coordinated shift from defensive to forward postures, country *a*'s best response would be the defensive posture at the lower left, the other Nash equilibrium. Thus, the lower left equilibrium is *strong* in the terminology of Aumann (CGT4, pp. 287–324), and it is also *coalition-proof* in the terminology of Bernheim et al. (1987), while the equilibrium at the upper right is neither.

Strong equilibria and coalition-proof equilibria differ in detail, but both reflect the stability of the equilibrium against the formation of coalitions. If an equilibrium could be disrupted by a coordinated shift of strategies by some group of players, the members of the group are better off as a result *and the shift is to strategies that correspond to another Nash equilibrium,* then the first Nash equilibrium is rejected as unstable. Notice that the alliance between *b* and *c*, in the example, needs no enforcement – Nash equilibria are self-enforcing, and either country can only lose by deviating from it. That is why the phrase in italics, *"the shift is to strategies that correspond to another Nash equilibrium,"* is crucial: otherwise the new situation could not be sustainable without some enforcement. A *strong* Nash equilibrium is one that cannot be disrupted in this way, either because it is unique or because no group can benefit by a coordinated shift to another equilibrium. A *coalition-proof* equilibrium is one that either is strong or, if not, nevertheless is unlikely to be disrupted in this way, because any group shift to a new Nash equilibrium would lead to an equilibrium that would itself be unstable, in that a subgroup of the original group could benefit, at the expense of the rest, by a further coordinated shift.

This idea seems to have particular relevance for relations among sovereign states, as in the example. Sovereignty means that there is no enforcement of agreements, so relations among sovereign states are essentially

non-cooperative. Nevertheless, treaties and alliances can be effective and rather stable among sovereign states. This example suggests a reason why they can: the treaty or alliance corresponds to a Nash equilibrium, but there are other Nash equilibria that might otherwise be realized, which would be less advantageous to the allies or treaty partners. At the same time, and more crucially, the example indicates the limits of this possibility: if the terms of the treaty or alliance do not correspond to a Nash equilibrium, then in all probability they will not be kept.

As conventional solutions to a coordination game, alliances may be persistent. And in the real world, of course, there is more to it than this symmetrical model. A new alliance could well create irreversible changes in political and other circumstances that could make it impossible to go back to the old equilibrium. But we will have to defer any further speculation along these lines until Chapter 6, on the game in extensive form. (For an intuitive discussion of a real historical example, see McCain, 2004, pp. 231–2, 245.)

## 4.6 REFINEMENTS

In addition, Game 4.7 illustrates a common problem of non-cooperative game theory. When there are many solutions, how are we to choose among them? In Game 4.7, we excluded the equilibrium at the upper right on the grounds that a coalition of two countries could improve on it – it is neither strong nor coalition-proof. This would be an instance of a *refinement* of the Nash equilibrium. In general, a refinement is any assumption additional to the definition of the Nash equilibrium that allows us to reduce the number of Nash equilibria considered as solutions. We have already mentioned, in passing, another very important refinement: an evolutionarily stable strategy is a Nash equilibrium that is stable with an evolutionary dynamics.

In some games, we may find that one Nash equilibrium is better for *all* players than another equilibrium: one equilibrium Pareto-dominates another. Certainly the dominated equilibrium would not be strong, but this dominance condition would give us an even more persuasive argument to rule it out. The fact remains that every pure-strategy Nash equilibrium has the consistent conjectures property – so that if each agent really believes that the other agent will choose a strategy that leads to the dominated equilibrium, he will do best to choose the corresponding strategy. Moreover, there are many proposed refinements of Nash equilibrium, and in some cases they conflict. Despite the large literature on refinements of Nash equilibrium, there is no refinement that assures us of a unique solution in

every case, nor that is applicable in every case. As we will see in later chapters, some refinements may be very important in particular cases.

## 4.7 EVOLUTIONARY GAMES

We have made passing references already to evolutionary interpretations of Nash equilibrium, primarily as a way of generalizing two-person interaction where the real interest is in the interaction of a large number of agents. Evolutionary models are an important branch of non-cooperative game theory with broader implications.

As we recall, the evolutionarily stable strategy (ESS) set is a refinement of Nash equilibrium, and the refinement might well be applicable to interactive decisions of human beings as well as to the evolution of species. As Friedman notes (1998), what distinguishes an evolutionary dynamics from other ways of looking at game theory is primarily a lack of foresight, in that agents do not anticipate or attempt to influence the future evolution of the decisions of others. Beyond that, the evolutionary dynamics might be linked to an otherwise completely rational decision process. However, one of the advantages of an evolutionary perspective is that we might instead approach game theory from the point of view of *bounded rationality*.

In neoclassical economics and classical game theory, individuals are supposed to be rational in the sense that they maximize their utility, profits or payoffs. This view has never been without critics, and Nobel Laureate Herbert Simon was one of the most important and widely recognized of them. The phrase "bounded rationality" stems from him, and reflects his judgment that the maximization of utility, profits or other payoffs requires decisions that are far too complex to be within the cognitive capacity of real human beings. Instead, human rationality is *bounded* by the computational capacity of the human mind and brain. The typical *rational* activities of real human beings are the setting of targets and the search for alternative activities that are satisfactory according to those targets. This has been expressed by saying that real decision-makers do not maximize but satisfice. To many economists, this is Simon's critique in a nutshell; but Simon was more than a critic and is known outside economics as one of the founders of artificial intelligence theory. In the tradition of artificial intelligence founded by Simon (with his collaborator Alan Newell: Newell and Simon, 1972; note also Simon, 1995), rational decisions are made by *heuristic rules*. These rules do not necessarily lead to maximization of profits or payoffs or anything else, but do lead to "satisfactory" results. Artificial intelligence envisions that real decisions are based on a large set of rules, a "knowledge base" or "rule base," not (as a rule) a single rule

per decision – although a simple rule like "do what the boss says (if he is watching)" can play a great role in decision-making and the organization of human action. In general, though, even a naive decision-maker will have to decide what rule to apply – before deciding what to do – and the final decision can draw on a large number of rules working together. Experts in a field will have still larger and more complex rule bases for their expert decisions. Think, for example, of a medical doctor's decision whether to recommend surgery to a patient: rules for the diagnosis of the condition, other rules that reflect the experience of the medical community about the usual results of alternative treatments for patients in particular categories, rules for the judgment of patient psychology and so on will all be brought to bear on the decision.

What perhaps needs stress is that bounded rationality is *not irrationality*. It is *rationality*, conceived in a way that reflects a realistic view of the limits of the human brain as a computer. (This is the way it is viewed by Simon and his followers if not by neoclassical economists.) Nevertheless it raises questions about the assumption of Nash equilibrium theory that agents infallibly choose the best response to the strategies chosen by others. Instead, we may link the theory of bounded rationality with that of evolutionary dynamics in game theory, yielding a more cognitively realistic non-cooperative game theory (for example, Aumann, 1997; Gintis, 2007).

The evolutionary bounded rationality approach to game theory would begin by assuming bounded rationality. At a particular moment, then, agents choose their strategies according to a given set of heuristic rules. Some of these rules may generate strategies that better approximate best responses to the strategies chosen by others, and some generate strategies that approximate best responses less well. However, people learn, both from their own experimentation and the imitation of others, so that over time the rules that lead to poor strategies will be replaced by other rules that do better. In this way, decisions would tend, over time, to approximate the best-response decisions. Not all Nash equilibria will be stable with this sort of process. Those that are stable will be ESS and we can predict, tentatively, that actual decisions will be approximated by the ESS decisions.

In this view, the heuristic rules play the role in social evolution that genes play in organic evolution. (They are the "replicators;" Hodgson, 2002.) The pure strategies in the game play the role that individual plants and animals play in organic evolution: they are the interactors, and their interaction determines (via the payoffs that result) the tendency for the corresponding replicators (heuristic rules) to be replicated as a larger or smaller proportion of the population of rules as a whole. Thus it is the rules and decisions that evolve in this evolutionary view.

But is it *really* evolutionary? Those who have a turning to philosophy may

question whether this sort of scheme really ought to be called evolutionary or whether evolution, which by definition advances no purpose or intention, can be applied to the decisions of human beings, decisions that are directed (within the limits of bounded rationality) to better realize the purposes and intentions of the human individuals. These issues are better avoided, so let us say that the scheme that combines ESS with bounded rationality is an *adaptive* game theory, based on the assumption that real human rationality is bounded but adaptive (compare Selten and Gigerenzer, 2001).

## 4.8 CONCLUSION

We find that a number of problematic cases in public policy can be traced to inefficient Nash equilibria with the games considered in strategic normal form. In this sense, we may say that non-cooperative game theory is a powerful diagnostic tool for public policy. For now we reserve judgment as to whether it may also be helpful in prescription. Nash equilibrium models have proved a powerful tool of problem identification for public policy. While the Prisoner's Dilemma has commanded the central position in this perspective, a wide variety of other Nash equilibrium models may result in inefficient equilibria in the absence of some public action. In conflict situations, and some others, it may be rational for agents to be unpredictable, and to randomize their strategies; but in other circumstances, coordination and anticoordination games, randomization may be something to be avoided, and avoiding it may require some information not within the game itself. The Nash equilibrium in pure strategies has a property of consistent conjectures that helps to explain how a Nash equilibrium, once established, can be persistent even when other Nash equilibria may be more efficient. Extensions of these models, with large numbers of players, trial-and-error adaptive learning in place of ideal rationality, and lack of symmetry among players, may seem more "realistic" than the simpler Nash equilibrium models. Where there are two or more Nash equilibria, refinements may eliminate some as less plausible. Pareto-dominance, evolutionary stability, and resistance to disruption by coalitional shifts of strategy provide criteria for refinement in particular cases.

## NOTES

1. For a case study see Hamilton et al. (2008).
2. I am indebted to my colleague Richard Hamilton for discussions that contributed to this example.

# 5. Correlated equilibrium

While Nash equilibrium is the central concept in non-cooperative game theory, and has many applications, it is not quite the whole story. There are rival solution concepts and applications that are prescriptive rather than diagnostic. This chapter will discuss a major alternative: correlated strategy equilibrium.

A few years ago in New Zealand (Bray, 2003), telecommunications companies Teamtalk Ltd and MCS Digital Ltd were embroiled in a lawsuit. If both were to pursue their claims in a court of law, the legal fees would be great enough that both would be worse off. If one knew for certain that the other would pursue his claim, then the best response would be to abandon the claim, to avoid the legal costs. However, neither was certain that the other would withdraw, nor was willing to be the one to withdraw unilaterally. They agreed to settle the difference by arm-wrestling, the winner to take the asset and the loser to abandon his claim. On the face of it, this may seem an irrational procedure, but on more careful consideration it is quite rational. The two businessmen had arrived at a correlated equilibrium solution to their problem.

## 5.1 INTRODUCTORY EXAMPLE AND DEFINITION

Eastonia and Westoria are neighboring townships that share a business district on their border. Each is considering building a parking garage to serve the business district, but only one garage is economically feasible, so that if both build, both will be worse off. If one town builds, then the other has the further option to (1) improve their infrastructure to make it easier for people to make use of the parking garage in the nearby town, or (2) do nothing, retaining the reduced fees for on-street parking. Option (1) will have some cost but will capture a small part of the benefits of the parking structure in the form of increased business traffic. Most of the benefit of the parking garage will flow to the town that builds it, though. This is shown as Game 5.1, with 10 indicating the maximum benefit and other payoffs reflecting the assumptions above (Table 5.1).

This game has two pure strategy equilibria, each where one town builds and the other improves its infrastructure. There is also a mixed strategy

*Table 5.1 Game 5.1: parking garage*

| Payoff order: Eastonia, Westoria | | Westoria | | |
|---|---|---|---|---|
| | | Build | Improve | Do nothing |
| Eastonia | Build | −5,−5 | 10,2 | 3,1 |
| | Improve | 2,10 | 2,2 | 0,1 |
| | Do nothing | 1,3 | 1,0 | 1,1 |

equilibrium at which each town builds with probability 8/15 and improves with probability 7/15. A probability of zero is assigned to "do nothing," since any increase in the probability of "do nothing" will reduce the expected value payoff of the decision-maker. Both of the pure strategy equilibria are efficient, but the benefits are very unequally distributed. The mixed strategy equilibrium is highly inefficient, however, with an expected value payoff of 2. In the absence of any custom or other signal to support a Schelling focal equilibrium in this case, as we have seen, the mixed strategy equilibrium would seem plausible as it applies to a case of complete ignorance.

In a *cooperative* arrangement, one could build and make a side payment to the other township so that both would share more equally in the benefits. Here, though, we are concerned with *non-cooperative* arrangements. As we have seen, non-cooperative equilibria must be self-enforcing, but can be randomized. The problem with the mixed strategy solution in this case is that it assigns a positive probability (0.28) to the lose-lose outcome at the upper left, where both towns build and lose 5. It also assigns a positive probability (0.22) to the outcome in the middle, where neither town builds. Suppose instead that the decision could be randomized in a way that would assign probabilities of zero to the upper-left and middle strategy combinations. This would assure that exactly one of the towns builds.

In fact, this is pretty easy to do. The two township supervisors could get together and flip a coin, with Eastonia building if the coin comes up heads, and Westoria building if the coin comes up tails. The result would be that the expected value for each township would be 6 – very much better than the 2 that would come from the mixed strategy equilibrium, and with the benefits equally shared in expected value terms. Then the township supervisor of Eastonia is choosing according to the rule "if heads then build else improve," and the Westorian supervisor is choosing according to the rule "if heads then improve else build." These rules are self-enforcing, in that the fall of the coin provides a signal for a focal equilibrium in Game 5.1.

This solution differs from a mixed strategy equilibrium in that the

decisions are correlated. If Eastonia chooses "build" then Westoria chooses "improve" with probability 1, and if Eastonia chooses "improve" then Westoria chooses "build" with probability 1. Thus, it is called a "correlated strategy equilibrium," or more briefly, "correlated equilibrium." This concept, and much of the discussion here, is due to Luce and Raiffa (1957), although papers by Aumann (1974; 1987) have stimulated much of the interest in it. Aumann showed (1987) that correlated equilibrium can be a result of Bayesian learning, rather than conscious randomization and maximization, and several papers written in the late 1990s (Foster and Vohra, 1997; Fudenberg and Levine, 1999; Hart and Mas-Colell, 2000) introduce adaptive (boundedly rational) procedures that lead to correlated equilibrium. Thus, correlated equilibrium is a very plausible adaptation in a game like Game 5.1.

The probabilities of the two Nash equilibrium pure strategy combinations would not necessarily be 50-50. Indeed, any probability between 0 and 1 would share the same self-enforcing property in Game 5.1. Figure 5.1 shows the range of correlated strategies for Game 5.1 as the probabilities assigned to the two Nash equilibria of the underlying game vary. In general there will be very many correlated strategies, and so it seems that we have only reproduced the problem of multiple Nash equilibria (though gaining some efficiency in the process). However, the equal probabilities do supply a cognitively salient focal point that could lead both agents to expect one correlated equilibrium rather than another. Moreover, in this case the principle of insufficient reason reinforces the equiprobable correlated equilibrium.[1] As we will see, there is some casual anthropological evidence that equiprobable solutions are very common. On the other hand, if

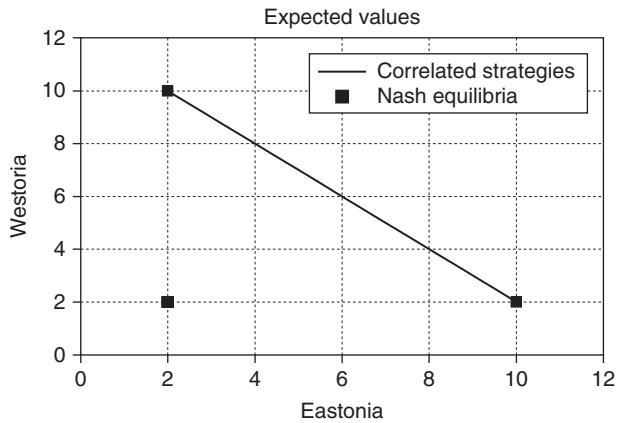


*Figure 5.1 Nash equilibria and correlated strategies in Game 5.1*

the agents have an opportunity to agree on a signal, whether it is flipping a coin or arm-wrestling, they will also have an opportunity to settle the probabilities by negotiation.

There have been few applications to public policy. However, rationing by lottery has been fairly common in human history, and the military draft and similar arrangements may be seen as correlated equilibrium solutions. However, equal treatment in terms of expected values may not be seen as ethically or politically adequate. It may be that the objectives of public policy or the ethical standards from which public policy objectives are derived make reference to actual outcomes, so that these objectives and standards cannot be satisfied by unequal outcomes, even when all individuals have the same probabilities of being advantaged or disadvantaged. More generally, if we feel that different outcomes for different people require some justfication, whether the justification is on the basis of different contributions, different needs, or different entitlements, differences resulting from a random mechanism may be seen as unacceptable. This consideration undoubtedly limits the application of correlated equilibria in public policy.

Thus, it may not be practical for the township supervisors in Game 5.1 to make their decision by flipping a coin. Yet they may nevertheless arrive at a correlated strategy solution to their interactive decision problem. Suppose they hire a consulting firm to study and compare the costs and benefits of the two alternative plans, for Eastonia to build and Westoria to improve and vice versa. Consulting the payoff table, we know that the net benefits are the same in either case, so any difference found by the cost–benefit study will be the result of errors in the study. We may suppose that the errors will be random and unbiased. Thus, when the township supervisors make their decisions on the basis of the cost–benefit study, they are carrying out a coordinated strategy solution! Of course, the cost–benefit study is likely to be a little more costly than flipping a coin. Still, it could be worthwhile as a politically respectable means of avoiding the impasse of the mixed strategy solution. We should note that the game is probably unrealistic in assuming that the two towns are perfectly symmetrical. Instead, we might want to assume that there actually are differences in the costs and benefits of the two proposals, but the town supervisors, not being specialists in cost–benefit analysis, do not know what they are. The payoffs are best-guesses, and the consulting firm is able to improve on them with more information and find real differences. However, the township supervisors are unable to anticipate that, and from their point of view, the subjective probabilities are 50-50. As Aumann has shown us, that is sufficient for the correlated equilibrium solution. When the New Zealand businessmen arm-wrestled to settle their impasse, each one probably thought that he

was most likely to win, while an outside observer's subjective probabilities would perhaps be 50-50.

What the coin flip, the cost–benefit study and the arm-wrestling do is to supply a common signal to the two decision-makers. The signal gives the decision-makers the information they need to choose one of the plural Nash equilibria in pure strategies, and the decisions are correlated so that they will in fact correspond to one of the Nash equilibria in pure strategies. In the games we have considered in this section, it is clearly in the common interest of both decision-makers to contrive such a common signal.

## 5.2 COORDINATION AND ANTICOORDINATION GAMES

Game 5.1, with its two Nash equilibria, has something in common with Games 4.5 and 4.6, the coordination and anticoordination games. Indeed, when the equilibrium of a coordination game is determined by custom, the custom might be considered as a signal that supports a correlated equilibrium. In some superspace of different possible histories, perhaps, driving on the left and driving on the right are equally probable – as witness the opposite customs in different countries. But, as we recall, anticoordination games present more difficult problems.

### 5.2.1 Stoplights as a Paradigm

The example of an anticoordination game is the intersection game – two cars approaching an intersection. Which will go through, and which will pause? Two Nash equilibria exist, where each car takes one of these roles. If they had time, the two drivers could get out and flip a coin to decide – but that would defeat the purpose of the exercise, which is to get through the intersection quickly. If the intersection is controlled by a stoplight, though, the car with the green light will go ahead and the car with the red light will stop. *This is a correlated strategy equilibrium.* The probability of getting a green light may be equal for the two, depending on when they arrive, or it may be unequal, if one of the streets is a major artery and so has longer green lights to accommodate heavier traffic; but, as we have seen, equal probability is not a requirement.

The traffic light is a fascinating twentieth-century innovation in the practice of interdependent decisions![2] It also illustrates the difference that makes anticoordination games more difficult than coordination games. For a coordination game, such as Game 4.5, the two decision-makers receive the same signal, such as a customary practice, and that is sufficient

so that they can coordinate their decisions. However, in the anticoordination game, different, correlated signals are required.

In a two-person game, given time, there is little difficulty in contriving this. However, the intersection game is a bit artificial, in that highway traffic is not really a two-person game. Rather it is a many-person game with drivers randomly matched to interact at intersections. In the individual matches, there will not be time enough for the drivers to get out and arm-wrestle. What is needed is a correlated set of signals for the entire population. The stoplights provide that correlated set of signals. But notice that the provision of common signals for this large population is a public good. It is no accident that stoplights are provided by government, although, in the early stages of motoring, some traffic direction was provided by private initiative.[3]

This defines a function of the public authority that has not been explicitly recognized, although it is implicit in some existing public activities such as the provision of stoplights and signage. As we will see it is also implicit in some aspects of economic policy. Perhaps explicit consideration of this role of the public authority will lead to innovations that can improve the results of private decisions in new ways.

### 5.2.2 Other Historical Instances

Lindow man is a famous mummy, the preserved cadaver of a Briton, seemingly a Celtic priest of pre-Roman times who was selected from among the members of a priestly community to be sacrificed to the gods and goddesses he served (Brothwell, 1987). His stomach contained the remnants of a burnt bannock. Interpreted in the light of Celtic tradition, it seems likely that he was chosen from among the community to be the sacrificial victim by the following procedure: those qualified to be the victim were required to draw a portion of bannock, a baked grain food, from a bowl without being able to see which fragment of the bannock they would take. One fragment was burnt, and the priest who drew the burnt bannock was sacrificed.

We may take it for granted that the priests felt that the sacrifice must occur; that the failure to make an appropriate sacrifice would be worse than death. At the same time we may take it equally for granted that each preferred not to be the sacrifice.[4] Thus, each faced an interdependent decision with three possible outcomes: in order of preference, another sacrificed, myself sacrificed, none sacrificed. If there are N priests this game has N Nash equilibria, one corresponding to the sacrifice of each of the priests. But how to decide among these alternatives? The priests adopted a correlated equilibrium, in which the probabilities attached to the different equilibria were roughly equal.

It may be that further discoveries will provide a different interpretation of this death. However, it will serve here to illustrate the antiquity of correlated equilibrium solutions in practice, and the tendency to choose equiprobable schemes. Drawing straws, flipping a coin, and playing scissors-paper-stone[5] are all–like drawing the burnt bannock, arm-wrestling, and having an error-prone cost–benefit study done–means of assigning subjectively equal probabilities to two or more pure strategy Nash equilibria. A Methodist minister preaches that there are many instances of casting lots to make decisions in the Hebrew and Christian scriptures, dating from the Pentateuch (Lamar-Sterling, 2006). It was the "custom of the sea" that when sailors were cast away and starving, the victim to be cannibalized was chosen by casting lots or drawing straws (Hanson, 2001). What these examples show is that correlated equilibria with equiprobable assignments are widely known to nonspecialists in game theory and have been so, probably, for thousands of years. A rigorous anthropological study that would document the cross-cultural width and temporal depth of this knowledge would be of interest but is beyond the scope of this book. In short, correlated equilibrium is not mysterious, but just the opposite. However, we will see that an example given by Aumann has difficulties that we have not yet discussed, and that example has motivated most of the discussion of correlated equilibria in game theory.

## 5.3 DIFFICULT CASES

Now let us consider another example, with Eastonia and Westoria at loggerheads once again. This time there is a proposal for a cellular-telephone tower to serve both towns. There is a location on the border between the two towns that will serve them equally, but with minor inconvenience to both. In order to build at that site, both townships will have to agree to it. If just one approves the building of the tower, it will be built at a site within the approving township, and that township will be greatly inconvenienced. The other township will then get service of a quality almost as good as that from the borderline site, without any inconvenience at all. If neither approves, no tower is built, and cellphone reception will remain poor in both towns. As usual we will assign payoff numbers that agree with those assumptions so far as their relative magnitude is concerned, and not worry much about real units of measurement. Then we have Game 5.2 (Table 5.2).

This is a game discussed by Aumann (1974) and has been the source of most research on correlated equilibria. (Aumann refers only to "player 1" and "player 2" and offers no application in which the problem might

*Table 5.2 Game 5.2: a cellular telephone tower (Aumann's game)*

| Payoff order: Eastonia, Westoria | | Westoria | |
|---|---|---|---|
| | | Approve | Reject |
| Eastonia | Approve | 6,6 | 2,7 |
| | Reject | 7,2 | 0,0 |

arise, but the numbers have been chosen to agree with Aumann's; in any case they have the right relative magnitudes for our cell-telephone tower story.) Aumann's game has pure strategy Nash equilibria at the lower left and the upper right, and a mixed strategy equilibrium with probabilities 2/3, 1/3. The expected payoff of the mixed strategy equilibrium is 4 2/3, 4 2/3. Aumann writes (p. 72) "Consider now an objective chance mechanism that chooses one of three points A, B, C [that is, upper left, lower left, and upper right] with probability 1/3 each. After the point has been chosen player 1 is told whether or not A was chosen, and player 2 is told whether or not C was chosen; nothing more is told to the players." Then (p. 73) ". . . the random device . . . is not at all difficult to construct. Given a roulette wheel, it is easy to construct electrical connections that will do the job." Aumann's example can be restated in terms of a more recent electronic technology, along the following lines.

We might construct a computer interface so that four possible events occur with probabilities determined by the program. The events are colored shapes shown on the screen: red triangle, red square, green triangle, and green square. The probabilities are 1/3, 1/3, 1/3, 0. Call the first three events, with positive probabilities, $E_1$, $E_2$, $E_3$. These probabilities are known to the players. However, the two players see separate screens, and Eastonia observes only the *shape* – triangle or square – while Westoria observes only the *color* – red or green. Thus, neither player actually knows what event has occurred. However if (for example) Eastonia observes a triangle, one event can be ruled out, namely red square; that is, he can infer "not $E_2$." Similarly Westoria can observe red and infer "not $E_3$."

Suppose Westonia is known to choose according to the rule "if red then approve, otherwise reject;" that is, "if not $E_3$ then approve, otherwise reject." Let us call this contingent rule "rule A." Eastonia is considering adopting the rule "if triangle then approve, otherwise reject;" that is, "if not $E_2$ then approve, otherwise reject." Let us call this "rule B."

Case 1: suppose also that Eastonia observes square. He can infer that Westoria has observed red with probability 1 and so Westoria will choose approve, in accordance with rule A. In that case Eastonia's best choice is

"reject" for 7 rather than 6. Case 2: suppose Eastonia observes triangle. He can infer that red square has not occurred (not $E_2$), but also that Westoria may play either accept or reject with equal probabilities, so that the conditional probability is ½. Again, this follows rule A. In this case the expected value of approve (that is, of following rule B) is $1/2(6 + 2) = 4$, while the expected value of reject is $1/2(7 + 0) = 3.5$. So Eastonia follows rule B and chooses approve. It follows that rule B is a best response to rule A, and similar reasoning establishes the converse.

If the decision-makers coordinate their decisions in this way, probabilities of one-third each are assigned to the upper left, lower left, and upper right cells, and zero to the lower right cell. The overall expected value for each player is $1/3(7 + 6 + 2) = 5$, so that this contingent strategy dominates the mixed strategy Nash equilibrium. Suppose instead that probabilities ½ had been assigned to the two pure strategy equilibria as in a stoplight equilibrium, and zero to "approve, approve." Then the expected value for each town would be 4 ½ – so the correlated equilibrium can do better than simply a weighted sum of the pure strategy Nash equilibria. This is Aumann's key conclusion.

Rules A and B provide a set of strategies (contingent on the private signals from the two computer screens) that are best responses to one another, thus need no enforcement; are not Nash equilibria because the strategy choices of the agents are not independent but partially correlated; and are more efficient than any Nash equilibrium or probability mixture of Nash equilibria. It is a correlated strategy equilibrium, like the ones we have seen in coordination and anticoordination games; but the game is not a coordination or anticoordination game and is more difficult in two senses.

First, in Aumann's game, the correlated strategy equilibrium is not efficient. A slight increase in the probability of "approve, approve" at the expense of the other two (Nash equilibrial) outcomes will increase the expected value payoffs of both players. However, there is a limit to how far this can be taken. If either township is certain that the other will approve the project, then its own best response is to reject. In Game 5.2, if the probability of the red triangle ("approve, approve") is more than one half, then rules A and B are no longer self-enforcing. And the correlated equilibrium mechanism can assure the two players an expected value no greater than 5.25 each, whereas an enforceable cooperative agreement would assure them of 6 each. For this reason, also, signals must not be public, but private, and cannot be perfectly correlated, but must be imperfectly correlated. Put otherwise, the effectiveness of the correlated equilibrium mechanism in this case requires that each agent be at least somewhat ignorant of the other's plans – but ignorant only to just the right degree! The provision of private, imperfectly but appropriately correlated signals

to different agents in a game is a difficulty that does not arise for coordination or anticoordination games.

## 5.4 SUNSPOT EQUILIBRIA AND ECONOMIC POLICY

Stanley Jevons (1835–82) was one of the great founders of neoclassical economics, and also pioneered statistical work in economics and made contributions to symbolic logic, including a prototype computer (New School History of Economic Thought Website, 2007). Before he became an academic economist and philosopher, he worked as a geologist and was an amateur meteorologist. He proposed that sunspots might be a cause of business cycles. By the late twentieth century "a sunspot" came to refer to any fluctuating quantity that might seem to have a correlation with business activity but could have no causal influence. A literature of the late twentieth century suggested that such "sunspot" variables might nevertheless partly determine economic activity.

By this time it was generally known that "rational expectations" equilibria in macroeconomics would not in general be unique, but that, indeed, a macroeconomic model could have many equilibria. In overlapping-generation models Azariadis (1981) and Cass and Shell (1983) showed that agents might correlate their decisions with a "sunspot" variable, so that a "sunspot equilibrium" might be observed – even though the sunspot would have no causal effect on any real economic variable. A quite large literature followed, of which the following are just a few suggestive contributions. Prima facie, sunspot equilibria seem to resemble correlated equilibria in noncooperative game theory. However, Maskin and Tirole (1987), with a fixed-horizon model, obtain largely negative results, in that the conditions for a correlated equilibrium to correspond to a sunspot equilibrium are quite limited. On the other hand, Peck and Shell (1991) provide an example with imperfect competition in which the sunspot equilibria are correlated equilibria of a market game. Chatterjee et al. (1993), in a two-sector overlapping-generations model, argue that complementarity can result in multiple equilibria and fluctuating sunspot equilibria. The following example is suggested by, but not necessarily an instance of, these contributions.

We will begin with a two-person game of market entry. Firms A and B can each choose between immediate entry (go) and postponement of entry to the next period (postpone). While their operations are complementary, in that each supplies a cheap or highly effective input to the other, their technologies are different. Firm B uses a roundabout method that will be

*Table 5.3 Game 5.3: complementary market entry*

| Payoffs: A, B | | Firm B | |
|---|---|---|---|
| | | Go | Postpone |
| Firm A | Go | 10,5 | 0,0 |
| | Postpone | 8,8 | 5,10 |

more effective in the second period, after a preparatory phase in the first period, so postponement by both firms favors Firm B, and this will be expressed by payoffs 5,10, where the first payoff is to Firm A. Firm A relies on a wasting resource which is more available in the first period, so the case in which neither firm postpones is favorable to Firm A, expressed by payoffs 10,5. If B goes ahead with entry and A postpones, B's first period production increases the stock of input to Firm A that offsets the wasting of its resource that would otherwise occur, so both benefit to some extent from complementarity in both periods and this is expressed by payoffs 8,8. However, if B postpones and A does not, neither benefits from complementarity in any period, expressed by payoffs 0,0. The resulting Game 5.3 is shown as Table 5.3.

This game has three Nash equilibria: pure strategy equilibria in the upper left and lower right, and a mixed strategy in which A assigns a probability of 0.714 to postponement and B assigns a probability of about 0.286 to postponement. In this mixed strategy equilibrium, each firm has an expected value payoff of 7.14.

Now, suppose that the presidents of both firms observe sunspots, and that the probability that sunspot activity is greater than its mean is just 0.5. Suppose each businessman believes that when sunspot activity is above mean, times are good for new entering firms, but when sunspots are below average it is best to postpone entry. Then both enter with probability ½ and both postpone with the same probability, ½. In that case, they have a correlated equilibrium with expected value payoffs of 7.5. Moreover, a market for financial securities, that is "contingent claims," can play a role in this correlated equilibrium. Suppose Firms A and B sign a contract that specifies that A will pay 2.5 to B if sunspot activity is greater than average, and B will pay 2.5 if not. Then, by playing the correlated equilibrium, the two agents each obtain a payoff of 7.5 with certainty.

We might make this a basis for a larger game. Suppose that at each time $t = 1, 2, \ldots, q$ new firms of each type come into existence. The firms are matched in complementary pairs, perhaps because of different locations, with payoffs determined as in Game 5.3. Now suppose that above-average

sunspots in periods $t$ and $t + 1$ are followed by below-average sunspots in period $t + 2$. Firms that have entered in period t cease to exist in period $t + 2$, and the potential new firms in period $t + 2$ postpone their entry until period $t + 3$. Thus, $2q$ firms of each type are active in period $t + 1$, but period $t + 2$ is a recession period with only $q$ firms of each type active. Now suppose that sunspots are again above average in period $t + 3$. Potential new firms from both periods $t + 2$ and $t + 3$ enter in period $t + 3$, and we have a recovery, with the number of active firms returning to $2q$ of each type. Thus we observe "business cycles" correlated with sunspots although sunspots have no causal influence on economic variables.

But it is possible to improve on the sunspot equilibrium in this instance. Suppose instead that a disinterested third party gives each firm an instruction either to enter or to postpone entry. The third party randomizes instructions, instructing both to enter with probability ¼, both to postpone with probability ¼, and A to enter and B to postpone with probability ½. The firms are also told to keep their instructions confidential. Suppose Firm B is instructed to enter without postponement. Computing the conditional probability that Firm A also will enter without delay as $(\frac{1}{2})/(\frac{1}{2} + \frac{1}{4}) = \frac{2}{3}$, Firm A's expected value from following the instruction is 7, while the expected value of acting against the instruction, that is postponing entry, is 6.7. Thus Firm B's best response is to follow the instruction. Moreover, Firm B has an incentive to keep his instruction confidential, as he is also instructed, since if he were to reveal that he is to enter immediately, A's best response would be to do the same, reducing B's payoff to 5 with certainty. Suppose B is instructed to postpone. In that case, he has nothing to lose either by following the instruction or by keeping it confidential, since he knows certainly that Firm A's instruction is also to postpone. Firm B can realize his maximum payoff by following this instruction without revealing it and revealing the instruction will not improve on that.

Suppose Firm A is instructed to postpone entry. Computing the conditional probability for Firm B's instructions as before, Firm A also finds that its expected value is greater from carrying out the instructions than from deviating from them; and moveover that she is better off to keep the instruction confidential, since if Firm B knew with certainty that firm A would postpone, Firm B would postpone, reducing Firm A's expected value from 7 to 5.

All in all, then, the third party's randomized instructions are self-enforcing, and since the overall result of following instructions is an expected value payout of 7.75, the instructions Pareto-dominate both the previous correlated equilibrium, with a public "sunspot" signal, and the mixed strategy Nash equilibrium.

In these examples, the signal that coordinates decisions for the two firms is an astronomical phenomenon or an instruction from an anonymous third party. As we noted in the discussion of stoplights, provision of a coordinating signal might be a function of the public authority, as the provision of a public good. To what extent, then, might the provision of coordinating signals for the macroeconomy be a public function? Indeed, arguably, it already is. The importance of the "announcement effects" of policy announcements by the Federal Reserve System in the United States has long been known (Waud, 1970). It has been suggested (Stein, 1989) that the Fed deliberately practices "cheap talk" as a means of influencing economic activity, though the generality of the reasoning has been challenged (Conlon, 1993). Federal Reserve announcements – and similar announcements from other public bodies – could play the role of public "sunspots" in determining a correlated equilibrium in the "game" of macroeconomics. In the example with private signals, the "third party" giving the instructions sounds a bit like an economic planning agency.[6] This is not to suggest that any economic planning agency has ever, in practice, had the knowledge necessary to construct a self-enforcing plan, using randomized strategies, as that example proposes. It is an open question whether this might be practicable in some future institutional context. The conclusion we may draw is that the public provision of coordinating signals is a public function about which very little is known, and that merits extensive future research.

## 5.5 PLURAL NASH EQUILIBRIA AND THE RATIONALITY POSTULATE

It has been argued (Hargreaves Heap and Varoufakis, 1995; Coleman, 2003) that the multiplicity of Nash equilibria impeaches the rationality postulate basic to game theory (as well as neoclassical economics). When there are plural equilibria, the agents cannot determine their decisions simply through rational procedures. At best, custom and convention come into their decisions, and at worst, the decisions are simply indeterminate. This is said to discredit the rationality postulate. This critique goes too far, however.[7]

In the case of coordination games, such as Game 4.5, custom can indeed be the decisive determinate of the decisions. But the best-response principle is an explanatory principle without which custom might not determine the decisions in these cases. In Game 4.5, for example, the key point is that following custom is a best response; if people do not predictably follow their best response in Game 4.5 then the custom itself

loses its predictive role. The nonrational alternative would be to suppose that people mechanically follow custom regardless of whether it is a best response or not. This in turn would mean that in Game 4.1, the pollution game, a custom of nonpollution would be sufficient to assure the efficient outcome. Game theory predicts the opposite, setting limits to the cases in which custom may be decisive. Indeed, the weakness of custom in limiting the deployment of polluting technologies does seem to contrast with the power of custom in determining that Britons drive on the left side of the road and Americans on the right. In any case, this is an elaboration, not a failure, of the rationality hypothesis.

The case is more difficult in an anticoordination game, such as Game 4.6. In such a game a uniform signal, such as a convention along the lines of "drive on the left," will not do, so that the decision is all the more likely to be indeterminate. Yet, in fact, a more complex custom may resolve the decision. If the participants in the game are ranked, so that the person of higher rank is given precedence, then the information as to which agent is of the higher rank makes the decision determinate. A difficulty is that the hierarchy of rank must be complete: it must be that every match is between two agents of different rank. Rank in military organizations illustrates this. Combat generally involves coordinated action but differentiated missions and objectives, with great and widely different risks. In battlefield conditions, coordination is crucially important and indeterminate decisions can result in disaster. Thus, it is crucial that some specific person is able to make authoritative decisions, and far less important to make fine calculations about who is best suited to make them. The system of military ranks, with persons of equal rank subordinated by seniority or even age, is well suited as a customary solution to this problem.

On its face, it may seem that the correlated equilibria make the case worse, because they are far more numerous than Nash equilibria in many games. When there are two or more undominated Nash equilibria, there is a continuum of correlated equilibria. However, this is misleading. When the two businessmen agreed to settle their difference by arm-wrestling, the probabilities (whatever they may have been) were probabilities they agreed on, and their agreement specified a single one of the infinitely many possible allocations of probabilities among the two different pure-strategy Nash equilibria. The same is true when the township supervisors of Eastonia and Westoria agree to base their decision on a cost–benefit study, although each is certain that his own town is the best choice, and when two drivers at an intersection happen unpredictably to arrive when the light is red one way and green the other, and of Celtic priests deciding who is to be sacrificed.

It is true that there may not be time to come to agreement on the probabilities or on a mechanism, such as arm-wrestling or drawing the short straw, so that a correlated equilibrium may simply not be available as a means of resolving an indeterminate Nash equilibrium. The meeting of two cars at an intersection has been given as an example – and probably most readers of this book have experienced impasses of this kind. It is also true that, in a game like Game 5.3, Aumann's game, there is potential of increasing efficiency beyond what an average of the Nash equilibria can support; but this potential can only be realized with private, imperfectly correlated signals, which may be difficult or costly to arrange. But it is not the rationality postulate that underlies these failures – it is the lack of sufficient time or communication.

The widespread observation of equiprobable mechanisms suggest that, in the absence of clear reasons against them, equiprobable correlated equilibria will usually occur when there are multiple undominated Nash equilibria. That equal probabilities are cognitively salient and determine a Schelling focal point, and require little knowledge but are consistent with the principle of insufficient reason, further points in favor of the prediction of an equiprobable solution. If there is a custom or convention that supplies a solution, then custom will take precedence over an equiprobable chance mechanism. In the rest of this book, cases of multiple Nash equilibria that cannot be resolved otherwise will be assumed to lead to equiprobable correlated equilibria.

## 5.6 CONCLUSION

Nash equilibria do not exhaust non-cooperative game theory. Correlated equilibria, with strategies that are randomized but not independent of one another, expand the set of possible non-cooperative solutions, but also resolve many of the puzzles about games with two or more undominated Nash equilibria. It should be added that the correlated equilibria can be thought of as Nash equilibria in enlarged games that include such things as arm-wrestling or drawing straws, with strategies contingent on events in those stages. It is the fact that they are Nash equilibria of enlarged games that makes them self-enforcing. Thus, the correlated equilibria are a structure built on the foundation of Nash equilibria. They cannot supplant Nash equilibria in non-cooperative game theory. Public policy applications of correlated equilibria are largely unexplored. Nevertheless, these studies enrich our understanding of the theory of non-cooperative games in strategic normal form in important ways and their role in the study of public policy deserves extensive research.

# NOTES

1. The "principle of insufficient reason" is at best controversial. If we reject it entirely, though, then multiple equilibria are not likely to be much of a problem – people will always have some grounds for thinking one equilibrium more likely than another, and the set of equilibria collapses to the one that is thought more likely. Schelling's (1960) discussion follows very much that line. On the other hand, if we take seriously the idea that the agents have no reason to expect one equilibrium rather than another, we might express that by saying that their information is minimal; and the minimum-information condition for an exhaustive set of exclusive alternatives is that the alternatives are equiprobable. (See McCain, 1972, for discussion and an application.)
2. The invention of the traffic signal has been widely attributed to Garret Morgan, an African-American inventor (Famous-inventors.com, 2006) although that has become a matter of controversy on the world-wide web (Crandall, 2007). It does seem clear that Morgan was, at least, *an* inventor of a device to control traffic at intersections, bringing about a correlated equilibrium solution to the anticoordination game of intersection traffic.
3. Keystone Automobile Club (1927).
4. It would make little difference if each wished the honor of being sacrificed, provided that the sacrifice must be unique.
5. Since the only Nash equilibrium of scissors-paper-stone is a mixed strategy equilibrium with equal probabilities, this method uses a game with an unique solution to generate the probabilities to choose among the equilibria of a game with plural Nash equilibria.
6. McCain (1991) proposed a model of economic planning for market economies based on a coordination game model which in turn was derived from Rosenstein-Rodan (1943). This model was extended to asymmetric information in McCain (1985). (The 1991 paper, though preliminary, was long delayed in publication.) McCain's model assumes only public signals, however.
7. I do not mean to suggest that there are no valid criticisms of the rationality postulate, as discussions at Chapters 3 and 8 may illustrate.

# 6. Non-cooperative sequential games and public policy

In Chapters 4 and 5, our focus was on non-cooperative games in strategic normal form. While (as von Neumann and Morgenstern showed) all games in extensive form can be represented in strategic normal form, to do so in general we may have to be careful to specify strategies as contingency plans. Thus, the strategic normal form will apply most naturally and with the best intuition to games in which simultaneous choices of behavior strategies must be made, such as the Prisoner's Dilemma. Conversely, when some decisions must in fact be made before other decisions are made, so that subsequent decisions are made with knowledge of the earlier decisions, the game represented in extensive form may be more natural and intuitive. Such games are said to be sequential. In this chapter we focus on sequential games and on the game represented in extensive form.

## 6.1 SUBGAME PERFECTION AND TREMBLING HANDS

Recall Game 2.2, Figure 2.1 in Chapter 2. We should notice that the decision by Firm A, to accommodate or retaliate, is a subgame in this game. Accordingly, we can define a *behavior strategy* locally at this decision point. The behavior strategy is just to accommodate or to retaliate, without specifying any conditions as to what previous decisions might be made. (Such conditions would be trivial in this case anyway.) In the spirit of Nash equilibrium theory, we might suppose that Firm A will choose the behavior strategy that leaves it with the larger payoff. This is "accommodate" for a payoff of 2 rather than 1. Moreover, the potential entrant, Firm B, can anticipate this. Therefore, Firm B expects that the payoff from the behavior strategy "enter" pays 1 while the behavior strategy "don't" pays 0, and accordingly Firm B chooses "enter." Thus the non-cooperative solution to this game would seem to be "enter, accommodate."

Four comments should be made on this reasoning.

First, it is an example of *subgame perfect Nash equilibrium*, a concept that is now central to the analysis of sequential games. A subgame perfect

Nash equilibrium is a sequence of behavior strategies that (1) is a Nash equilibrium in behavior strategies in the game as a whole, and (2) is also a Nash equilibrium in every subgame. In this case, we have just one proper subgame, and that is Firm A's decision whether to retaliate or accommodate. The fact that Firm A chooses the behavior strategy that maximizes its payoffs at that point means that we do indeed have a Nash equilibrium in this subgame. That Firm B maximizes its own payoff based on anticipation of that decision means that each firm is choosing its best response to the other's strategy (sequence of behavior strategies); we have a Nash equilibrium in the game as a whole.

Second, the example illustrates an algorithm for finding subgame perfect Nash equilibria. The algorithm is called "backward induction." In this case, notice, the first step is the last decision to be made, resolved by determining a Nash equilibrial behavior strategy as if the subgame stood alone. We then treated the first decision as a "reduced game" in which the payoffs were 0 for "don't" and 2 – the equilibrial payoff in the first step – for "enter." The Nash equilibrial decision for the "reduced game," "enter," then completes the subgame perfect Nash equilibrium. For more complex games and in general, the algorithm would be as follows: (1) Among all subgames, determine those that are *basic.* A basic subgame is one that has no other subgames within it; in Game 2.2, Firm A's decision is the only basic subgame. (2) Determine the behavior strategies that constitute a Nash equilibrium for the basic subgames. (3) If the Nash equilibrium is unique, form the *reduced game* by eliminating the basic proper subgames, replacing the basic proper subgames by their equilibrial payoffs. If the Nash equilibrium for a particular subgame is not unique, replace the subgame by one or another set of equilibrial payoffs. (4) Repeat until the reduced game is the first decision to be made, and determine the Nash equilibrial behavior strategies and payoffs for that decision. (5) The sequence of behavior strategies are then the subgame perfect Nash equilibrial behavior strategies, and the payoffs yielded by this sequence are the subgame perfect equilibrium payoffs.[1] If one or more of the equilibria determined at stage 3 are non-unique, then the subgame perfect Nash equilibrium is non-unique.

Third, another way to express the result in this analysis is to say that the threat of a price war in this case is *incredible.* This recalls, yet again, Nash's comment that a threat is often something that a person would not want to do for themself, and that is the case with respect to the price war in Game 2.2. In general, in non-cooperative game theory, a threat is *credible* only if it is subgame perfect. If the threat is part of a subgame perfect equilibrium sequence of behavior strategies, then it is Nash equilibrial in the subgames of which it is a part, and if so then it is an exception to Nash's comment – it is something the person would want to do for themself.

Fourth, the application in this case is itself very important for public policy and economics. It supports the argument that market entry is irrepressible in a market economy without government restrictions. Since entry tends to increase price competition and price competition in turn tends to induce efficient pricing and resource allocation, this would be an element of an argument for free market policies. On the other hand, if in a special case market entry were to have negative consequences, it could be an element of an argument for public policies that would restrict entry. Patent rights would be an instance of the special case.

Although we are concerned here with the representation of the game in extensive form, it will be helpful here to digress on the strategic normal form. Recall the normal form of this game, Table 2.2 in Chapter 2. Notice that this game has four Nash equilibria: "don't" with strategies 1 and 4, and "enter" with strategies 2 and 3. The latter two correspond to the subgame perfect Nash equilibrium, since both require Firm A to choose the behavior strategy "accommodate." But the other two formally are Nash equilibria as well.

Is there any basis to exclude these equilibria? We do notice that for Firm A strategies 1 and 4 are weakly dominated. A strategy is *weakly dominated* if there is another strategy the payoff to which is never less, and is greater for at least one strategy that the other player might choose. Since behavior strategy "accommodate" pays 2 if Firm B enters, and 5 if Firm B does not, the contingent strategies leading to "accommodate" weakly dominate those leading to "retaliate." Indeed, we see that this game has only three distinct outcomes: price war, accommodated entry, and continued monopoly. Behavior strategy "don't enter" always leads to the same outcome, therefore to the same payoffs. In general, when we translate a game in extensive form into a game in strategic normal form, we will find that there are many fewer outcomes than strategy combinations, since many different combinations of contingent strategies will lead to the same basic subgames, and therefore to the same outcomes. Thus, weakly dominated strategies are likely to be quite common in games in extensive form. The question thus becomes: is there any basis to exclude equilibria that are based on weakly dominated contingent strategies?

In 1975, Selten (CGT, pp. 317–54) introduced a refinement of Nash equilibrium called the *trembling hand* equilibrium: suppose there is some small positive probability that a player will fail to choose his best-response strategy, so that the player will choose any other specific behavior strategy instead. This is a *perturbed* game. We can define equilibrium for the perturbed game in the usual way, with the expected values of payoffs determining the best responses. The equilibria of a perturbed game may differ from those of the original game. Now define a sequence of perturbed games in

which the probability of errors approaches zero in the limit. Selten shows that the limit of the equilibria of such an (appropriately constructed) sequence of perturbed games is an equilibrium of the original game, but not all equilibria are the limits of such sequences. Those equilibria that are the limits of such sequences are *perfect equilibria*. This means that equilibria are excluded if they depend on behavior that is rational only on the assumption that other players are themselves perfectly rational.

In a perfect equilibrium, a Nash equilibrium is realized in every subgame of the original game, including of course the game itself. Thus a perfect equilibrium is, in particular, subgame perfect. In Game 2.2 revised, for example, suppose that the probability that Firm B chooses the "wrong" strategy is $p$. In such a case the payoff of contingent strategies 1 and 4 is $2p + 1(1 - p)$. The payoff to strategies 2 and 3 is $5p + 2(1 - p)$. Clearly the second is larger for any positive $p$, so strategies 1 and 4 are not best responses in any perturbed game. Consequently, the subgame perfect equilibrium of Game 2.2 is the only perfect equilibrium of Game 2.2 revised.

But the perfect equilibrium can also be applied to games that do not have subgames. As an example, Selten discusses the Horse game (Figure 2.3, Game 2.3, Chapter 2). As in Chapter 2 we will encode the behavior strategies as follows: for Firm A, "License" is R1, "Don't" is L1, for Firm B, "Don't" is R2, "Enter" is L2, and for Firm C, "License" is R3 and "Don't" is L3. Notice that for this game, L1R2R3 is an equilibrium: but it is so only because Firm B does not get an opportunity to play at all. If Firm B were to get an opportunity to play, he would know that player 1 had not played L1 but R1 and, on an expectation that Firm C would play R3, player 2's best response is not R2 but L2. This is unreasonable, Selten argues, writing (p. 328) "Player 2's choices should not be guided by his payoff expectations in the whole game but by his conditional payoff expectations" at decision node B. In fact, L1R2R3 is not a perfect equilibrium.

In Game 2.2 revised, the Nash equilibria with contingent strategies 1 and 4 can be excluded because they are not perfect equilibria. As we noted, they are formally Nash equilibria, but their exclusion is very much in the spirit of Nash's non-cooperative game theory. Retaliation is a threat strategy, and would not be "something A would want to do, just of itself." The term "perfect Nash equilibrium" is quite apposite: rather than restricting, Selten has perfected Nash's reasoning.

## 6.2 PRAGMATICS: PROBLEM SPECIFICATION

As before, one of our concerns is with problem identification. Accordingly, we consider some cases in which the game in extensive form helps us to

specify a problem of interactive decision-making that may be relevant for public policy.

### 6.2.1 Ulysses and the Sirens

We recall that the entrenched monopolist in Game 2.2 was unable to prevent the entry of new competition, because the threat of a price war was not subgame perfect. Yet the entrenched monopolist might not be entirely helpless. The monopolist might consider a large-scale investment that would increase both its production capacity and its costs, creating a situation in which the restricted output corresponding to accommodation of the new entering firm would be less profitable, and the increased production incident on a price war more profitable. If it anticipates competitive entry (for example, following deregulation) the firm might undertake such an investment, in the hope that the entry would be prevented. This is *strategic investment to deter entry,* and Game 6.1 is an example. As before, the first payoff is to Firm B. (See Figure 6.1; the numbers in parentheses will be explained below.)

In Game 6.1, if Firm A decides not to invest, then we have Game 2.2, but if Firm A decides to invest, we have a different subgame. To solve this more complex game, we again identify the basic subgames, and they are A's second round of decisions. Both are basic. For the lower one, we already know that the perfect behavior strategy is "accommodate." For the upper subgame, however, "retaliate" is the perfect response. Thus, Firm B can anticipate that the behavior strategy "enter" will pay 1 in the lower subgame but −1 in the upper. "Enter" is a perfect behavior strategy in the lower subgame but not the upper. Anticipating all this, Firm A expects that investing will lead to profits of 4 while not investing will lead to profits of 1. Thus, the subgame perfect sequence of behavior strategies is "invest," "don't enter."

The investment may or may not be efficient. To know the answer to that, we need to know more about the benefits to customers. In economics, the *consumers*' *surplus* measures the net benefit to the buyer from buying at a particular price. The consumers' surplus plus the total profits of the two firms measures the net social benefit for this industry, in the absence of externalities. For this example, the numbers in parentheses represent consumers' surpluses corresponding to the different degrees of price competition and output capacity in the various possible outcomes of the game. The largest total, 8 on our arbitrary scale of measurement, occurs if Firm A does not make the investment, Firm B does enter, and entry is accommodated. In that case, the consumers have the benefits of both expanded production capacity and increased price competition. By contrast, in the

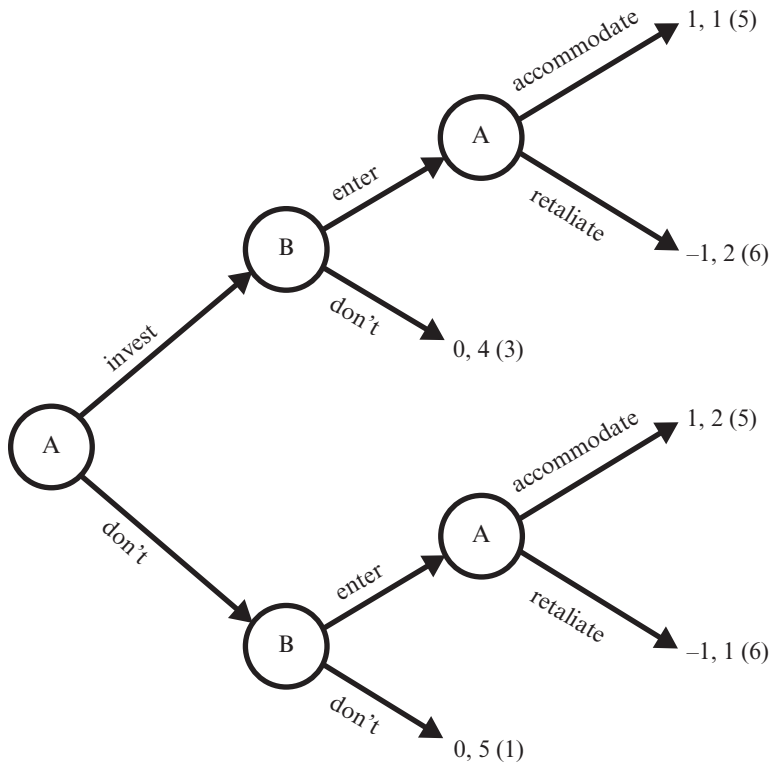


*Figure 6.1 Game 6.1: strategic investment to deter entry*

subgame perfect equilibrium, consumers benefit from increased capacity but not from increased price competition, so the total, 7, is less. Efficiency is improved – without either entry or new investment, at the bottom of the diagram, total benefit is 6 – but the resulting outcome is not fully efficient because of the entry-limiting investment.

The key point is that the investment has been successful in deterring entry. In Game 2.2, Firm A was unable to deter entry because the threat of a price war was incredible: Firm B could anticipate that Firm A would not undertake such an unprofitable step. If Firm A were able to do so, it would choose the best response behavior strategy of accommodation. However, by making the strategic investment, Firm A has deprived itself of the opportunity to accommodate profitably.

This illustrates a more general point that emerges from the study of the game in extensive form. In some cases, an agent may be better off with fewer opportunities, fewer options. Schelling (1960) stressed that in order

to bind others, we might have to find a way to bind ourselves. This is what Firm A has done in Game 6.1. Elster (1977) drew the analogy to Ulysses having himself bound to the mast so that he could hear the song of the sirens – and not succumb to it.

By considering a strategic investment to deter entry, Firm A has *nested* Game 2.2 within a larger game. Since Game 2.2 is a subgame of Game 6.1, it is an *imbedded* game. That is important, since it means that the subgame perfect solution of Game 2.2 is perfect also in the context of the nesting Game 6.1. That is to say, (for purposes of non-cooperative game theory) we can analyze the imbedded game as a game in its own right, expecting that its equilibrium will be equilibrial also within the larger game – while that might not be true for a game that is nested but not imbedded in the larger game.

### 6.2.2 Agency

A large body of literature in economics and game theory has grown up with respect to relations between a principal[2] and an agent. These models are sequential in that the principal sets the conditions for the decisions taken by the agent. There are some aspects of the agent's activity, such as effort, that the principal is unable to observe. It is this that makes the relationship (at least partly) non-cooperative. Agency models are also more general than the phrase may suggest. The agent may sell a home on behalf of the owner, but may be a corporate executive when the principals are shareholders, or may be a professional and the principal a client, and so on. As usual we will illustrate the idea with a simplified (and nonmathematical) example; the agent will be a lawyer and the principal a client.

For the example, there is a third player: chance. The lawyer can pursue the lawsuit with great or slight effort, and these are the lawyer's strategies. Chance can provide good or bad conditions for the lawsuit, at random with equal probabilities ½. If the lawyer makes a great effort and chance is favorable, the lawsuit yields 14 to the plaintiff. (As usual the reader may add as many zeros as seem appropriate.) For the lawyer, great effort is equivalent to a deduction of 2 from her fee, while slight effort is equivalent to a deduction of 1. If the lawyer makes a slight effort and chance is unfavorable, the lawsuit yields only 2. If the lawyer makes great effort and chance is unfavorable, *or* if the lawyer makes slight effort and chance is favorable, the lawsuit yields 6. Since the client cannot observe either effort or the random variate, he will not know whether the intermediate outcome is a consequence of slight effort or of adverse chance. As a result, the payment for legal services cannot be conditioned on effort. The client is considering whether to pay the lawyer a flat fee of 3 or a contingent fee of 1/3 of the award.

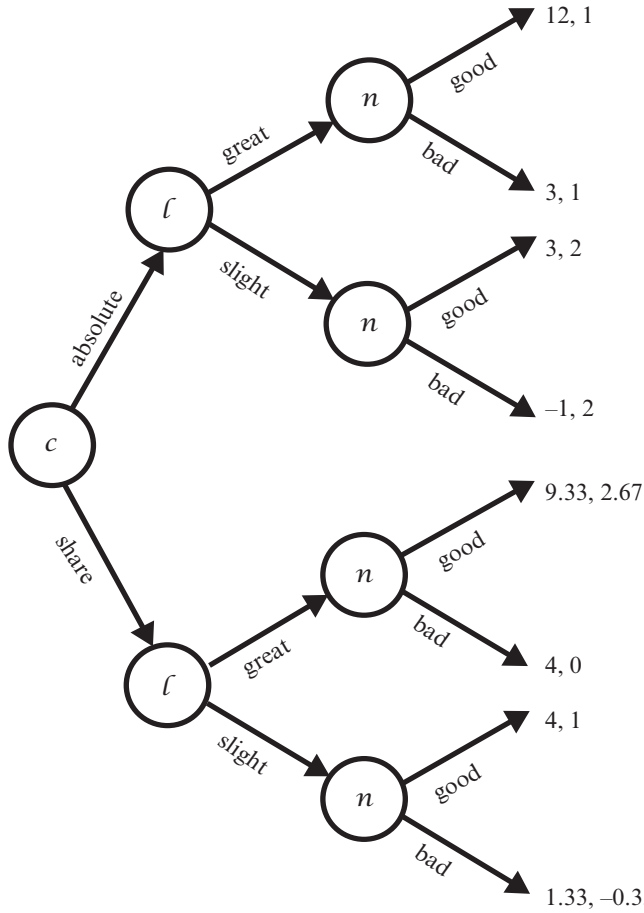


*Figure 6.2 Game 6.2: agency*

The resulting game in extensive form is shown as Game 6.2, with $c$ denoting the client, $\ell$ the lawyer and $n$ (for nature) chance (Figure 6.2). The first payment is the net benefit to the client, the second to the lawyer. Chance "decisions" are not subgames (since chance has no intentions) and the basic subgames are the lawyer's decisions. With a fixed fee of 3 she is always better off making slight effort, but with the contingency fee can anticipate an expected value of 1 1/3 with great effort, but 1/3 with slight effort, so in the lower part of the diagram, the lawyer will choose to make a great effort. Accordingly, the client will face an expected value of 1 (net of the fee of 3) in the case of a flat fee but 6 2/3 in the case of a contingency

fee. The subgame perfect equilibrium is for the client to offer a contingency fee and the lawyer to make great effort.

In Williamson's informal work on the economics of the firm and similar cooperative arrangements (for example, 1964), he put some stress on "opportunism," arguing that many seemingly inefficient contractual arrangements could be explained as means of avoiding or limiting opportunism. He was sometimes criticized by economists who suggested that, after all, opportunism is nothing more than self-interest. In the agency example, suppose that the lawyer promises to make a great effort in exchange for a fixed fee. This would generate an expected value of 7.5 for the client, so that, if she trusts the lawyer, the client would accept that offer. If, however, the lawyer nevertheless makes slight effort, we would say, in a sense of everyday usage, that the lawyer had acted opportunistically, and that the choice of a contingency fee is made to avoid the consequences of opportunism, very much along the lines of Williamson's thinking. It is true enough that opportunism is *non-cooperative* self-interest, but also that opportunism is not identified by the character of the behavior alone but also by the sequential structure of the game. Opportunism in this usage will be important in some later chapters.

## 6.3 IMBEDDED GAMES

In Chapter 2, a game was said to be imbedded in another if the first game is a proper subgame of the second. The imbedded game can be studied as a stand-alone non-cooperative game. Since we cannot represent the universe as a single game, this is essential to any applied game theory. In particular, it is important for public policy applications. If we think of the public sector as setting the "rules of the game," then private sector (interdependent) decisions are imbedded within the game of public policy determination. We will illustrate the point, as usual, by example.

Let us return to an example from Chapter 2: two agents own property on a river. If Joe Upstream diverts the stream for some project of development of his own land, then Irving Downstream's water supply from the river will be reduced. How may we represent this as a game in extensive form? It depends on the regime of property rights. Under *riparian rights,* Irving will be able to sue Joe and recover any damages that result from the diversion. (This will imply some legal costs that will have to be borne by one landowner or the other.) On the other hand, if the regime is non-riparian, a landowner is allowed to develop his own land as he chooses, regardless of the results for other landowners up or downstream. Thus, in a non-riparian regime, Irving would not be able to sue to recover damages. If he did file a lawsuit he would lose.

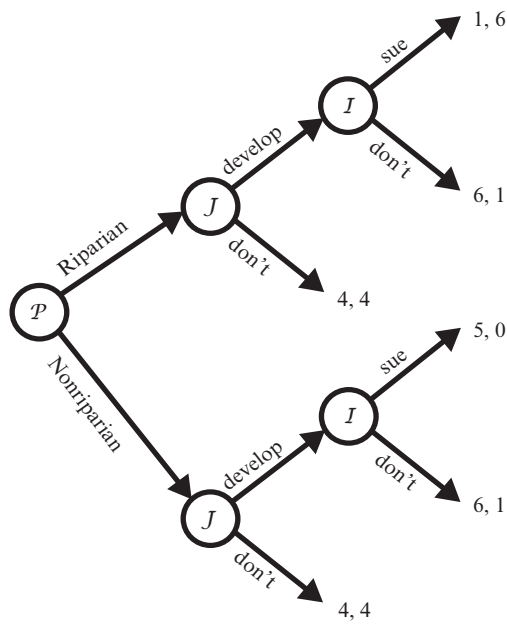


*Figure 6.3 Games 6.3, 6.4, 6.5: a public decision process*

Let us consider first a case in which the project is inefficient, that is, it imposes costs (damages) on Irving that are greater than the benefits that Joe obtains. Then we have Game 6.3, shown as the upper branch of Figure 6.3, beginning with node *j*. As usual, the payoffs are arbitrary but chosen to be consistent with the assumptions of the problem; in this case total benefits are evaluated on a scale from 0 to 10. The first payoff is to Joe, the second to Irving. The subgame perfect equilibrium of this game is that Joe does not develop his property, and in this instance it is an efficient result; though we should stress that this depends on the particular numbers in this case and if (for example) the benefits of the development were greater, this result could be inefficient.

Now suppose instead that the regime were one of non-riparian rights. In that case Joe has the right to develop his property as he may please, and Irving has no grounds for any lawsuit. If he does sue he will lose. We then have Game 6.4, the subgame shown as the lower branch of Figure 6.3 beginning, again, with node *j*. In this case the subgame perfect equilibrium is that Joe goes forward with the inefficient development, and Irving does not sue.

Riparian rights provide an example of alternative systems of property

rights, with no clear rationale other than efficiency for choosing one system over the other. From a normative or ethical point of view, we have two arguments that seem to offset one another: (1) one ought to be able to enjoy an unaltered river frontage; and (2) one ought to be able to develop one's own property as one may choose. In any case the decision between two property rights regimes can only be made by the public authority.

Now let us assume that these two examples are typical: that is, that on the average, riparian rights will deter inefficient projects and admit efficient ones (where the damages are less than the benefits to the developer). This is an independent judgment of fact and would have to be verified by empirical research. We suppose that this has been done. We treat the public authority as a player in the game, and assume that the payoff to the public authority is the sum of the payoffs to Irving and Joe. Then the decision of the public authority is represented by Game 6.5, Figure 6.3 as a whole.

The subgame perfect equilibrium of Game 6.5 is that the public authority chooses riparian rights and Joe elects not to proceed with the inefficient project. Note that for this example, both the two-person games with riparian and with nonriparian rights are imbedded games in the public decision. Conversely, in order for non-cooperative game theory to be applied validly in the formation of public policy, it is essential that the private sector decisions be imbedded as subgames in the larger game comprising the public policy decision – that is, that the interdependent decisions in the private sector be analyzed as complete games, and not as fragmentary nested (but not imbedded) games. If we think of public policy as setting "the rules of the game" for the private sector, then clearly private sector decisions must be imbedded in the game of determining public policy.

In this example, the strategies available to Joe and Irving do not have any influence on the decision of the public authority. That is, strategies involving lobbying and bribery have not been taken into account. If these instances of "rent-seeking behavior" are available to the private agents, then their game (that is, the private sector) is no longer imbedded in the public policy decision, although it is nested in that larger game. In that sense, the imbedding of private sector decisions in the larger game of determination of public policy is an ideal case.

## 6.4 REPEATED PLAY

In a very early experimental study in economics, two experimental subjects played a non-cooperative game 100 times in succession. The results did not agree with the predictions of Nash equilibrium for the individual plays. John Nash's response to this was that it was not a valid test of the theory,

but that the entire series of 100 repetitions would have to be solved as a single game. (This account follows Poundstone, 1992, Ch. 6.) It was widely suspected that repeated play would make a difference (von Neumann and Morgenstern conjecture along these lines in passing) and could result in cooperative play in an otherwise non-cooperative game. This suspicion was so strong that it is sometimes called a "folk theorem," although it has not been proved, nor even formally stated, and indeed is not true in general. However, the tools to deal with the question did not exist until the 1970s. A game consisting of repeated plays of a simpler game is best dealt with as a game in extensive form, and subgame perfect equilibrium is a key tool.

### 6.4.1 The "Folk Theorem"

Nutter's (1964) theory of oligopoly envisions price competition among a small number of firms as a Nash equilibrium in which the competitive price is the unique equilibrium. This equilibrium can be arrived at by iterated elimination of dominated strategies,[3] in which a lower price always dominates a higher price (at or above the competitive price). However, typically, oligopolists face one another in price competition again and again, period after period. Thus, perhaps they will be able to attain a cooperative solution (among themselves, not considering the customers as players) and share monopoly profits based on higher prices. If so, there would be important implications for antitrust policy. This is one of the most important (and common) applications of the theory of repeated play.

As usual we begin with a simple example, designed to minimize mathematical difficulty. Firms A and B will choose between just two prices, high and low, and the prices are the behavior strategies for the duopolists. The high price corresponds to a shared monopoly and so higher profits for both firms, and the lower price corresponds roughly to a competitive price. We will not explore the details of cost and revenue that generate these profits and, for mathematical simplicity, will ignore intermediate, lower and higher prices that might be charged. The example is Game 6.6, shown as Table 6.1. However, the two firms will play the game again and

*Table 6.1 Game 6.6: duopoly*

| Payoffs: A, B | | Firm B | |
|---|---|---|---|
| | | High | Low |
| Firm A | High | 7,7 | 3,10 |
| | Low | 10,3 | 4,4 |

again, and this repeated play, taken together, is a sequential game. The strategies for this game are behavior strategies, while a (von Neumann and Morgenstern) contingent strategy for this game would be a sequence of conditional decisions for each play of the game.

**Case 1 M repetitions**

We first consider a case in which the play is repeated for a definite number of iterations, indicating the number of repetitions as M. The repetitions of Game 6.6 are indicated as 1, 2, . . ., M. In the larger game of repeated play, subgames[4] are sequences j, . . ., M, where $1 \leq j \leq M$. In particular, play M, the last play, is the only basic subgame. We know the equilibrium of that subgame: it is "low, low" for payoffs 4, 4. We then consider the reduced game consisting of repetitions 1, . . ., M − 1, with payoffs augmented by 4 each.[5] The only basic subgame of this game is repetition M − 1. Once again, we know the equilibrium and it is the non-cooperative solution in behavior strategies to Game 6.6. We continue in this vein until repetition 1 is the reduced game; for it, too, the non-cooperative solution is the equilibrium. The conclusion is this: for a game with a definite number of repetitions, the folk theorem is false, and the only subgame perfect equilibrium is a sequence of plays of the Nash equilibrium behavior strategies in the original game.

**Case 2 Indefinite repetitions**

Aumann (CTG4, 1959, pp. 287–324) defined a *supergame* for a game $\Gamma$ as an infinite sequence of repetitions of $\Gamma$. Clearly, the reasoning in the previous case will not apply to a supergame, since the supergame has no basic subgames. Every subgame is a sequence of repetitions of $\Gamma$ indexed as j, j+1, . . ., without limit, so every subgame contains other proper subgames. We will have to use different methods to deal with supergames.

But is this realistic? After all, nothing lasts forever! However, Case 1 seems a little artificial in assuming a *definite* number of repetitions. How likely is it that oligopolists, or others engaged in a repeated game, would anticipate the exact number of repetitions that will occur? Suppose instead that at each repetition of $\Gamma$, the players can expect that there will be yet another repetition with probability $\delta$, but the probability that there will be no further repetitions whatever is $1 - \delta$. Let $y_j$ be the payoff to a player in the $j$th repetition of the game. Then at repetition $t$, the player wants to maximize the mathematical expectation $\sum_{j=t}^{\infty} \delta y_j$. This formula is the same as a formula for the discounted present value of the series of payments at a discount factor $\delta$, and accordingly $\delta$ is referred to as the discount factor.[6] Although the probability of more than M rounds of play approaches zero as M increases without bound, a game such as this has to be analyzed as an infinite game and has no basic subgames.

We may suppose that the players in the game choose behavior strategies for each repetition according to some rule. The "tit-for-tat" rule is an important possibility: begin by playing the cooperative behavior strategy, "high" price in this case, and continue playing it unless the other player plays non-cooperatively ("low" price). If the other player plays non-cooperatively, then retaliate by playing once non-cooperatively on the following round. Notice that the threat of retaliating by playing non-cooperatively is credible, since non-cooperative play is an equilibrium behavior strategy on any particular round.

Tit-for-tat is called a "trigger strategy," since non-cooperation triggers a retaliatory act of non-cooperation. However, properly speaking, the tit-for-tat rule is not itself a strategy.[7] It is neither a behavior strategy nor a contingent strategy as understood by von Neumann and Morgenstern. Rather, it characterizes an infinite family of contingent strategies or of sequences of behavior strategies for this game. However, because the retaliation is itself Nash equilibrial, each of the contingent strategies in the family is subgame-perfect, provided that the threat is sufficient to deter the other player from choosing the non-cooperative strategy. That is the question to which we now turn.

The question is this: supposing Firm A plays according to the tit-for-tat rule, will firm B will be deterred from a *single* opportunistic non-cooperative play, that is, from playing "low" at round $t$, taking advantage of A's cooperative play, and then returning to playing "high, high, high" so long as the game continues. This implies a sequence of payoffs $y_t = 10$, $y_{t+1} = 3$, $y_{t+2} = y_{t+3} = \ldots = 7$. The alternative is to play cooperatively on every round, which implies payoffs $y_t = y_{t+1} = y_{t+2} = y_{t+3} = \ldots = 7$. The expected value of the first sequence is $10 + \delta 3 + \delta^2 7 + \delta^3 7 + \delta^4 7 + \ldots$. The expected value of the second sequence is $7 + \delta 7 + \delta^2 7 + \delta^3 7 + \delta^4 7 + \ldots$. Since only the first two terms differ, the second sequence of payoffs is greater if $7(1 + \delta) > 10 + 3\delta$. A little algebra tells us that this will be true whenever $\delta > 0.75$.

What if Firm B plays non-cooperatively again and again? If so, then A will respond by also playing non-cooperatively on every turn, in accordance with the tit-for-tat rule. Thus, Firm B's sequence of payoffs is $y_t = 10$, $y_{t+1} = y_{t+2} = y_{t+3} = \ldots = 4$, which can be written as $10 + (\delta/(1 - \delta))4$. For any $\delta > 0.25$, the expected value of this sequence will be less than the expected value for the sequence of payoffs for a single non-cooperative play. Conversely, if the threat implicit in tit-for-tat play is sufficient to deter a single round of non-cooperative play, it is undoubtedly sufficient to deter systematically non-cooperative play. With $\delta > 0.75$, playing against tit-for-tat, Firm B will simply find that more non-cooperation means lower expected value payoffs.

What we have found is that, if the probability of another round of play is great enough,[8] in this example, a tit-for-tat rule by one player will make it unprofitable for the other player to deviate from cooperative behavior. If each player plays tit-for-tat, then the play is always cooperative, and neither player can gain anything by deviating from the tit-for-tat rule.

Unfortunately, that is not the whole story. Mutual play of a tit-for-tat rule is only one of many equilibria of an indefinitely repeated social dilemma. In particular, pure non-cooperation by both players is always also an equilibrium. There are many others at intermediate levels of efficiency. Nor is the tit-for-tat rule dominant over all other rules by which the game might be played. Suppose, for example, that Firm A plays tit-for-tat while Firm B plays a more "forgiving" trigger strategy rule, tit-for-two-tats. That is, Firm B plays cooperatively unless Firm A plays non-cooperatively for two rounds in succession, and then responds with one round of retaliatory non-cooperative play. These two rules would lead to cooperation, and Firm B can do no better so long as Firm A sticks to tit-for-tat. But Firm A can do better by deviating from tit-for-tat. In particular, suppose Firm A adopts the rule of alternating cooperative and non-cooperative play. Then Firm B never retaliates and Firm A alternates payoffs of 10 and 7, a sequence that dominates the sequence from steady cooperative play. The point is that there are some strategy rules (for example, tit-for-two-tats) against which the tit-for-tat rule does not produce best responses.

The tit-for-tat strategy rule and variants of it, such as a tit-for-two-tats and two-tits for-a-tat (retaliate with two rounds of non-cooperative play for one round by the other player) are all *forgiving trigger strategy rules*, which means that the retaliating player will eventually return to cooperative play if the other player does so. A rule that plays cooperatively until the other player initiates non-cooperative play and then retaliates by playing non-cooperatively on all successive plays is called the *grim trigger*. The grim trigger may deter non-cooperative play where tit-for-tat would not. The grim trigger played a key role in warfare in the twentieth century. Poison gas was used as a weapon of war in World War I, and in the Iran-Iraq war of the 1980s, but not in World War II. The use of a weapon such as poison gas may be a social dilemma for the belligerents (McCain, 2004, pp. 43, 277–9). In a long war, with repeated battles, perhaps restraint might be based on fear of retaliation from an opponent playing according to a grim trigger rule. In fact, historical evidence makes it clear that Germany, the United States, and Britain (with pressure from the United States) were following a grim trigger rule with respect to gas (Harris and Paxman, 2002). This example may illustrate the real possibility of cooperation in games of completely opposed interest, but also underscores that

there is nothing inevitable about this, and that non-cooperation is always among the equilibria of repeated games.

This discussion assumes a two-person game. The extent to which the results may be extended to games of more than two persons remains a somewhat open question. What is clear is that the relatively simple argument along the lines of the previous example is not applicable to more than two players. Difficulties arise with as few as three players (Fudenberg and Maskin, 1986, p. 543). Allowing for correlated strategies (with public signals) and assuming sufficient diversity in the payoffs to the different players, Fudenberg and Maskin do extend the model to $n$ players. Abreu et al. (1994) follow Fudenberg and Maskin with a more precise characterization of the conditions for cooperation in $n$-person games. In a working paper, Haag and Lagunoff (2005) find that diversity in subjective rates of time discounting makes cooperation less likely, though it grows more likely in larger groups. Nevertheless, it seems widely felt that larger groups are less likely to cooperate, on the basis of experience in the applications to price competition.

### 6.4.2 An Extension

In some ways the probabilistic repeated play model seems very plausible. After all, retaliation is a matter of common experience. However, it allows very little for changing circumstances outside the control of the agents in the game. Suppose, for example, that two firms play according to the tit-for-tat rule for a number of years, and then it becomes known that one of them is financially impaired and may go bankrupt. As a result, it seems far less probable that there will be further rounds of "play," and the cooperative agreement breaks down. The repeated play model, with its constant discount factor, does not seem to allow for this sort of possibility. This section will sketch a modest extension of the model that will "realistically" allow for such changes of circumstances to affect the continuation of cooperation.

A key tool for this purpose is the state transition matrix. We suppose, for example, that there are just three possible states of the world: state 1, in which both firms are financially sound and the "game" of price competition takes place; state 2, in which the "game" takes place but one firm is financially impaired; and state 3, in which there is no play, perhaps because one firm has gone bankrupt. There are two players. In states 1 and 2 they play Game 6.6. In state 3 they do not interact at all, and payoffs for both players are zero.

Given that the world is in state $i$ in period $t$, the probabilities[9] that the world will be in state $j$ in period $t + 1$ are known constants summarized in the state transition matrix. Suppose the probabilities are as shown in Table 6.2.

*Table 6.2 Transition matrix 1*

| | | Transition to | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| Transition | 1 | 0.8 | 0.2 | 0 |
| from | 2 | 0.6 | 0.2 | 0.2 |
| | 3 | 0.1 | 0.1 | 0.8 |

The number in a given cell tells us the probability that the state represented by the row will be succeeded by the state represented by the column. Thus, for example, this transition matrix tells us that state 1 will be followed by state 1, 80 percent of the time, by state 2, 20 percent of the time, but never directly followed by state 3. Nevertheless, we might see the system in state 1 in the first period, in state 2 in the second period (with 20 percent probability) and in state 3 in the third period. The probability that the system would transit from 1 to state 3 so quickly is the compound probability, $0.2 * 0.2 = 0.04$ – a small probability. But, given more time, the probability could be greater since there are very many more ways that the transition could occur. Using compound probabilities, we can compute the probability that any one of the states will occur in any future period, starting out from state 1 (or indeed any other state). For example, the probability that we will observe state 1 steadily approaches a stable value of 0.64; and similarly for the other states approaches the constants 0.18, 0.18. In fact, many such models have equilibria of this kind, and the equilibria can be found by a fairly simple exercise in linear algebra, solving a system of three equations with the three constant probabilities as the three unknowns. We shall skip the details. We can also compute the probability of yet another round of play, that is, the probability that either state 1 or state 2 will occur in the period $n$, if play took place in period $n - 1$. (This reflects the probability both that state 1 or 2 will occur in period $n - 1$ and the probabilities in the state transition matrix.) This approaches a constant value of 0.78.

For a case like this, we might just take the equilibrium probabilities and treat the model as if it had constant probabilities, at least as a first approximation. Let us do that, asking whether tit-for-tat play will deter defection in Game 6.6. We find that if the probability of yet another round of play is greater than 0.675, indeed it will. If we begin from state 1, the probability of another round of play is greater than 0.675 in *every single period*, so we can be confident that cooperation is feasible based on the tit-for-tat strategy rule.

*Table 6.3 Transition matrix 2*

| | | Transition to | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| Transition | 1 | 0.8 | 0.2 | 0 |
| from | 2 | 0.4 | 0.2 | 0.4 |
| | 3 | 0 | 0 | 1 |

For our example an advantage of this approach is that state transition models can represent irreversible events, such as bankruptcy and death. Consider the transition matrix in Table 6.3. Row 2 tells us that when a firm is financially impaired, it will return to financial health with a probability of 0.4, go bankrupt with probability 0.4, or continue impaired in the next period with probability 0.2. As for the third row, it reminds us that liquidation is irreversible: once you are dead you stay dead, and the probability of coming back from the dead is zero.

If we repeat the enumeration of the probabilities of the three states for future periods, beginning in state 1, we see that the probability that state 3 will be observed approaches 1, and the probabilities of the other states, and of another round of play, approach zero. That is, state 3 is what is called an "absorbing state:" sooner or later we are all dead. As a result, the probability of another round of play keeps dropping and approaches 0 in the limit. It may seem that we can apply backward induction so that there will be no cooperation.

However, this is a mistake, or at least hasty. Assume that the agents can observe the state of the world. At the very least, agents will be able to tell whether anyone is bankrupt or not. We will assume that they can also observe whether they are in state 1 or state 2. Therefore, they can make their strategies contingent on the state. Thus, in place of tit-for-tat, suppose both parties play according to Rule 1:

*Rule 1. Cooperate IF the state is 1 AND (it is the first round of play OR the state in the previous period was other than state 1 OR the other agent played cooperate on previous round) ELSE defect.*

Now suppose we are at state 1 and one player defects on the current round, planning on returning to "cooperate" thereafter. His expected payoff is 10 + 3*0.8 + 4*0.2 = 13.2. On the other hand if he cooperates the expected payoff is 7 + 7*0.8 + 4*0.2 = 13.4. Cooperation pays better and defection is deterred. The term 4*0.2 is the payoff of the mutual defection

that is sure to occur in case state 2 is realized in the next period times the probability that this will happen. (The example does not allow for time discounting and with time discounting the result might be different.) In this case the probabilities 0.8 and 0.2 are always applicable because they are conditional probabilities, conditional on the observation that state 1 has occurred.

Suppose instead that state 2 has been realized. Then the conditional probabilities of state 1 and state 2 in the following period are 0.4 and 0.2. Suppose the player defects once while the other player plays Rule 1. Then the defector's expectation is 10 + 7*0.4 + 4*0.2 + 0*0.4 = 13.6. If he plays cooperate it will be 7 + 7*0.4 + 4*0.2 + 0*0.4 = 10.6. (Since there is no play in state 3 we assign payoffs of zero.) Cooperation does not pay.

Thus, whatever state occurs, there is no incentive to deviate from Rule 1 – Rule 1 is subgame perfect. (Here, again, we are assuming the rate of time discount is sufficiently small.) But notice what it means. We start from state 1, with cooperation. Over the next few rounds, the probability (as seen from period 1) that we will remain in state 1 declines. We can foresee that within several rounds, with high probability, the system will transit to state 2, and at that point cooperation will break down. If the firm in trouble manages to return to financial health (the system transits back to state 1), cooperation will be resumed. On the other hand, if one firm is liquidated, there will be no more opportunities for cooperation; and since this will occur sooner or later, cooperation will be repeated only a finite number of times.

There could be a range of other applications and contingent rules. For example, the agents might be playing different games in different states, with play in one game contingent on the other's strategies either in the last play of the game now being played, or in the last play of the other game, or both.

### 6.4.3 Interim Summary

We see that repeated play can be a link from Nash equilibrium to cooperative play. If agents are involved in interactions that are likely to be repeated, and the agents are patient enough and have some foresight, then cooperative play may emerge as one of the equilibria in a supergame, that is, a game repeated an indefinite number of times.

## 6.5 ON SOME EXPERIMENTAL STUDIES

A number of experimental studies have addressed the predictions of the perfect equilibrium model in non-cooperative game theory. Two games

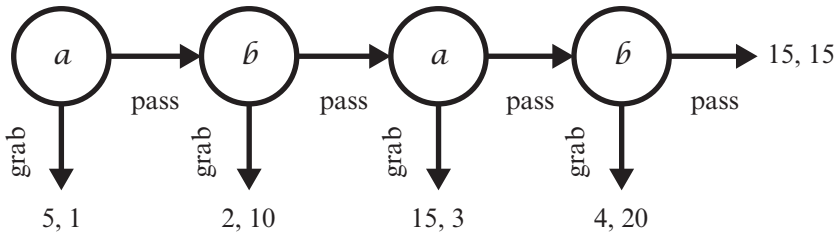


*Figure 6.4 Game 6.7: a Centipede Game*

are particularly important in this connection: the Ultimatum Game and the Centipede Game. On the whole, the experimental results disagree strongly with the predictions of the perfect equilibrium model, if it is considered as an empirical hypothesis. The Ultimatum Game has been discussed in historical context in Chapter 3. Here we will focus on the Centipede Game.

The "Centipede Game" (Rosenthal, 1981; McKelvey and Palfrey, 1992) is illustrated by Figure 6.4. The centipede is a game with two participants and a pot of money payoff dollars. The two participants will be $a$ and $b$. The game proceeds in stages, and at each stage one or the other of the participants must make a commitment. At the first stage player $a$ can either take or pass a money payment. If he takes it $b$ also gets a smaller payment. If $a$ passes at the first stage, $b$ has an opportunity to take a larger share of the payment, leaving $a$ the smaller share. However, if $b$ passes at the second stage, $a$ in turn gets an opportunity to take the larger share, and the game proceeds in this way. The two players alternate, as shown in Figure 6.4, where the numbers show the payoff to $a$ first and then to $b$. The game ends after some finite number of steps with each participant getting a specified share of the pot. The size of the total payment to the two players may increase with the number of stages the game continues. This could be a model of "roundabout" production in economics, in that "passing" the pot on an early round allows the resources generated in the first round to be compounded in the later rounds. In some studies the game has subsequent stages, and it may have many stages. If we visualize a game with 100 stages rather than four, the basis of the name "Centipede Game" becomes clear.

A cooperative solution to this game requires a sequence of behavior strategies "pass." Using backward induction it is clear that the subgame perfect equilibrium in this game is for $a$ to "take the money and run." Since $a$ knows that $b$ is a rational player, $a$ cannot expect that $b$ will pass on the second round and allow $a$ to grab the larger amount, 15, at the

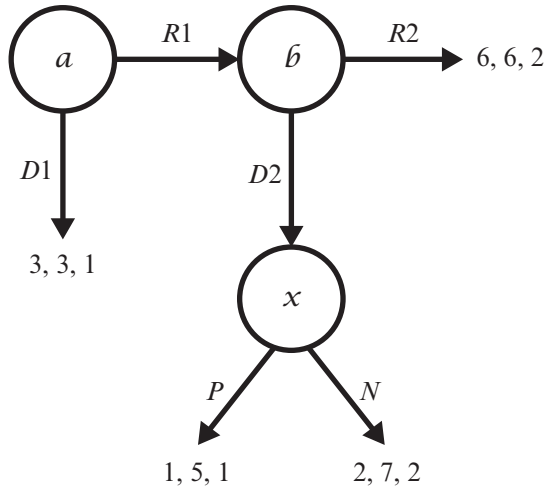


*Figure 6.5 Game 6.8, 6.9 in extensive form*

third stage, nor can *b* expect that if he passes *a* will allow him the opportunity to grab 20 at the fourth stage. In short, a rational agent cannot "outsmart" another rational agent.

Here, again, the experimental evidence does not agree, and a variety of outcomes (but only rarely the cooperative outcome of continuing play to the end) are observed. Many of the observed sequences of play are consistent with the possibility that one or both players are trying to "outsmart" one another – with at least one of them failing to do so. Suppose (as von Neumann and Morgenstern assumed, but Nash and Selten to the contrary) that many experimental subjects can commit themselves to contingent strategies such as "If *a* passes then I will pass at the second stage and then, if he passes again, I'll grab at the last stage for 20 rather than 10." If *a* conjectures that *b* has adopted that strategy, *a*'s best response would be "Pass at the first stage and then, if *b* passes, grab at the second for 15 rather than 5." On this interpretation, the evidence suggests that, at least in some circumstances, experimental subjects are able to commit themselves to particular contingent strategies.

In studies of reciprocity, variants on the centipede have given rise to important results. Figure 6.5 will serve as a generic diagram for two extensions of the centipede. These games take place in a maximum of three stages, although it can be cut short by either player.

In Game 6.8, decision node *x* is player *a*'s second decision node.[10] It introduces a "punishment" or "threat strategy," *P*, that gives *a* the

option of reducing *b*'s payoff at the cost of some reduction in his own, or of not doing so (arrow *N*). The basic subgame is the punishment node, and its Nash equilibrium is *N*. The game is then reduced to a two-step centipede with payoffs of 2, 7 at the second stage and *D*1 is the subgame perfect behavior strategy for *a*. The prediction of the subgame perfect equilibrium model is that the punishment node will make no difference and *a* will grab at the first opportunity, just as in a Centipede Game without the punishment node. However, such punishment is often observed, and cooperative outcomes (with payoffs such as 6,6) are more common in this game than they are when the third stage does not exist in the experiment. Since a decision for *P* at the punishment node leaves *a* worse off than he would be otherwise, *P* would be an instance of negative reciprocity.

Now suppose instead that the third node in Figure 6.5 is not the decision of player *a* but of a third player, *c*. The payoffs to *c* are shown third. This is an example of third-party punishment. As the choice of *P* makes agent *c* worse off, it would be an instance of altruistic punishment. It may also be referred to as "third-party reciprocity" (Fehr and Fischbacher, 2004). Once again, non-cooperative game theory predicts that such third-party punishment will never occur, and consequently the strategies of *a* and *b* will be the same as they would be if there were no third stage; but the experimental evidence does not confirm this prediction. Rather, third-party punishment is observed, and seems consistent with the hypothesis that the third parties place some value on reciprocity between the original two players, and punish deviations from it (Fehr and Fischbacher, 2004).

On the whole, experimental evidence does not favor the subgame perfect equilibrium as a general empirical hypothesis. On the other hand, we should observe that these experiments have themselves arisen from the subgame perfect equilibrium model. By drawing on that analysis, they have provided more precise evidence on human motivation than earlier experiments were able to supply. Thus, we may best regard the subgame perfect equilibrium as defining one extreme of a spectrum of forms of rationality that we may observe in human action, a point to which we return in Chapter 10.

## 6.6 SUMMARY AND CONCLUSION

Games in which play takes place as a sequence of decisions, so that the decision-maker at any stage can condition his/her decisions on the previous decisions of others, are called sequential games. For these games, Nash equilibrium can be refined, limiting the non-cooperative equilibria

to those in which the play for each subgame is Nash equilibrial or which are the limit of a series of games in which play is perturbed by a "trembling hand." These are "perfect equilibria." For sequential games, the "perfect equilibria" can be found by backward induction. Perfect equilibria are important in many economic applications, including games of market entry and entry limiting strategies, and principal-agent interactions. When we consider private decisions as games imbedded in the larger game of public policy, we rely on perfect equilibria to assure us that the non-cooperative behavior of the private agents will itself be equilibrial. Repeated play is also analyzed in terms of perfect equilibria, with the idea that a threat of retaliation will be credible only if it is subgame perfect. On this basis, conclusions can be drawn as to whether cooperative play is likely to emerge from repeated non-cooperative play in a particular case. However, experimental results suggest some caution, in that subgame perfect equilibria may not be realized when they conflict with reciprocity motives.

## NOTES

1. For further examples, see my introductory textbook, *Game Theory: A Nontechnical Introduction to the Analysis of Strategy*, South-Western, 2004, esp. Chapters 14, 15.
2. In this as in other cases, "principal" and "principle" may be confused, especially since "principal" is more usually an adjective. However, standard dictionaries concur that this is the correct usage for one who authorizes an agent to act on her or his behalf.
3. Note Game 4.2, Chapter 4.
4. In a more complex game such as Game 6.2, the set of subgames will include the repetitions of the original game, but may be a much larger set. The reasoning is slightly more complex in an example of that kind, but the conclusion is not changed.
5. It may be appropriate to discount the payments at this last stage to present value at some specific discount rate. The economic literature on repeated games tends to stress this, but it does not affect the qualitative results and we shall ignore discounting here.
6. If there is a definite time period between repetitions, then $\delta$ should reflect the time discount rate as well as the probability of repetition. See McCain (2004, p. 273) for details. If there is no definite time period between repetitions, or it is very brief, then time discounting may not be possible or significant. Some scholars avoid the idea of maximizing the time-discounted value of a sequence of payments in general, and substitute for it a criterion that a sequence of payments is to be preferred to another if the first eventually overtakes the second. The folk theorem has also been developed with the overtaking criterion (Rubinstein, 1979).
7. In what follows, I will use the term "trigger strategy rule," accordingly.
8. If payoffs are discounted for time, then a great enough time discount might offset the high probability of repetition, so that the tit-for-tat strategy rule will fail. Thus, many authors express the point this way: if the players are patient enough (have low enough time discount rates), then cooperation can be attained via tit-for-tat play. In this example, though, time discounts are likely to be small relative to probabilities that play will be discontinued that are in the range of 0.25 to 0.5.

9. Note that this discussion differs from Shapley's (1953) "Stochastic Games", in that for Shapley's model the probabilities depend on the strategies chosen, while for this discussion the probabilities are given.
10. For now, ignore the third payoff number.

# 7. Social mechanism design

The 2007 Nobel Memorial Prize in Economics was awarded just at the time this chapter was being drafted, and was awarded to Hurwicz, Maskin, and Myerson for their contributions to social mechanism design. Maskin and Myerson are well known as game theorists, and the scientific background document for the prize (Royal Swedish Academy of Sciences, 2007, p. 1) states "By using game theory, mechanism design can go beyond the classical approach," so arguably this is the third Nobel for game theory. "Mechanism design, Professor Maskin explained, can be thought of as the 'reverse engineering part of economics.' The starting point, he said, is an outcome that is being sought, like a cleaner environment, a more equitable distribution of income or more technical innovation. Then, he added, one works to design a system that aligns private incentives with public goals" (Lohr, 2007). Much of the work of social mechanism design has been within the scope of game theory, though (like bargaining theory) social mechanism design has a longer history than game theory does. Hurwicz (1973, p. 2[1]) traces the idea to utopian socialism.

Hurwicz's founding lecture was concerned (in its first three sections) with dynamics and with the exchange of information necessary to achieve efficiency or other objectives by particular mechanisms, supposing that people report honestly. Some mechanisms will require more or bigger (thus more costly) messages to attain efficiency, or even feasibility. The feasibility of central economic planning (at that time allegedly practiced in the Soviet Union) was a major concern. In the fourth and last section he addresses *incentive compatibility*, noting that agents may have incentives to lie, so that honesty may not be taken for granted. A mechanism free of the incentive to lie is "incentive compatible." This, he says (Hurwicz, 1973, p. 23) ". . . is a problem in the theory of games, in this case non-cooperative games without side payments." This has been the focus of most of the subsequent work denoted as mechanism design or implementation theory.

Mechanism design can be thought of in terms of imbedded games. We have a population of $n - 1$ "real" players, human agents with their own intentions, preferences, and potential courses of action. The $n$th player is the designer, an "artificial player." For the designer, the various strategies are the various "rules of the game" according to which the $n - 1$ "real

players" might play their game, such as riparian or non-riparian rights in Game 6.5. Since these various rulesets define games that are imbedded in the designer's game, perfect equilibrium in the designer's game means that each is in a Nash equilibrium. The "artificial player's" payoff may be an index of the efficiency or equity of the perfect equilibrium, or it may be 1 in case the outcome corresponds to a particular cooperative solution (in implementation theory) and 0 otherwise.

## 7.1 CUTTING THE CAKE

As an instance of mechanism design or implementation, we can turn to an idea much older than game theory: cutting a cake. The objective is that the cake should be divided equally – or if not equally, then fairly. Suppose the cake is to be divided between two identical persons, each of whom prefers more cake to less. One of the two is assigned to cut the cake, and the other gets to choose which piece he will take. Then the cake-cutter knows that he will get the smaller piece, and thus has an incentive to make the division as equal as possible. Equal division is incentive compatible, that is, consistent with non-cooperative decisions by the two recipients of cake. The rules of the game – one cuts and the other chooses – implement the objective of an equal division of cake.

Of course, the devil is in the simplifying assumptions, as usual. Let us make the problem a little more complex. We suppose the cake contains nuts, and the nuts are not randomly distributed – there are more of them on the right (let us say). Now suppose that the two agents are not alike but are of different *types*: agent $a$, who is to choose, likes nuts and might prefer a smaller piece if it had more nuts, while agent $b$ is indifferent with respect to nuts and just wants a bigger piece, regardless of the quantity of nuts. What is a "fair division" in this case? A division may be fair in this sense if it is non-envious (Foley, 1967): each person gets a piece of cake that he prefers to the piece the other person has, rather than vice versa. Now suppose $b$ is to cut and $a$ is to choose and $b$ knows $a$'s preferences. Then $b$ can cut the cake into two unequal pieces, with more nuts in the smaller piece, such that $a$ will choose the piece he prefers while $b$ is left the piece that he prefers. Thus a fair division (in this sense) is incentive-compatible.

But what if agent $b$ does not know which *type* agent $a$ is – whether agent $a$ likes nuts or is indifferent with respect to them? (Perhaps he even hates them.) $b$ needs this information to know how to cut, for the benefit of both. $b$ can ask $a$ what type he is – but can he trust the answer? Suppose for a moment that $a$ is, like $b$, indifferent to nuts. By saying that

he likes nuts, $a$ would be able to fool $b$ into cutting unequally. $a$ would then choose the bigger piece, and so be better off than he would be if he told the truth. Thus the rules of the game will have to be designed so that each agent will truthfully reveal what type he is (and that is to say, reveal whatever is relevant to the correct decision). This is the *revelation principle*. It may be that (for a particular class of games, or in general) there are no "rules of the game" that will do this, and this issue has been central to mechanism design.

## 7.2 NASH AND OTHER EQUILIBRIA AS OBJECTIVES OF MECHANISM DESIGN

In the simpler cake-cutting game, we have a non-cooperative game and a Nash equilibrium that, at the same time, satisfy the criteria for an equitable division; criteria that in themselves have nothing to do with the non-cooperative game. Broadly speaking, this is the objective of social mechanism design. With multiple Nash equilibria, as we have seen, there may be some uncertainty as to whether the agents will find their way to the right refinement of Nash equilibria. The mechanism would be more reliable if, for example, the Nash equilibrium were unique, or if it could be shown that every Nash equilibrium implements the cooperative solution concept or normative objective (Royal Swedish Academy of Sciences, 2007, p. 13; Maskin, 1999). Many studies in implementation theory seek a non-cooperative game for which the objective social state is the dominant strategy equilibrium. This would provide a very reliable implementation, since the dominant strategy equilibrium is essentially unique,[2] has a compelling justification in terms of self-interest, and is cognitively relatively easy. However, implementation as dominant strategies is a difficult objective, and it may not be possible even in principle to find a non-cooperative game that implements the objective as a dominant strategy equilibrium. This is illustrated by the game theoretic study of elections and voting, an important topic in itself for public policy. Where implementation in dominant strategies is impossible, another common standard is implementation as a Bayesian Nash equilibrium, since such equilibria lend themselves to learning by trial and error. However, multiplicity of equilibria may be a problem in this case, apart from the special case of "Maskin-monotonicity," (Royal Swedish Academy of Sciences, 2007, p. 13; Maskin, 1999), which is a key condition particularly in the study of elections.

## 7.3 A NEGATIVE RESULT: NON-COOPERATIVE GAME THEORY AND ELECTIONS

Discussion of alternative voting procedures has a long history, with eighteenth-century contributions from de Borda and Condorcet and in the nineteenth century by the author of *Alice in Wonderland*, the mathematician Charles Dodgson. Here is an example. Suppose there are three types of voters on a committee that will vote among alternatives A, B, and C. There are three of type 1, two of type 2, and two of type 3. Their preferences are shown as Table 7.1. Voting will be by majority rule with an agenda: at the first step a choice is made between A and B, and at the second step between the first stage winner and C.

If each voter votes his sincere first preference, A will prevail over B 5–2 and, at the second stage, C will be chosen 4–3. However, this is not a Nash equilibrium. By shifting the first stage votes to B, voters of type 1 can bring about a contest between B and C at the last step, and B will then prevail, leaving type 1 voters better off with B rather than C. Since neither types 2 nor 3 can then improve their outcomes by shifting from sincere first preference voting, this is the Nash equilibrium in this particular voting game. It is an instance of voting manipulation, and, since at least two of the three type 1 voters must shift to voting for B in order for this to work, it is manipulation by a coalition. In this connection, the question for mechanism design is: can we design a better voting mechanism that would be immune to manipulation in this sense? And the general answer is no.

### 7.3.1 Arrow's Impossibility Theorem

Any account of the study of elections in game theory and economics requires a background account of twentieth-century developments in welfare economics (McCain, 2007b). In *The Economics of Welfare*, Pigou had set out a systematic normative economics in terms derived from Mill's rule-utilitarian ethics. Many economists (following Pareto, 1971 [1906]) felt that Pigou's (1920) welfare economics assumed too much. In

*Table 7.1 Preferences for three types*

| Type | 1 | 2 | 3 | Votes |
|---|---|---|---|---|
| First | A | C | B | 3 |
| Second | B | A | C | 2 |
| Third | C | B | A | 2 |

particular, Millian utilitarianism assumed that utility is interpersonally comparable and additive, so that utilitarian values could support (limited) arguments for equalization of income. In place of utilitarianism, the "new welfare economics" of the 1930s and 1940s assumed only that individuals had transitive preferences, but that the preferences could not be expressed in numerical terms that could be compared interpersonally and therefore could not be aggregated as a basis for welfare judgments. However, many studies assumed that preferences could be aggregated to form a social-welfare function, that is, group preferences. This problem was addressed by Arrow (1951), and he proved a negative result: adopting a set of axioms that seemed to express reasonable conditions for a social welfare function, he proved that no function could satisfy them all. Arrow's conditions were (here I follow Satterthwaite, 1975, pp. 203–4)

(1) rationality, that the social preference is transitive, free of cycles;
(2) that it is nondictatorial, that is, that there is no individual whose preferences are decisive regardless of the preferences of all;
(3) independence of irrelevant alternatives, that is, the social preference between two alternatives depends only on the individual private preferences between them;[3]
(4) citizen sovereignty, that is, that any alternative might be chosen if it is widely enough preferred;
(5) non-negative responsiveness, that is that a shift of preferences by which one alternative rises in the preferences of some individuals could not result in its being lower in the social preference.

This "general possibility theorem" led to a great deal of ferment in welfare economics.[4] In particular, it must be that Pigou's maximization of aggregate utility violates one or another of them. In the discussion that followed Arrow's contribution, all of these were questioned, but the independence of irrelevant alternatives was especially a target of critics.

### 7.3.2 Elections

Arrow's result was not limited to voting systems. Indeed, his result dismissed judgments of the efficiency of markets even in ideal conditions, no less than judgments of the efficiency of elections. Nevertheless, Arrow made extensive use of election-type rules as examples, and it was widely conjectured that similar problems might arise in elections. This conjecture was given formal proof by Gibbard (1973) and Satterthwaite (1975), independently. Their contributions were different in detail, but complementary, and both drew importantly on Arrow's work. Feldman's (1979)

very readable exposition is also helpful, especially to the less mathematical reader.

Satterthwaite defined manipulation as a failure to vote in accordance with the individual's own preferences. The person might report a different order of preferences than his own if, due to the balloting procedure, such a report would give rise to a decision he prefers to the one that would follow from his own decision. If, for a particular voting procedure, this can never happen, then the voting procedure is "strategy-proof." For a strategy-proof voting procedure, "every set of sincere strategies is an equilibrium as defined by Nash" (Satterthwaite, 1975, p. 188). But Satterthwaite proves directly that this cannot be the case if the procedure satisfies the Arrow conditions. In fact, Arrow's independence of irrelevant alternatives and nonnegative responsiveness together are equivalent to strategy-proofness. Satterthwaite demonstrates that if a voting procedure is strategy-proof then it is dictatorial. Satterthwaite then shows that a rational, citizen-sovereign social welfare function can be derived from any nonmanipulable voting scheme, and conversely, and uses his own theorem to construct a new proof of the Arrow theorem.

Gibbard (1973, p. 589) expresses skepticism about the identification of nonmanipulation with "honest" voting: "Nothing in the structure of a game form tells us which strategy 'honestly' represents any given preference ordering." Accordingly, he characterizes nonmanipulation by an arbitrary function from preferences to votes, and asks whether such a rule of honest voting can be implemented as a dominant strategy. The answer is no, he argues, since a nontrivial voting game cannot have any dominant strategy equilibria whatever – honest or otherwise. Gibbard uses Arrow's result in his proof.

The connection of this "Gibbard-Satterthwaite theorem" to Arrow's result is quite close in a mathematical sense, but substantively it is less close than it is sometimes thought to be. Recall, Arrow's purpose was normative, and the objective was aggregation of preferences. Thus, the last three conditions assert a dependence of social preferences on actual individual preferences. By contrast, Gibbard and Satterthwaite assert (with the same conditions) a dependence of social decisions on preferences as expressed in voting. Their objective is to pose an empirical hypothesis, and one that seems to be true: that voting is always manipulable. The issue of misrepresentation of individual preferences does not arise for Arrow, but, on the other hand, the assumption of nonnegative responsiveness seems much more natural in the normative than the positive framework.

Manipulable voting processes can easily be seen to violate some of the Arrow conditions in simple examples. Return to the example of Table 7.1. Note that the best response of a type 1 voter depends on the fact that

voters of type 3 rank C below B. Suppose alternative B is eliminated, so that the choice is strictly between A and C. Type 1 voters can no longer gain by voting strategically and C becomes the winner. This violates the independence of irrelevant alternatives. The example points up a strong connection between the independence of irrelevant alternatives and manipulation of voting.

The negative result in the Gibbard-Satterthwaite analysis is not quite the whole story. Maskin remarks "The Arrow Theorem is too negative" in the context of voting theory.[5] Maskin proposes that elections with more than two alternatives should be conducted by a Condorcet evaluation: (1) voters record their preference orderings, rather than a single choice; (2) all possible pairwise choices are evaluated to determine which would win against the other; (3) if one alternative wins over all the others, it is the choice. But this procedure is not *decisive*. That is, there may be preference profiles such that for at least three alternatives, A is preferred to B, which is in turn preferred to C, and C is preferred to A. Indeed Table 7.1 gives an example of this. Maskin concedes that some other criterion would have to be used as a "tie-breaker" in such a case, but argues that, regardless of the tie-breaker, a procedure that always chooses a Condorcet winner will be more workable than any procedure that does not.

He draws this conclusion despite the Arrow theorem on the following reasoning: the Arrow theorem demands that a particular decision rule satisfy very demanding assumptions for *all possible* profiles of preferences over voters. However, he suggests, a procedure incorporating the Condorcet evaluation could satisfy the same assumptions over *a large proportion* of all possible profiles, while failing only on a subset of the possible profiles. In Maskin's terminology, a decision rule "works well" for a particular subset of preference profiles if it satisfies a list of assumptions[6] including independence of irrelevant alternatives and adding decisiveness for any of those profiles. In Dasgupta and Maskin (2008) Maskin compares simple majority rule (which he understands as the Condorcet evaluation) with plurality rule and the Borda count, a system that awards points depending on the individual's preference ranking and chooses the alternative with the most points. He shows that there are sets of profiles on which each of the three "works well," but that that Condorcet evaluation "works well" on the sets in which either of the others work well, while the others fail on some profiles for which Condorcet evaluation "works well." Thus, despite Arrow's negative conclusion, Maskin makes a strong case for group decisions according to the Condorcet evaluation.

There are a number of other important papers in the literature that reconsider one or another aspect of the Arrow-Gibbard-Satterthwaite impossibility theorems, generally in favor of some kind of majority rule.

In addition, there is some reason to believe that in elections with a large number of voters, the likelihood of manipulation decreases with an increasing number of voters (for example, Baharad and Neeman, 2002). On the other hand, casual empiricism suggests that even in elections with a large number of voters, manipulation is quite common.

## 7.4 CAP AND TRADE REGULATION

Tradable emissions controls have become a popular means of dealing with environmental pollution. This sort of program is often called a cap-and-trade program (Colby, 2000, p. 639). This approach does not really arise from the program of social mechanism design and relies less on game theory than on the much older competitive market theory of neoclassical economics (Hahn, 1989, p. 99). Nevertheless, it is similar in its overall objectives and seems to be one of the few ideas that commands a good deal of consensus across the divided American political spectrum (Broder, 2007).

In a "cap-and-trade" regulatory framework, polluters are permitted to emit pollutants up to some limit, or "cap," while the permits may be bought and sold among the different polluters who are regulated. The hope is that competitive markets for the permits will develop, allowing a market price that would indicate the least social cost of reducing pollution by one unit, so that pollution objectives would be met more cheaply and the polluters would face approximately socially optimal incentives to create and adopt less-polluting technologies. (See McCain, 1978, on the influence of market prices on technological trends toward environmental degradation.)

We now have some experience of cap-and-trade programs. In the 1980s Hahn (1989) found some impact on costs, though less than theory would have anticipated, but no significant impact on environmental quality and few trades. To some extent that could be attributed to limits and imperfections in the regulatory programs and lack of competition in the permit markets. Lack of competition in permit markets is particularly likely to be a problem if the objective of policy is to reduce local concentrations of pollutant, since the localities are quite likely to contain only a few large sources of pollution. A decade later, Colby (2000, p. 642) reported that, after a delay of years, trading (in sulphur dioxide emission permits) "took off," and "Twenty years after regulators, utilities, and environmental advocates were introduced to air emissions trading opportunities . . . cautious exploration has evolved into a mature, productive allowance market." Trade in water access has been less successful (ibid., pp. 643–5).

Despite intense resistance, some success is reported in markets for tradable fishing quotas (p. 647).

This experience indicates that tradable permits may be helpful, but that the form and consequences of such policies are as much a matter of political economy as of neoclassical economics or game theory, that oligopoly in permit markets is often likely to be a problem, and that successful permit programs may require a long learning period and may face crucial obstacles in customary norms, particularly (as in fisheries) where neither regulation nor trading is customary. Considering these difficulties, non-cooperative game theory could be useful in taking oligopoly and thin markets into account, but the others are as foreign to non-cooperative game theory as they are to neoclassical economics. A discipline of social mechanism design will need to be enriched with elements from cognitive science, political theory, and perhaps cultural anthropology if it is to address these problems.

## 7.5 ASSESSMENT OF MECHANISM DESIGN

Marx (1845) wrote, "Philosophers have hitherto only interpreted the world in various ways; the point is to change it."[7] That is the standard by which mechanism design invites evaluation. While a number of designed mechanisms, such as the Groves-Ledyard mechanism for public goods supply (Groves and Ledyard, 1977) remain untried in the real world, the long history of elections provides some experience on the reliability of mechanism design. The universally acknowledged success story for mechanism design and implementation theory, however, is the design of auction mechanisms (Royal Swedish Academy of Sciences, 2007, p. 15; Lohr, 2007; Glenn, 2007; Zaretsky, 1998; McMillan et al., 1997). Accordingly, this seems to be the appropriate testbed for the theory of mechanism design.

Of course auction theory, like bargaining theory and voting theory, has a history before game theory or the literature on mechanism design. In particular, though, the work of Vickrey (1961) gave rise to the "Vickrey auction," which fits the description of mechanism design. The success of Ebay can partly be attributed to its adaptation of Vickrey's ideas. Vickrey suggested a sealed-bid, second price auction as having desirable properties of efficiency and seller revenue maximization, though (in an expected value sense) several major auction designs could be seen to be equivalent. But Vickrey auctions have their desirable properties in the case of individual private value auctions, that is, auctions (such as those for collectibles) in which individual values are subjective, uncorrelated, and known only to the

individuals themselves. For auctions in which the items sold have an objective value, imperfectly known to the bidders, inefficient overbidding might occur – in effect, the auction being won by the bidder who most overestimates the value of the item being sold. This is known as the "winner's curse" (Camerer, 1987). Otherwise, when more than one item is sold, and the items sold might be complements or substitutes, sealed bid auctions could be less efficient than ascending auctions, since the latter would provide more flexibility to coordinate purchases. In general, of course, if one buyer were large enough to exercise monopsony power, or if there were collusive (cooperative) behavior among the bidders, Vickrey's results might not apply.

If we consider auctions of resources or privileges that would enter into production or marketing, such as the privileges of drilling for petroleum on government-held territory or of using electromagnetic spectrum, it is unlikely that the values will be subjective and uncorrelated; and it was the auctioning of electromagnetic spectrum that put mechanism design "on the map." As early as the 1980s, New Zealand had begun to allocate electromagnetic spectrum by auction, using a Vickrey auction design. However, some of these auctions had disappointingly small yields as buyer monopoly power and/or collusion limited the bidding (Zaretsky, 1998). In 1993, three well-known economists associated with the consulting firm MDI Associates invented the simultaneous ascending auction, according to the company's own history (Market Design, Inc., 2007). The simultaneous ascending auction was intended to avoid the "winner's curse" and facilitate adjustment of bids to complementarities among the items being sold, by learning from one another's bidding. This auction design was adopted by the Federal Communications Corporation for the allocation of licenses to use electromagnetic spectrum in the United States. This series of auctions was considered very successful and largely gave rise to the conception of mechanism design as a practically successful field.

In 1994–97, there were 13 auctions based on the simultaneous ascending auction. It seems clear that they were indeed successful on the whole, and two were quite remarkably successful (Cramton, 1997). However, a series of European auctions of third-generation telecommunications spectrum licenses were less successful, on the whole, and Klemperer (2002) put the blame on mechanism design, saying (p. 3) "Good auction design is really good undergraduate industrial organization; the two issues that really matter are attracting entry and preventing collusion," and adding in a footnote, "By contrast, a graduate knowledge of modern auction theory is at best of lesser importance and at worst distracting from the main concerns." Klemperer observes that ascending auctions permit collusion for the same reason they avoid the winner's curse: they allow bidders to learn from the bids of others, and this learning may facilitate collusion. Thus, he argues for

sealed-bid auctions or auctions with a sealed-bid stage although "Auction design is not 'one size fits all'" (pp. 3–4). Among the European auctions, all but one used an ascending-auction design. Of those, the first, the English, was a major success, but by contrast the following Netherlands, Italian, and, especially, Swiss auctions could be characterized as failures. German and Austrian auctions differed somewhat from the others, and of those, the Austrian auction too was characterized as a failure. The one exception, Denmark, used a sealed-bid auction design and was successful.

To some extent the European failures were the results of bad timing. They took place during or after the dot-com collapse of 2001. However Klemperer's discussion makes it pretty clear that they failed also because of neglect of the "two issues that really matter." Certainly collusion is a cooperative phenomenon: as we recall, price competition is an exceptional case in which cooperative behavior (among those on one side of the market) is undesirable. The other issue that really matters, entry, is somewhat complex in itself. Klemperer seems to mean two things: first, entry into the auction itself, and second, entry into the market. On the one hand, several of the European auctions failed in part because so few of the potential competitors actually participated in the auction. Broadly speaking, in assembling a group to participate in the auction, the governments who conducted the auctions were assembling a cooperative coalition to act on a common strategy, although the common strategy itself demanded non-cooperative action from the bidders. On the other hand, for a firm to enter an industry is to seek to form cooperative coalitions (for exchange) with potential buyers. Thus, it seems fair to say that the failed European auctions failed because auction design was narrowly based on non-cooperative game theory and neglected cooperative game perspectives.

On the other hand, the relatively successful British and Danish auctions incorporated Klemperer's concerns, and the US auctions clearly were characterized by predominantly competitive bidding (Cramton, 1997). The American market, larger than those of the European countries, was (on Klemperer's reasoning) more favorable to competitive bidding. Moreover, the two auctions that were most successful had special circumstances promoting competition that were not predicted. In the second auction, of narrowband licenses, special arrangements to encourage minority and female participation led to vigorous bidding by those groups that spilled over into the bidding of others (Cramton, 1997, pp. 433–4, 451). In the December 1994 broadband auction the participation of Craig McCaw, seemingly seeking an opportunity to re-enter the telecommunications market after selling his company, was a highly competitive factor (Cramton, 1997, pp. 454–5). All in all, it seems that Plott (1997, p. 637) goes a bit far when he writes "The overall success of the auctions must be attributed to . . .

economic theorists, applied economists, FCC lawyers, and the FCC staff." Or if we accept this assessment, then the failures of some of the European auctions must be similarly attributed.

On Marx's criterion, then, mechanism design must be seen as partially successful. It has indeed changed the world. However, the change in the world has not entirely been the one intended: mechanism design has not quite "made history consciously." Yet this is hardly a damning conclusion. Errors occur in all human activities, and the key to progress (from a pragmatic point of view) is that they are corrected. An improved non-cooperative game theory, using alternatives to the Nash equilibrium or selectively incorporating ideas from cooperative game theory and other disciplines, may yet provide public policy with a reliable tool of prescription that can be used to specify policies that could advance a wide range of policy priorities. There is some literature in implementation theory that points in that direction (for example, Moulin and Peleg, 1982). Nevertheless, it seems fair to say that non-cooperative game theory has not yet become that tool.

## 7.6 CHAPTER SUMMARY

In mechanism design, non-cooperative game theory is turned from diagnostic to prescriptive use. This demands more of a theory than diagnostic application does. Non-cooperative game theory assumes that human decisions are always non-cooperative, and there is little doubt that they sometimes are, but it is at least possible that they are also sometimes cooperative. If this is so, then problems that would arise in a hypothetical non-cooperative world may also arise in the actual world. In the actual world, these problems may be less pronounced, or it may be that non-cooperative behavior even on the part of a minority of the population will result in a non-cooperative outcome, in particular cases. On the other hand, a mechanism design based on assumptions of non-cooperative behavior may be undermined when cooperative coalitions are formed to exploit it, as in the case of collusive pricing and bidding strategies. Thus, non-cooperative game theory may be far more reliable as a diagnostic tool than it is as a prescriptive tool, and the experience of auction theory seems to support this conjecture.

## NOTES

1. Hurwicz's paper is sometimes dated as 1972. It was presented as the Richard Ely lecture at the annual conference of the American Economic Association in 1972, and published

in the proceedings volume. At that time, the American Economic Association met in late December of each year and the proceedings volume was issued in the following May. Thus, 1972 is the correct year for the lecture but 1973 is the year for the print publication.

2. If more than one strategy sets are dominant strategy equilibria, all will have the same payoffs.
3. This assumption is stated in different ways, for mathematical analysis, by different authors. Following Dasgupta and Maskin (2008), if alternative A is chosen from the set of available choices X, and if X′ is a subset of X, A is an element of X′, then A is chosen from X′. In other terms, dropping alternatives that are not chosen out of the set of alternatives does not change the set chosen.
4. Important further developments arose from the work of Foley (1967), who demonstrated that the ordinal preference theory could be used to make distributive judgments. Indeed, the distributive norms arising from models like Foley's tend to be more equalitarian than the utilitarian ones are, and parallel the ideas of the philosopher John Rawls (1971). John C. Harsanyi (1975) drew on Rawls's ideas but also on the reformulation of utility theory in the von Neumann-Morgenstern tradition and on Bayesian concepts of rationality, and argued that social welfare could after all be based on a summation of individual utilities. Amartya Sen (1985) has proposed conditions less limiting than Arrow's that allow the possibility of a consistent majoritarian social welfare function. Sen, however, rejects what he describes as the welfarism of both the old and the new welfare economics, by which he means the supposition that the goodness of a social system depends only on the welfares of individuals in those social systems. In addition, Sen would have data on the capacities and perhaps freedoms of individuals reflected in the normative evaluation of economic society.
5. This quotation is from memory, from an oral address at Haverford College, Haverford, PA, 16 November 2007.
6. While the list differs somewhat from (1)–(5) above, it is similar in conception. The literature includes a number of different equivalent and near-equivalent axioms expressing the desiderata of good collective decision processes.
7. One finds a considerable variety of versions of this quotation, partly, no doubt, due to alternative translations from the German.

# 8. Superadditive games in coalition function form

This chapter reviews some concepts from what might be called near-consensus cooperative game theory. The objective of the chapter is primarily expositional. Apart from expression, examples, arrangement, and some critical comments, the chapter is not intended to be original.

For this chapter, the game is primarily represented in coalition or characteristic function form. That is, the game comprises a set N of players, $a_1,\ldots,a_n$; $a_i \in N$; the enumeration of all subsets of that set, the potential coalitions, and a mapping from subsets to real numbers, the characteristic or coalition function, which gives us the value attainable by each coalition. The value of a coalition C will be denoted as v(C) or $v\{a_1, a_2, \ldots\}$ where $a_1, a_2, \ldots$ are the members of coalition C. As we recall, von Neumann and Morgenstern identified this with the assurance value. The key point is that the value of a coalition is well defined and depends only on the membership of the coalition. We also adopt the assumption, from von Neumann and Morgenstern, that the game in coalition function form is superadditive; that is, that the value of a merged coalition is no less than the sum of the values of the merged coalitions acting separately.

## 8.1 SOLUTION CONCEPTS

For a superadditive game in coalition function form, the only rational arrangement is the grand coalition. If the grand coalition is formed, nothing can be lost (since the grand coalition must have a value no less than those of any proper coalitions into which it can be decomposed) and something will usually be gained. All that remains is to determine how the value of the grand coalition will be divided among the decision-makers. As we recall from Chapter 3, there are several such solution concepts.

### 8.1.1 The Core and Related Concepts

Probably the most widely discussed solution concept for games in coalition function form is the *core*. The simple idea behind the core of a cooperative

game is that no group can be denied the value that they could obtain if they were to form a coalition and act independently of the rest. As an illustration, consider Game 2.5. A singleton coalition that would produce the public good would then have a value of at most 4, so will not produce the public good. The singleton coalition would face a unified opposition, a two-person coalition, capable of producing two units of the public good. The opposition coalition would refuse to produce the public good, presumably in order to increase its bargaining power, since producing the public good would raise the value of the singleton to 7 or 9. Therefore, $v\{a\} = v\{b\} = v\{c\} = 5$. A two-person coalition that would produce the two units of public good would be worth 12, whereas if it does not produce its value is 10. Moreover there is nothing the opposition singleton coalition can do to reduce the two-person coalition's payoff below 12, so the two-person coalition will choose to produce the public good[1] and $v\{a,b\} = v\{b,c\} = v\{a,c\} = 12$. The grand coalition of all three agents will be worth 24 if it produces three units of the public good and less if not, so it will produce them and $v\{a, b, c\} = 24$.

The payoff to agent $j$, after side payments are made, is denoted by $x_j$. A set of payments $x_j$ for the n players in the game, consistent with the value of the grand coalition, is called an *imputation*.[2] Accordingly, suppose $x_a = 5$, $x_b = 5$, $x_c = 10$. Then $x_a$ and $x_b$ can instead form a two-person coalition and earn 12, which they can divide among themselves. Thus we exclude the imputation 5,5,10 from the core. In general, by the same reasoning, we exclude from the core any schedule of payoffs that does not satisfy:

1.1. $x_a \geq 5$
1.2. $x_b \geq 5$
1.3. $x_c \geq 5$

1.4. $x_a + x_b \geq 12$
1.5. $x_a + x_c \geq 12$
1.6. $x_b + x_c \geq 12$

1.7. $x_a + x_b + x_c \leq 24$

Adding inequalities 1.4–1.6 we obtain $2(x_a + x_b + x_c) \geq 36$, that is, $x_a + x_b + x_c \geq 18$. Comparing this with inequality 1.7, we see that there are infinitely many imputations that satisfy the criteria for the core in this example. In particular 8,8,8; 8,6,10; and 12,6,6 all are members of the core.

Here is another example, Game 8.1. Once again it will be a three-person game and all singleton coalitions are worth 5 if no production takes place. There are two techniques of production, both of which have economies of

scale so that they can be undertaken only by coalitions with two or more members.[3] Technology 1 generates profits of 4 for those who undertake it but produces a polluting waste that has to be assigned to some individual agent (who need not be a member of the group that undertakes production with technology 1) and reduces that person's payoff by 5. Technology 2 generates a profit of 3 and no waste.

A singleton will face a united opposition that can reduce the singleton's value to zero by producing with technology 1 and assigning the waste to the singleton. Therefore, $v\{a\} = v\{b\} = v\{c\} = 0$. A two-person coalition can achieve a value of 14 by producing using technology 1, and there is nothing the opposing singleton can do to reduce the two-person coalition's payoff below 14. Therefore, $v\{a,b\} = v\{b,c\} = v\{a,c\} = 14$. The grand coalition has a value with no production of 15, with technology 1 of 14, and with technology 2 of 18. Therefore technology 2 will be used and $v\{a,b,c\} = 18$.

For this game, in order to prevent any two-person group from dropping out and shifting to technology 1, $x_a + x_b + x_c \geq 21$ is necessary. Since, however, $x_a + x_b + x_c \leq 18$ is also necessary, there are no imputations that satisfy the criteria for the core of the game. That is, the core for this game comprises the null set. This is often expressed by saying "the core does not exist," but strictly speaking, the core always exists, although (as in this case) it may be null.

Both of these games are symmetrical, but this needs not be so. Of course, nonsymmetrical games, in which coalition values depend on the individual members of the coalitions, will be more complex, and in some cases very much so.

There are a number of properties that we might like a solution to have. Two of the most important are that it should never be null and should correspond to a unique imputation. Clearly the core satisfies neither of these. However, there are a number of desirable properties that it does have.

Suppose we have two games played by the same set of players, $\Gamma = (N,v(C))$ and $\Theta = (N,w(C))$, and there are constants $\alpha$ and $\beta$ such that for any coalition C $w(C) = \alpha + \beta v(C)$. The two games are said to be *strategically equivalent.* Suppose then that whenever $\mathbf{x}$ is a solution of $\Gamma$, $\alpha + \beta\mathbf{x}$ is a solution of $\Theta$. Then the solution is *covariant under strategic equivalence.* In more ordinary terms, it says that the solution will be unchanged by a change in the scale of measurements of payoffs, and that is persuasively a good property for a solution to have. The core has this property[4] (Peleg and Sudhölter, 2003, p. 25).

The core also has a property of anonymity, which means that the solution does not depend on the identities of the players except so far as their contributions to the values of coalitions are concerned. To see how this

might fail, consider a "toy" solution concept, which is meant only as a bad example. Assign $x_a = v\{a\}$, $x_b = v\{a,b\} - v\{a\}$, $x_c = v\{a,b,c\} - v\{a,b\}$, and so on if there are more than three players. Now consider another game, $\Theta$, that is a permutation of $\Gamma$; that is, we simply take $a$, $b$, and $c$ in a different order, such as $c$, $b$, $a$. Nevertheless $w\{a, b\} = v\{a, b\}$ and so on. But if we apply the toy solution concept using the new order we have $x_c = v\{c\}$, $x_b = v\{a,c\} - v\{c\}$, $x_a = v\{a,b,c\} - v\{a,c\}$. For the toy solution concept, the solution depends on how the players are ordered. A solution concept that is independent of this ordering is *symmetrical* and if it is similarly independent of any identity the agents may have apart from what is expressed in the coalition function, it is *anonymous.* The core has these properties (Peleg and Sudhölter, 2003, p. 26).

If a solution concept gives each agent a payoff no less than he could get acting as a singleton coalition, it is said to have *individual rationality*. The core has this property (Peleg and Sudhölter, 2003, p. 27).

Thus, despite its shortcomings, the core has some properties that we do want to find in a solution, and it captures the idea that a group of players will in general get at least what they can obtain by acting independently.

### 8.1.2 Arbitrational Concepts

In 1950–53 two solution concepts for cooperative games were proposed, both of which (unlike the core) provide unique solutions that are never null. These were due, respectively, to Nash and Shapley. Both were derived from systems of axioms that describe properties of a solution that might be considered reasonable or appropriate. Luce and Raiffa (1957) suggested that they might be interpreted as frameworks for arbitration, in that an arbiter would consider the properties of the decision in deciding on the distribution of payoffs among the group.

#### 8.1.2.1 Nash bargaining

Nash begins by assuming that the payoffs to two interdependent decision-makers must fall within a feasible set that is convex and compact.[5] These properties assure that the assignment of payoffs to the two persons will be unique and are trivially satisfied for games in coalition function form. (Nash did not assume transferable utility.) He assumes that the decision will have the properties of:

(1) Individual rationality, that is, each agent receives at least as much as he could obtain if there is no agreement.
(2) Pareto-optimality, that is, the decision cannot be improved on for both decision-makers simultaneously.

(3) Independence of irrelevant alternatives (see Chapter 7.2.1).
(4) Solutions are covariant under strategic equivalence.
(5) Symmetry: that is, if the two bargainers are interchanged, the chosen payoffs are interchanged accordingly.[6]

Given these assumptions Nash shows that the vector of chosen payoffs can be computed as the solution to a maximum problem. Let $x^*$, $y^*$ be the payoffs to the two bargainers if there is no agreement and $x$, $y$ be their payoffs from the agreement. Then $x$ and $y$ will be chosen so that the product $(x - x^*)(y - y^*)$ is maximized among all the feasible pairs $(x, y)$. This solution assigns unique payoffs to the bargainers and is never null. However, it directly applies only to two-person games and makes no allowance for differences in bargaining power. There are several proposals of extensions to more than two players, but none seems widely accepted (for example, Harsanyi, 1963; CGT4, pp. 325–56).

#### 8.1.2.2 Shapley value

Shapley proposed a solution concept that would associate a payoff with each of n agents in a coalition function game, where n could be any positive integer. Shapley first adopts a series of three axiomata (CGT, p. 71) that are regarded as necessary characteristics of a solution. In ordinary language, these are that (1) nothing depends on the identity of a player, as distinct from the role (dealer, maker of the opening lead play, and so on) that the player takes in the game, (2) the payoffs add up to the total payoff of the grand coalition in the game, and (3) if the same players play two different games, the values in the merged game are the sum of the values in the two games. He then shows (CGT, pp. 71–4) that these conditions are uniquely satisfied by an algebraic, permutational formula,

$$\phi_i(v) = \sum_{\substack{S \subseteq N \\ i \in S}} \gamma_n(s)\,(v(S) - v(S - \{i\})) \tag{8.1}$$

where $\phi_i(v)$ is the value assigned to player $i$ in the game characterized by $v$, and $\gamma_n(s)$ is

$$\gamma_n(s) = \frac{(s-1)!(n-1)!}{n!} \tag{8.2}$$

where $s$ is the number of players in coalition $S$, $n$ the total number of players in the game, and ! denotes the factorial of the number. Thus, in ordinary terms, the value will be the weighted sum of the individual's contributions to the value of a coalition, for all coalitions in which he might

participate. The weights do not lend themselves to a simple ordinary language explanation. However, Shapley also offers (CGT, pp. 78–9) what he describes as a bargaining process that would generate the value (perhaps following Nash's 1953 example). He writes:

> The players . . . agree to play the game v in the grand coalition, formed in the following way: 1. Starting with a single player, the coalition adds one player at a time until every player is admitted. 2. The order in which the players are to join is determined by chance, *with all arrangements equally probable.* 3. Each player, on his admission, demands and is promised the amount his adherence contributes to the value of the coalition . . . The expectations under this scheme are easily worked out. (emphasis added)

These expectations are just equation (8.1) above, where $\gamma_n(s)$ is the probability that the corresponding payment will be the one offered.

The Shapley value has a number of desirable properties:

(1) It is Pareto-optimal.
(2) It is covariant under strategic equivalence.
(3) While it is not necessarily anonymous, the Shapley value has a property of symmetry: if the players are permuted, without any other change in the game, their Shapley values are unchanged. (Note that this is not true for the "toy" solution concept of Section 8.1 above. The difference arises from the fact that the "toy" assumes a particular ordering of the agents, while the Shapley value averages over all such orders.)
(4) It has an equal treatment property: suppose that two agents make the same contribution to all coalitions; then they are assigned the same Shapley value.
(5) It is additive (this is assumed in the introductory paragraph of this section).
(6) It has a null player property: if an agent adds nothing to any coalition then his Shapley value is zero.[7]

The additivity property is crucial and is a defining property of the Shapley value. It may also be the least compelling to intuition.

### 8.1.3 Nucleolus

The core and the Shapley value seem to be the most widely discussed and applied solution concepts in cooperative game theory. Of the several other concepts, this book will discuss only the nucleolus. Like the Shapley value, the nucleolus is never null and assigns a unique net payoff to every agent

in the game. The nucleolus has the further property that if the core is not null, the nucleolus is an element of it. Thus, the nucleolus can be thought of as a core assignment algorithm; that is, as a basis for singling out a particular imputation in the core when the core is not unique. The nucleolus will not be discussed here but will be considered in detail in Chapters 12 and 13. The nucleolus can be computed by linear programming, although there are some complications, even for a game as simple as this one. The nucleolus has some desirable properties. Like the Shapley value and the core, it is covariant over strategic equivalence and has an equal treatment property. Like the core, the nucleolus is anonymous. It is not additive and lacks the null player property.[8]

### 8.1.4 Interpretations of the Solution Concepts

The solution concepts are usually presented as mathematical forms, with their interpretation largely left open. Indeed, they may be susceptible of at least two interpretations, and it may be that more than one solution concept might be adopted for different interpretations.

#### 8.1.4.1 Stability interpretation

Returning to Game 2.5, we have said that imputation $x_a = 5$, $x_b = 5$, $x_c = 10$ should be excluded, since $x_a$ and $x_b$ could then secede and earn 12. But what are we to make of this argument? Critics of the core concept have questioned this criterion along the following lines: if *a* and *b* are members of the grand coalition, then they have committed themselves to it, and to some prearranged imputation. For them to opportunistically abandon the coalition to increase their payoffs as a group is then seen as inconsistent for a cooperative game analysis. The cooperative game solution should represent a binding contract. This *binding contract interpretation* might support Nash bargaining, the Shapley value, or the nucleolus, against the core and related concepts.

Indeed, we might say that the core concept is based on coalitional egoism. But isn't coalitional egoism something we are likely to see in the real world, sometimes?

One possible response to the criticism is that commitments, however binding, are not forever. Thus, even if *a* and *b* were to remain with the grand coalition for a time, eventually their commitments would expire and they would be likely to make other arrangements. On this sort of interpretation, the core is a concept of stability. The empirical prediction would then be that imputations outside the core are unstable, and therefore *relatively* unlikely to be observed, since they will be short-lived if they do occur.

#### 8.1.4.2 Rhetorical Interpretation

Another possible interpretation is that the abandonment of the grand coalition by $\{a,b\}$ might not literally take place, but might be a threat made in the course of bargaining over the division of the value created by the grand coalition. This interpretation has been primarily associated with other solution concepts, such as Aumann and Maschler's (1964) bargaining sets (which will be beyond the scope of this chapter) and the nucleolus. Both Nash and Shapley referred to bargaining power as motivation for their models. But the rhetorical interpretaton could be applied to the core as well. The empirical prediction then would be that imputations outside the core would be rarely or never observed, since they would be rejected in the bargaining that precedes the formation of coalitions.

## 8.2 THE PROBLEM OF APPLICABILITY

By comparison with non-cooperative game theory, at least, there have been relatively few applications of cooperative game theory. In chapters to follow this book will argue that the simplifying assumptions lead to a theory that is simply too abstract to be useful in a wide range of applications. However, there are a few important applications in economics. We will review three: games of exchange, games of production, and applications to the allocation of cost in a multidivisional organization. The first two are applications of the core, while the last begins from the Shapley value.

### 8.2.1 The Market as Implementation of the Core

Among the most important applications of cooperative game theory is the study of games of exchange. In a game of exchange, there are two or more types of players, who differ in their endowments of particular goods and services or in their preferences or both. Coalitions are formed for the purpose of making reciprocal transfers of the goods and services, that is, exchanges. The benefit of doing so is that each agent may at the end find himself with a collection of goods and services that he likes better than the endowment he began with.

Most contributions to this literature do not assume transferable utility, but instead adopt the nontransferable utility approach of Shapley and Shubik (1952). For simplicity, we will consider an example of trade in indivisible units of two goods.[9] Suppose, then, that we have two traders interested in exchanging olive oil for wine. (This will be Game 8.2.) Traders of type $a$ are endowed, at the beginning of the game, with three

*Table 8.1 Preferences for a game of exchange*

| Barrels of oil | Barrels of wine | *a*'s preferences | *b*'s preferences |
|---|---|---|---|
| 0 | 0 | 16 | 16 |
| 0 | 1 | 15 | 15 |
| 0 | 2 | 14 | 14 |
| 0 | 3 | 12 | 12 |
| 1 | 0 | 13 | 13 |
| 1 | 1 | 11 | 11 |
| 1 | 2 | 7 | 8 |
| 1 | 3 | 5 | 5 |
| 2 | 0 | 10 | 10 |
| 2 | 1 | 8 | 7 |
| 2 | 2 | 6 | 4 |
| 2 | 3 | 3 | 2 |
| 3 | 0 | 9 | 9 |
| 3 | 1 | 4 | 6 |
| 3 | 2 | 2 | 3 |
| 3 | 3 | 1 | 1 |

barrels of olive oil, while traders of type *b* are initially endowed with three barrels of wine, and the barrels cannot be divided. With just two traders, then, an individual may find himself with 0, 1, 2, or 3 barrels of oil and 0, 1, 2, 3 barrels of wine. There are sixteen such combinations and, for the purposes of our example, we need to know the preferences of both players with respect to all sixteen. These are shown in Table 8.1. The preferences are expressed as first, second, and so on, so smaller numbers are better. Thus, for example, the third column tells us that a player of type *a* prefers one barrel of oil and three of wine (fifth preference) to two of oil and one of wine (eighth preference), while the last column tells us that a player of type *b* prefers two of oil and one of wine to one of oil and two of wine.

Suppose, then, that it is proposed to exchange two barrels of oil for one of wine. This would please the type *b* trader, raising him from his 12th to his 4th preference, but it would reduce the type *a* trader from his 9th to his 11th preference; so the type *a* trader would veto the trade. Suppose, on the other hand, that it is proposed to trade one barrel of oil for two of wine. This would raise the type *a* trader to his 6th preference and the type *b* trader to his 11th. Thus we may say that the allocation[10] that results from the one-oil-for-two-wine trade, two oil and two wine for type *a* and one oil and one wine for a type *b*, *dominates* the initial allocation. In general, an allocation **x** will dominate an allocation **y** if there is a coalition at least one member of which is better off, and none worse off, with **x** than with **y**.

With only two traders in the game, we need consider only the grand coalition and the singleton coalitions.

As before, the core will consist of all allocations that are undominated. As before, the core may not be unique. For this game, with just one trader of each type, we have

2, 2 for $a$ and 1,1 for $b$ resulting from a 1-for-2 trade
1, 2 for $a$ and 2,1 for $b$ resulting from a 2-for-2 trade
1, 3 for $a$ and 2,0 for $b$ resulting from a 2-for-3 trade

In each of these cases, neither person can do better as a singleton, that is, if no exchange takes place. Thinking in terms of prices, or rates of exchange, we see that the price of a barrel of wine can vary from half a barrel of oil to 1 barrel of oil within the core. (At a one-to-one exchange rate, the exchange of one for one is dominated by the exchange of two for two, as an exchange of one for one leaves each with his 8th preference while the two-for-two exchange leaves each at his 7th. Put otherwise, at a price of 1, each person will offer two units for trade, and this is the market equilibrium.)

Now suppose that we have two traders of each type, and suppose that the allocation proposed is that type $a$'s get 2, 2 and $b$'s 1,1, corresponding to a price of ½. Instead, consider a coalition of one $a$ and two $b$'s, and of the $b$'s, suppose $b_1$ transfers 2 of wine to $a$, $b_2$ transfers 1, and $a$ transfers 1 of oil to each. This leaves $a$ with 1, 3, his 5th preference; $b_1$ with 1,1, his 11th, and $b_2$ with 1,2, his 8th. $a$ and $b_2$ are better off than they were in the proposed allocation, and $b_1$ is no worse off. (If we allowed wine to be divided, $b_2$ could offer $b_1$ a cup or two from $b_2$'s second barrel, to make it worth $b_1$'s while to join the coalition.) Thus, the three-person coalition permits an allocation that dominates the proposed 1-for-2 allocation, and the 1-for-2 allocation (and the price of ½) is no longer in the core. What we see is that with more traders, we may have more complex coalitions, and these impose more constraints on the allocations that can belong to the core, so that the core is smaller in a larger game.

Now, suppose that we have three agents of each type, and the proposed allocation gives $a$'s 1,3 and $b$'s 2, 0, as in the 2-for-3 exchange and the price of 2/3. Suppose instead that a coalition is formed of two $a$'s and three $b$'s with the nine barrels of each type allocated so that each $a$ gets 2,1, two $b$'s get 1,2, and the third $b$ gets 1,3. This means that the $a$'s are at their 4th preference, rather than their 5th as in the proposed allocation, and two $b$'s are at their 8th and one at his 5th preference rather than their 10th, as in the proposed allocation. We now have a $2a$ and $3b$ coalition dominating the 2-for-3 exchange, and the price of 2/3 is no longer in the core. Once

again, the larger game allows for more complex coalitions that impose more constraints on the allocations in the core, resulting in a smaller core.

In the literature on market games, as in this example, the objectives are vectors of quantities of different goods and coalitions are formed for reallocation of the initial endowments of goods. Usually goods and services are assumed to be divisible and the conventional neoclassical assumptions are made: the individual may be indifferent between two vectors of goods and services, but preferences are convex, meaning that a weighted average of two vectors of goods and services will be preferred to either of the two vectors that are averaged.[11] It is then demonstrated that

(1) The core of a market game is never null.
(2) While the core is usually not unique, increasing the number of agents in the game tends to eliminate some allocations from the core, so that the size of the core is smaller for larger games, as in the example.
(3) The supply-and-demand equilibrium allocation and ratio of exchange is always a member of the core.

This is a striking result. It says that, for games of exchange, the noncooperative game defined by competitive markets yields a result in the core, and that a great multilateral contract by way of the grand coalition could not improve on bilateral trade mediated through competitive markets. In the language of implementation theory, markets implement the core for this class of games. We should recall, however, the special nature of games of exchange (Chapter 3, note 11).

### 8.2.2 Telser on the Core in Games with Production

When we introduce production into the game, the case is quite different. If there are increasing returns to scale, it is quite likely that the core will be null. On the other hand, decreasing returns to scale seems prima facie to conflict with superadditivity, and decreasing returns in the neighborhood of the efficient imputation can also result in a null core. Even if the core is not null, the competitive equilibrium may not be an element of the core, and the core may require a "natural monopoly," with price discrimination or some other measure to efficiently recover overhead costs (Telser, 1978, pp. 129–31). Following Telser we will again assume transferable utility and denote candidate solutions as imputations rather than allocations.

From a mathematical point of view, it is reasonable to consider a null core as a failure for the theory of the core. A solution concept that can never be satisfied for an important class of problems seems mathematically

impotent. For Telser, however, the nullity of the core is an explanatory principle. He writes (1978, p. 65):

> These results can improve our understanding of the restrictions that are necessary for an equilibrium . . . These constraints assume a variety of shapes in the real world. The state may intervene either by outright ownership of the plants or by regulation of the activities of the single firm supplying the outputs from its plants. Sometimes the state intervenes by acting on behalf of the buyers, or the buyers may form their own coalition to act in concert . . . in the case of a natural monopoly or a natural monopsony.

A probable example of a null core is the airline industry. Telser (1997, pp. 5–7) gives an illustrative example with two airline companies and just three consumers. However, the conditions of this "toy" model are recognizable in the airline companies of the real world: a very high proportion of costs are overhead costs and profits depend on filling a high proportion of the seats. Telser demonstrates in his small-scale example that the core is null so long as competition among the airline companies is unrestricted. Another example along similar lines is the film industry. Telser describes (1997, p. 263) the arrangements by which producers controlled the showing of their films by theaters in the period 1920–40. These arrangements were considered collusion in restraint of trade and were abandoned under a federal consent decree in 1940. Telser argues (1997, p. 264) that they were optimal, however, as a means of stabilizing what would otherwise have been a game with an empty core.

Despite his passing comment that "Sometimes the state intervenes . . .," Telser is primarily interested in restrictions on competition that arise within the private sector, as the film industry example shows. Nevertheless his ideas supply a key resource for the understanding of public regulation. In general, public regulation may improve the functioning of the economy when it operates to prevent a class of transactions that, if permitted, would result in an empty core. If we adopt the stability interpretation of core theory, we would say that in such a case regulation operates to stabilize an economy that would otherwise have no stable state. Of course, this is not the only function of public regulation, which may be necessary (when there are externalities) to avoid economic states that are stable but inefficient. It is, though, a function of public regulation that is less well understood, and the theory of cooperative games in coalition function form is a key tool to understanding it.

Telser writes (1997, p. viii) "People facing empty cores try to devise suitable restrictions and rules in order to obtain efficient outcomes." In the absence of public regulation, they may not succeed. The airline example seems instructive. Telser suggests that vertical integration could resolve

the empty core in this case: if each airline were owned by a customer, the core would not be null (Telser, 1997, pp. 10–11). In a real world of airlines with more than three customers, this would mean that the airlines come to be operated by consumers' cooperatives. Of course, this has not occurred in the real world, and does not seem likely to.

Consumers' cooperatives can be successful in operating natural monopolies, as the many examples of rural electric utility and telephone service cooperatives in the United States shows. These cooperatives were established not where service was unstable but where it was not profitable enough for investor-owned companies to offer the service. It may be, though, that the investor-owned companies stayed out of the rural markets because they anticipated empty-core instability. Otherwise it is a bit difficult to explain why services that could be operated at a profit by consumer cooperatives would be refused by profit-maximizing investor-owned companies. In any case, government initiative was central in the establishment of these cooperatives.

Taking the stability interpretation of the theory of the core, an empirical prediction would be that the airline industry would have no stable configuration. That seems to agree with the facts. Bankruptcies and reorganizations of the industry seem to continue in the case of airlines. Airline bankruptcies have become so common that some are jokingly said to be "in Chapter 33" – the third time in Chapter 11! (*USA Today*, 2006). Noted investor Warren Buffet has been widely quoted (and has quoted himself) as saying, ". . . if a capitalist had been present at Kittyhawk back in the early 1900s, he should have shot Orville Wright. He would have saved his progeny money" (*The Age*, 2002).

It is sometimes suggested that what is needed in the airlines is more competition, not less. On this view, smaller, flexible, innovative new airlines operating in the high-volume markets or flying smaller aircraft will supplant the inefficient "legacy airlines" that are badly managed or fatally burdened by union contracts or both. Older people may recall an advertising jingle for Allegheny Airlines: "It takes a big airline – Allegheny!" In the 1970s, Allegheny was a smaller, flexible, innovative airline flying smaller aircraft and introducing competition in the higher-volume markets. At the time of deregulation (during Jimmy Carter's administration) it grew rapidly at the expense of then-established legacy airlines. As the jingle suggested, its earlier brand image as a regional airline was no longer optimal. In 1979 it adopted a new name: USAir (Lehman, 2007). It is, of course, the same USAir that is often cited as an example of "uncompetitive" legacy airlines.

If free competition has resulted in instability of the airline industry, as seems to be the case, it does not follow that regulation before 1977

was very successful either. Perhaps the regulation (or other public initiative) appropriate to the airline industry has to be considered an unsolved problem.

Telser's discussion suggests that natural monopoly and monopsony, and other empty-core cost conditions, are widespread in a modern economy. In such cases regulation in some sense is unavoidable; and the choice (as indeed Galbraith, 1973, argued) is between public and private regulation. Moreover, efficient private regulation in cases of natural monopoly and monopsony will usually involve price discrimination. Conversely, deregulation has often meant, not a competitive market, but increased and unregulated price discrimination. Statistical studies that show lower average prices following deregulation ignore this, and may not be representative of the prices available to smaller or spot-market traders. All in all, we are unlikely to understand regulation or deregulation without the insights of cooperative game theory.

### 8.2.3 Values, Power and Accounting

Important applications of the Shapley value include the measurement of power in committees and governments (Shapley and Shubik, 1954) and the allocation of shared costs (Shubik, 1962). Shubik's example for cost allocation supposes that two plants share a joint overhead cost. Here is an example with some similar features to illustrate Shubik's argument. West Philadelphia University is a coalition of an engineering school, E, a school of media arts, A, and a business school, B. Each requires the support of a school of arts and sciences, which (for simplicity) generates no tuition revenue. In addition there are other overhead costs such as a computer center. Any coalition, including a singleton (stand alone school of engineering, art or business) must bear 25 million of overhead costs for these purposes. Operating costs are E, 20; A, 25; B, 5. Tuition revenues are E, 40; A, 35; B, 25. The coalition function is shown as Table 8.2. Suppose, for example, we take the order E, A, B in assembling West Philadelphia University from its parts. The engineering college then must (as the first and so stand-alone unit) bear the overhead alone and generates a value of −5, which is 5 less than no university at all, so $v\{E\} - v\{\varnothing\}$ is –5. Adding an art school creates {E,A}, so $v\{E,A\} - v\{E\} = 5$. Now, add B, creating {EAB}, worth 25. Therefore $v\{E,A,B\} - v\{E,A\} = 20$. Proceeding in this way, considering all possible orders and the appropriate weights, we find that the Shapley values for this game are 10, 1.67, 11.67.

The Shapley values in this case are net of all costs – essentially the target profitabilities of the three colleges, after the overhead cost. The allocations of the shared cost will be the allocations that leave each college with its

*Table 8.2 Game 8.4: coalitions and values in a game among colleges*

| Coalitions | Values |
|---|---|
| ∅ | 0 |
| {E} | −5 |
| {A} | −15 |
| {B} | −5 |
| {A,B} | 5 |
| {E,B} | 15 |
| {E,A} | 5 |
| {E,A,B} | 25 |

*Table 8.3 Allocation of costs, revenues, and profitabilities for three colleges*

| | E | A | B |
|---|---|---|---|
| Revenue | 40.00 | 35.00 | 25.00 |
| Operating cost | 20.00 | 25.00 | 5.00 |
| Allocation | 8.33 | 8.33 | 8.33 |
| Net | 11.67 | 1.67 | 11.67 |

target profitability. Table 8.3 shows the revenues, operating costs, shared cost allocations and profitability targets (Shapley values) for the three colleges. As we see, in this case the overhead cost is allocated equally, even though the operating costs, tuition revenues, and profitability targets are quite different.

There have been a few other applications of Shapley values to cost assignment, with examples such as the divisions of the Tennessee Valley Authority and the different classes of aircraft that use an airport. This has not been adopted as general accounting practice, but it does provide (at least in principle) an objective standard for the sharing of common fixed costs.

## 8.3 SUMMARY

Much of the literature of cooperative game theory relies strongly on simplifying assumptions that originated with von Neumann and Morgenstern. There are a number of properties one might like a solution to have: needless to say, no one solution will have all of them. One of the most common

solution concepts is the core, which rests on the idea that no group can be denied the payoffs they could obtain if they formed a coalition and chose a joint strategy. The core may, however, be null or may have many potential solutions within it. Three solution concepts that are never null and are unique are the Nash bargaining solution, the Shapley value, and the nucleolus, though the bargaining solution is inapplicable to more than two agents. Applications of these models are largely in economics (and to some extent in political science) and include a theory of exchange, a theory of restrictions on competition, measurement of power and the allocation of shared costs.

## NOTES

1. This appears to be inconsistent with the previous paragraph, as Telser (1978) notes in a similar context. We might say that the valuations are subjectively consistent in that each coalition is equally (and utterly) pessimistic about the decisions of those outside the coalition. Nevertheless, this will be reconsidered in Chapter 10.
2. Some of the literature would use the term *preimputation* at this point, and consider it as an *imputation* only if every agent obtains at least what he would get as a singleton. However, that distinction will not be made here.
3. This may occur because the techniques of production involve division of labor (Smith, 1994 [1776]; Kaldor, 1934), so require a certain minimum work force to be put into effect.
4. This is demonstrated in the game theory literature by forming a set of axioms one of which is the property, and showing that these axioms are equivalent to the solution concept.
5. These are technical terms from mathematical analysis and will not be discussed in detail here.
6. This follows Forgo et al. (1999).
7. This listing follows Peleg and Sudhölter (2003).
8. This listing follows Peleg and Sudhölter (2003).
9. Extension to divisible goods, drawing in the economic concept of a preference system, would demand a bit of mathematics.
10. For this discussion, an *allocation* of available goods and services among individuals replaces an *imputation* of value to a coalition. The term "allocation" may replace "imputation" in some other applications, following the example of games of exchange.
11. In some contributions this latter assumption is relaxed.

# 9. Imperfect recall and aggregation of strategies

We recall that Kuhn (CGT, pp. 193–216) extended and refined the treatment of games in extensive form, including games of "imperfect recall," that is, games in which a player may not be aware of some of its own earlier moves. (Ch. 3). For non-cooperative analysis, Selten (CGT, pp. 312–54) argues, this multiplicity of agents should be excluded. However, any nontrivial coalition is a compound of two or more agents, so that imperfect recall arises naturally in coalitions. Nevertheless, coalitions seem never to have been discussed for games with imperfect recall. In this chapter we consider two implications of imperfect recall.

## 9.1 SUPERADDITIVITY

In their founding paper of the literature on cooperative solutions for games with given coalition structure, Aumann and Dreze (1974) question the assumption of superadditivity in games in coalition function form.[1] The argument for superadditivity is essentially that any vector of strategies available to the two coalitions separately is also available to the merged coalition, so that they can do no worse than to adopt the strategies adopted by the two coalitions separately. Let us call that argument "argument A." Aumann and Dreze (1974, p. 233) question the argument, although they concede that "superadditivity is intuitively rather compelling." Nevertheless, they go on to write ". . . 'acting together' and sharing the proceeds may change the nature of the game. For example, if two independent farmers were to merge their activities and share the proceeds, both of them might work with less care and energy; the resulting output might be less than under independent operations, in spite of a possibly more efficient division of labor"[2] (Aumann and Dreze, 1974, p. 233). Aumann and Dreze make no reference to imperfect recall (although Aumann and Maschler, 1964, questioned superadditivity and, 1972, used examples with imperfect recall) but here is an example with imperfect recall that is consistent with their point.

Suppose we have two farmers, a small farmer $s$ and a large farmer $\ell$.

The small farmer has three acres of land and the large farmer has 27. (The reader may add zeros to these numbers if he or she lives in a more developed country.) Each farmer can choose to work with great effort or with slight effort. Working with great effort increases output by 50 percent, but has a subjective cost equivalent to decreasing the farmer's produce by 10 units. Output also depends on land. Working independently, on his three acres, the small farmer works with great effort and produces 30, for a value net of effort cost of $v\{s\} = 20$. The large farmer also chooses to make a great effort on his larger landholding, and doing so can produce 150, leaving $v\{\ell\} = 140$ net of the effort cost. But this is an inefficient allocation of resources. The output of one farmer working 15 acres of land with great effort is 100. (With slight effort 15 acres will produce 65.) Thus, a reallocation of land could raise the total output of the two farmers to 200, and net of effort cost their total benefits would be 180.[3]

We suppose the two farmers form a coalition to realize that potentiality. Specifically, $\ell$ rents the land of $s$ and hires $s$, forming a farm of 30 acres on which labor and effort will be efficiently allocated, with each farmer working 15 acres and $\ell$ receiving the output but making a side payment to $s$. For an extensive-form game, a side payment is simply another strategic move at the last stage. For simplicity, we suppose that $\ell$ considers only two lump sum payments, large = 40 and small = 15. If both work with great effort, $\ell$ will find himself with 200 at the end of the year and a transfer of 40 to $s$ will leave $\ell$ with 160. Deducting 10 for the cost of $\ell$'s labor, $\ell$ has $150 > v\{\ell\} = 140$, $s$ has 40, and after we deduct 10 for his effort he is left with $30 > v\{s\} = 20$. If the coalition game is a game of perfect recall, $\ell$ will commit himself to the (contingent) pure strategy $\Sigma_1$ = "make great effort, and in case $s$ makes great effort transfer 40 to $s$, but otherwise transfer only 15 to $s$;" and the best response for $s$ is strategy $\Sigma_2$ = "transfer three acres of land to $\ell$ and make great effort on the land he allots to me to work."

But the coalition of $\ell$ and $s$ is now a compound player with two "agents," who may or may not know the strategy commitments of one another. Suppose, for example, that $\ell$ is unaware of the effort made by $s$. Then $\ell$ is unable to condition his side payment on the effort made by $s$. Strategy $\Sigma_2$ simply is not available to the coalition. Alternatively, suppose that $\ell$ can monitor the effort of $s$, but $s$ is unaware whether or not $\ell$ has committed himself to strategy $\Sigma_1$. Farmer $s$ may be concerned that $\ell$ will offer a small side payment of only 15, which is the optimal behavior strategy for $\ell$ at that point in the game (that is, payment of the small side payment is subgame perfect). Then strategy $\Sigma_2$ is not available to the coalition either. It may be that both of these imperfect recall conditions exists. If so, then the game in extensive form between $\ell$ and $s$, is shown by Figure 9.1. If they play "offer" but not "refuse" they find themselves playing a subgame equivalent to the

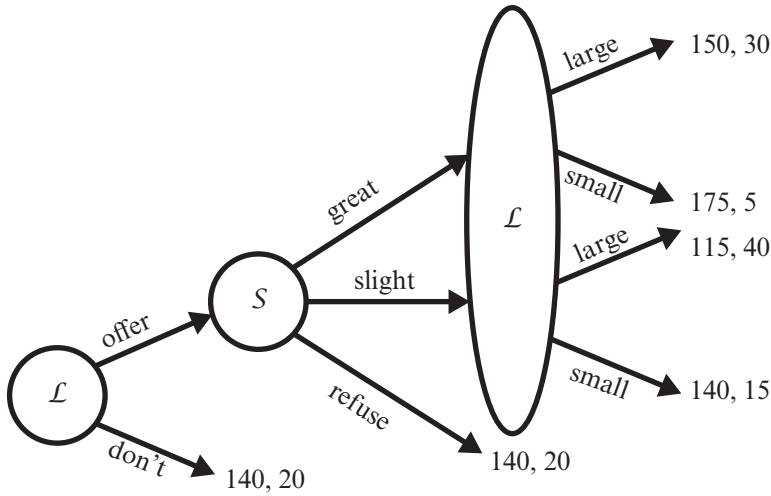


*Figure 9.1* *Game 9.1: the farmers' game in extensive form (with imperfect recall)*

*Table 9.1* *Effort and side payment*

| Payoff order: Agents $s$, $\ell$ | | $\ell$ | |
|---|---|---|---|
| | | Large payment | Small payment |
| $s$ | Great effort | 30,150 | 5,175 |
| | Slight effort | 40,115 | 15,140 |

game in normal form shown as Table 9.1, a social dilemma (as represented in behavior strategies).

How can a social dilemma arise within a cooperative game? We have seen that strategies that condition the side payment to $s$ on his effort are not available to the coalition, because of imperfect recall, so the great effort from $s$ is simply not enforceable. On the other hand, if $\ell$ cannot be confident that $s$ will make a great effort, he can only "break even" with a small payment, and a large payment leaves $\ell$ worse off than he would be in the singleton coalition. Conversely, if he cannot get the large side payment, $s$ can never be as well off than he would be in his singleton coalition, and making great effort only makes that worse. In these circumstances it does seem possible that each person would protect himself against the worst by refusing any contract on terms that would be viable. This would lead us to assign the coalition $\{s,\ell\}$ a value $v\{s, \ell\} = 155 < v\{s\} + v\{\ell\} = 160$.

Now recall argument A: "any vector of strategies available to the two coalitions separately is also available to the merged coalition, so that they can do no worse than to adopt the strategies adopted by the two coalitions separately." The two agents each chose great effort when acting as singletons, but (due to imperfect recall) no strategy is available to the coalition that would elicit great effort from *s*, indeed it is not possible for the grand coalition to adopt the strategy vector the two singleton coalitions adopted separately. Now, a defender of argument A might respond that this mis-states the point: that the argument, as applied in this case, means that *s* and *ℓ* could instead simply continue each to work his own land and take possession of his own product, with the result that the grand coalition would produce just the value that the two produced separately. This cannot be controverted, but what does it mean? I would suggest that we might say instead, "because of imperfect recall the game is not superadditive, so the grand coalition does not form in this game." What is the difference of meaning between the two phrases?

Another defense of argument A might be: "But this example depends on the assumption that *ℓ* rents land from *s* and hires him. There are other ways the coalition might be formed. For example, *ℓ* might rent 12 acres to *s*, either for a fixed sum or a share of his product, and each take possession of his own product." The counterargument to this defense is that the defense is simply describing a different game, and of course a different game may have a different solution.[4] But this defense actually points up the pragmatic advantages of a formulation that allows us to say that the game is not superadditive. We might want to explain the widespread practice of sharecropping along the following lines: "Family farms and sharecropping (and other forms of renting of agricultural land) are widespread because larger farms are less efficient: that is, the game of consolidating farms is a game with imperfect recall, and consequently is not superadditive." But if, accepting argument A, we assume superadditivity, this explanation is not possible.

## 9.2 AGGREGATION OF STRATEGIES

As we have seen, the "plausible" argument for superadditivity draws on an assumption that the set of strategies available to a coalition is comprised of the strategies available to the individual members of the coalition. For a coalition C with $r>1$ members, a coalitional strategy would then be a vector of strategies available to the individual members. The coalitional strategy vector would have dimension $r$. Drawing on Section 2.3 in Chapter 2, let $\sigma_i$, denote an element of $S_i$ the set of strategies available to

individual *i*. Then a strategy for coalition C is a set $\{\sigma_i\}$ with one element for every *i* who is a member of C. The set of all strategies available to coalition C, $\sigma_C$, then is the set of all[5] such $\{\sigma_i\}$. *This is an assumption* that will be called the assumption of aggregation of strategies in what follows; and it is an assumption we will now wish to question. In imperfect recall, we have one reason to question it. A second reason is division of labor, as Aumann and Dreze note in passing. An illustrative example will be given in Chapter 14, Section 14.6.

Division of labor presupposes that different agents perform different tasks. In general a mode of production with M tasks will be available only to a coalition including at least M agents. If, as Smith argued, more complex division of labor leads to higher labor productivity, then there will be increasing returns to employment scale, as Kaldor (1934) pointed out. (This is a "long run" argument – that is, it assumes that nonhuman inputs are optimal for every employment scale, so that there are increasing returns to the labor input *mutatis mutandis*, not *ceteris paribus*.) Other kinds of indivisibilities may also lead to increasing returns to scale so that, as a practical matter, some strategies will only be available to coalitions above a certain size. With perfect recall, this would increase the tendency toward superadditivity, even though it is not consistent with aggregation of strategies. If both division of labor and "imperfect recall" of effort commitments can occur, they might offset one another in various ways, including (among many others) the u-shaped long-run average cost curve of Marshall.

Thus, in general, it will be necessary to specify both the game in strategic (or extensive) form and the set $\sigma_C$ for each coalition. For some games the assumption of aggregation of strategies will be acceptable: these will be called *aggregative games*. Other games will be called *non-aggregative games*. For games such as the one in the previous section, $\sigma_C$ will be a subset of the set of all vectors of strategies available to individuals. However, economies of scale and division of labor may make strategies available to a coalition that are not available to any set of individuals as singleton coalitions.

## 9.3 EXCHANGE GAMES AND IMPERFECT RECALL

The study of exchange games is one of the great successes of cooperative game theory. In part, this reflects the fact that exchange games are assumed to be free of externalities. Nevertheless, monopoly remains mysterious in the context of exchange games (Aumann, 1973). The conventional monopoly result from the economics textbook, which is inefficient, cannot be consistent with any cooperative solution. Stuart (2001) resolves

this problem by imposing the "law of one price," but this assumption is ad hoc, and it is not clear that the law of one price is consistent with cooperative game theory.

The literature of exchange games, like almost all traditional cooperative game theory, assumes perfect recall, although the assumption is never explicitly stated. In this literature, the value of a coalition corresponds to the potential gains from trade among the members of a coalition. Since exchange games are usually modeled as NTU games, even this coalitional value cannot be expressed as a scalar; but in any case no explicit account is likely to be given of the choice of strategies in an underlying game. We will find that if recall is imperfect, the value of a coalition may be less than the potential gains from trade among its members. This section will explore the point by means of a small-scale example.

To assess the implications of imperfect recall for exchange games it will be necessary to be more explicit about strategies in an exchange game. Presumably the behavior strategies in an exchange game are transfers of particular quantities of goods or of money. Certainly each person's agreement to an exchange is an agreement to make a *conditional* transfer and hence is, in a broad sense, a contingent strategy. In order to model exchange as a game in extensive form, we might think of it as a two-stage game. At the first stage each party to the exchange makes a *revocable commitment* to transfer certain amounts of goods and/or money to certain other parties on the condition of certain transfers from them. At the second stage, if the conditions are consistent they are carried out, but if the condition has not been met, the offer is revoked and no exchange takes place. The contingent strategy then is of the form "Commit to a transfer of $X$ to $i$ and if $i$ does not commit to a transfer of at least $Y$ to me, then revoke; otherwise do not revoke." This will be called the game of transfers.

With that background consider the following four-person game. The four agents are of four different types and are denoted $a$, $b$, $c$, and $m$. Player $m$ is initially endowed with four widgets and his payoff depends on the number of widgets and the quantity of money he has at the end of the game. The payoff is

$$y_m = \begin{cases} 225 \text{ if no widgets are sold} \\ \text{money} + 50z \text{ where } z \text{ widgets are not sold} \end{cases}$$

Agents $a$, $b$, and $c$ are initially endowed each with 100 units of money and their payoffs are

$$y_a = \begin{cases} \text{money} + 100 \text{ if one widget} \\ \text{money if no widgets} \end{cases}$$

$$y_b = \begin{cases} \text{money} + 70 \text{ if one widget} \\ \text{money if no widgets} \end{cases}$$

$$y_c = \begin{cases} \text{money} + 40 \text{ if one widget} \\ \text{money if no widgets} \end{cases}$$

It will be apparent that agent $a$ will be better off to acquire a widget at a price of less than 100 and $b$ at a price of less than 70; however $c$ is not benefited by a widget at any price above 40, while $m$'s reservation price of a widget is 50. Note also that if agent $m$ were to sell four widgets at the reservation price of 50 each his payoff would be 25 less than it is if there are no sales at all. This 25 is an overhead cost of entering into the market.

Now suppose $m$, $a$, and $b$ were to form a coalition and commit themselves to the following strategies:

1.m $m$: Transfer 1 widget to $a$ provided $a$ transfers 60 monetary units to me AND transfer 1 widget to $b$ provided $b$ transfers 60 monetary units to me.
1.a $a$ and $b$ : Transfer 60 monetary units to $m$ provided he transfers 1 widget to me.

On this basis the payoffs would be 220 for $m$, 140 for $a$, 110 for $b$, and 100 for $c$, for a total of 570, an efficient outcome. However, with a payoff of 220, $m$ is worse off than he would have been with no trade, and so will refuse the coalition.

Suppose instead that the strategies were:

2.m $m$: Transfer 1 widget to $a$ provided $a$ transfers 82.5 monetary units to me AND transfer 1 widget to $b$ provided $b$ transfers 60 monetary units to me.
2.a $a$: Transfer 82.5 monetary units to $m$ provided he transfers 1 widget to me.
2.b $b$: Transfer 60 monetary units to $m$ provided he transfers 1 widget to me.

This yields the same efficient 570 of final payoffs but $m$, $a$, and $b$ are each better off with the trade than without, so the coalition might be formed on this basis. From the point of view of the economist, $m$ , a monopolist with a fixed cost, has increased his total revenue (in order to cover the fixed cost) by means of price discrimination.

However, strategies 2.a and 2.b are dominated by:

3.a $a$: Transfer 80 monetary units to $b$ provided he transfers 1 widget to me.

3.b $b$: Transfer 60 monetary units to $m$ provided he transfers 1 widget to me AND transfer 1 widget to $a$ provided $a$ transfers 80 monetary units to me.

Thus, $m$ sells only 1 and the payoffs are 210 for $m$, 120 for $a$, 120 for $b$, and 100 for $c$, for a total of 550. This is an inefficient outcome, but $a$ and $b$ are both better off than they were in the previous case.

The monopolist faces competition from his own customers when he practices price discrimination in this example. He might forestall this by making his strategy:

4.m $m$: Transfer 1 widget to $b$ if and only if BOTH [$b$ transfers 60 monetary units to me AND $b$ makes no transfers of the widget to other players] AND transfer 1 widget to $a$ provided $a$ transfers 82.5 monetary units to me.

An alternative in this small game would be to make both transfers of widgets contingent on the money transfers from both players, but in a larger game, in which the monopolist might not be aware of the types (different demands) of the various players, this might not be feasible. On the other hand we see many real-world examples of monopolists attempting to prevent the resale of their products, such as sale contracts that make air tickets non-transferable, anti-scalping laws, and frequent revision of textbooks.

If the game of transfers is a perfect-recall game, then commitment 4.m, with 2.a and 2.b, will result in the same efficient and mutually beneficial outcome as 2.m, 2.a, and 2.b as noted above. Suppose, however, that the game is of imperfect recall so that $m$ cannot verify that $a$ and $b$ are committed to 2.a and 2.b rather than 3.a and 3.b. Then strategies 4.m are not available to a coalition of exchange. In such a case we may observe:

5.m $m$ : Transfer 1 widget to $a$ provided $a$ transfers 87.5 monetary units to me.

5.a $a$: Transfer 87.5 monetary units to $m$ provided he transfers 1 widget to me.

In this case $m$ makes no offer to $b$ because any price acceptable to $b$ would, via secondary transfer from $b$ to $a$, make it impossible for $m$ to obtain an overall revenue that would cover his overhead as well as marginal cost. This results in final payoffs of 237.5 for $m$, 112.5 for $a$ , 100 for $b$ , and $c$,

for a total of 550. This is again an inefficient outcome, but $a$ and $m$ both find it preferable to no trade, and the others lose nothing by it.

This is a case of monopoly restriction of output such as may be found in any principles of economics textbook. The monopolist has restricted sales in order to charge a high price, and the inefficient payoff of 500 rather than 520 reflects a "deadweight loss" or "welfare triangle" equivalent to 20 monetary units. This is a consequence of "imperfect recall" within the cooperative coalition for exchange.

Suppose that we double the game, so that there are two identical agents of each kind, and again assume perfect recall. The value of a coalition of all active agents, $\{m_1, m_2, a_1, a_2, \bar{b}_1, \bar{b}_2\}$ will be 965. This is more than double the value of $\{m, a, \bar{b}\}$ in the previous game because of economies of scale: one of the two producers can supply all four of the active buyers, and the other remains inactive, contributing a value of 225 to the coalition. As before, agents of type *c* are dummies, contributing their singleton value of 100 to whatever coalitions they join. Now consider a payoff schedule with $y_{m_2} > 225$ and suppose a coalition $\{m_1, a_1, a_2, \bar{b}_1, \bar{b}_2\}$ – is formed, that is, $m_2$ is expelled from the grand coalition. The new coalition will increase its value by $y_{m_2} - 225 > 0$, so that the payoff schedule is not in the core. For a core allocation, each producer must obtain exactly its alternative cost, 225. By introducing a second producer, and thus a possibility of what economic theory would call market entry, we have introduced a "contestability" constraint (Baumol et al., 1982). An "established monopolist" (the active producer) faces the risk of entry that could precipitate losses or force him out of the market. To forestall this, the established monopolist must moderate his prices to the point that the potential entrant cannot increase his profits by entry. In this symmetrical case, that means the monopolist's own economic profits are reduced to zero. The result is efficient (because of price discrimination) and the buyers benefit from the economies of scale.

Suppose instead that recall is imperfect in the doubled game. On the one hand, by the same reasoning as before, the payoff to each producer must be exactly its opportunity cost, 225. On the other hand, this game is large enough that a producer can profitably sell to agents of type $\bar{b}$ despite the law of one price. A price consistent with these two constraints is 56.25. There is no monopoly inefficiency, although this is a result of the simplicity of the model, and with additional buyer types – such as a type *d* who can benefit by buying only at a price of 55 or less – the usual monopoly inefficiency will be restored, on the assumption of imperfect recall.

The conclusions are that, on the one hand, the "law of one price" and monopoly as it is customarily treated in economics are phenomena of "imperfect recall," so we should not expect any counterpart to them in conventional cooperative game theory in which perfect recall is assumed,

and (on the other hand) the results of exchange games may be sensitive to deviations from the assumption of perfect recall. We may modify other assumptions, such as introduction of a second potential producer and consequent contestability constraints, and this will change the bargaining power among the agents in the game; but monopoly inefficiency and the law of one price are to be expected only in case imperfect recall is assumed.

## 9.4 SUMMARY

Kuhn's games with imperfect recall provide a rationale, within game theory, for what pragmatically appear to be non-superadditive cooperative games. Plausible as the argument for superadditivity is, cooperative game theory seems potentially a more powerful explanatory tool if the assumption of superadditivity is not made. Similarly, it seems that the theory of exchange games is not the settled field that it might have seemed to be for the last forty years. While research on more general models of exchange with imperfect recall could be rewarding, further discussion will have to be beyond the scope of the book.

## NOTES

1. A search of Science Citation Index and Social Science Citation Index for papers that cite both the papers of Kuhn and Aumann and Dreze returned no references. I am indebted to my wife, Katherine McCain, Professor in the College of Information Science and Technology, Drexel University, for this information. Aumann and Sorin (1989) is not an exception and argues that a repeated non-cooperative game with some limitation on recall may for that reason be more likely to realize a cooperative outcome. However, it is not concerned with coalitions, nor is it concerned strictly with the condition Kuhn defined as imperfect recall.
2. This concern with shirking in organizations was central to some contributions to economics at about the same time; see, for example, Alchian and Demsetz (1972).
3. The numbers in this example are derived from a spreadsheet example in which the dependence of output on the land input and effort is Cobb-Douglas, although effort is considered (for simplicity) as indivisible. There are many models in the literature that allow effort to vary continuously (for example, McCain 1980; 2007a) but this complicates the mathematics somewhat without making any qualitative difference in the results.
4. There is, of course, a large literature on sharecropping and its advantages and disadvantages by comparison with renting agricultural land for a fixed sum. See, for example, Stiglitz (1974), Cheung (1968).
5. Formally $\Sigma_C = \{\{\sigma_i | i \in C\} | \sigma_i \in S_i\}$.

# 10. Strategy, externality, and rationality

It has been observed that much literature in game theory relies on simplifying assumptions that can frustrate the application of the theory, particularly to public policy. "Perfect recall" is one instance here. The objective of this chapter is to give arguments why several other assumptions are problematic. We will begin with a common (often tacit) assumption of non-cooperative game theory and then proceed to explore two further issues of cooperative game theory and an ambiguity in the concept of rationality.

## 10.1 "BEHAVIOR STRATEGIES SUFFICE"

We now have the technical apparatus to reconsider the role of contingent and behavior strategies in game theory, and the idea that, thanks to Kuhn's demonstration, "behavior strategies suffice." As we recall from Chapter 3, Kuhn had demonstrated that an important family of games in extensive form can be analyzed by using behavior strategies only, choosing local (generally randomized) best responses. It was noted, however, that this analysis is not applicable to games of imperfect recall (CGT, 1997, pp. 146–68), nor to any cooperative game (Selten, 1964), nor does it recover all Nash equilibria (CGT, 1997, pp. 312–54). It was also stated in Chapter 3 that Kuhn's reasoning does not apply to non-cooperative equilibrium concepts other than the Nash equilibrium. This will now be discussed.

In particular, correlated strategies cannot be derived from the local determination of behavior strategies as best responses at each information set. Consider Game 10.1, shown in extensive form by Figure 10.1. No "story" or application will be given for this game, which is offered strictly to illustrate the relation between contingent and behavior strategies. The agents are $a$ and $b$ and the game proceeds in just two stages. First, $a$ chooses between behavior strategies $u$, $c$, and $d$, and then (depending on $a$'s play at the first stage) $b$ chooses between $t1$ and $b1$ (at information set $b1$) or $t2$ and $b2$ (at information set $b2$).

As usual, agent $a$'s contingent strategies need not be distinguished from

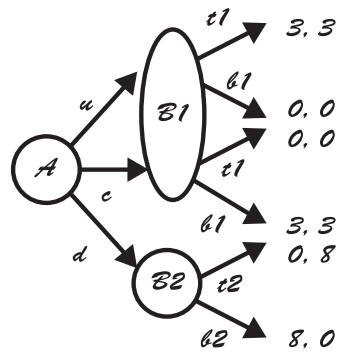


*Figure 10.1 Game 10.1 in extensive form*

*Table 10.1 Game 10.1, in strategic normal form*

| Payoff order: A,B | | B | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| A | *u* | 3,3 | 0,0 | 3,3 | 0,0 |
| | *c* | 0,0 | 3,3 | 0,0 | 3,3 |
| | *d* | 0,8 | 0,8 | 8,0 | 8,0 |

his behavior strategies, since he makes the first play. Agent *b* has four contingent strategies:

1. If *u* or *c* then *t*1, else *t*2
2. If *u* or *c* then *b*1, else *t*2
3. If *u* or *c* then *t*1, else *b*2
4. If *u* or *c* then *b*1, else *b*2

Notice that this list is highly redundant, as always when behavior strategies are translated to contingent strategies. The reason for this redundancy is that contingent strategies 1 and 3 differ only with respect to play at information set *b*2, which is never reached in Nash-equilibrial play, and similarly strategies 2 and 4. Notice also that any cooperative solution to this game will correspond to a strategy of *d* by player 1, followed by any behavior strategy of agent *b* and an offsetting side payment. But this can never be realized if behavior strategies are chosen as local best responses.

Game 10.1 in strategic normal form is given by Table 10.1. This game has a number of Nash equilibria, due to the redundancy that has been mentioned, but they fall into two categories: pure strategy equilibria

yielding 3,3 and randomized strategies yielding expected values of 1.5, 1.5. The pure strategy equilibria provide relatively good outcomes but, as usual, they raise questions since they seem to require consultation or information that the agents are assumed (in Figure 10.1) not to have. However, this game has a simple correlated equilibrium solution. Let the two agents flip a coin and play according to the rules:

For agent *a*: "If *H* then *u* else *c*."
For agent *b*: "If *H* then 1 else 2," or "if *H* and (*u* or *c*) then *t*1 else *t*2 else if *T* and (*u* or *c*) then *b*1 else *t*2."

What a correlated strategy does, in effect, is to imbed the game in a larger game in which the first step is the signal, in this case flipping the coin. Neither agent has any reason to deviate from play by these rules: in *a*'s case, a deviation to *d* or to play the "wrong" strategy from *u* and *c* will leave him with nothing, and in *b*'s case, a deviation to play *b*1 on *H* or *t*1 on *T*, a "wrong" behavior strategy, will similarly leave him with nothing. However, whether the original game is expressed in contingent or behavior strategies, the strategies of play in the larger game must themselves be contingent strategies. To be specific, unless agent *b* knows that agent *a* will play according to the contingent strategy "if *H* then *u* else *c*," agent *b* has no information that would allow him to choose a behavior strategy that would produce a pure strategy equilibrium. This is the logical issue in coordination games, and local choice of behavior strategy offers no escape from it. Local choice of best-response behavior strategies cannot produce a correlated equilibrium.

## 10.2 EXTERNALITY

Shapley and Shubik (1969) extended Shubik's model of the core of an exchange game to the case of externalities. It will be worthwhile to explore that model, since it has been asserted that externalities call for treatment in terms of a theory of games in partition function form (or perhaps some non-transferable utility extension of it) but Shapley and Shubik (1969) do not do so. Among their conclusions were that the competitive equilibrium would be in the core if externalities were positive, though a system of taxes and subsidies could be constructed that would also be in the core; but that negative externalities could result in an empty core.

Consider, again, Game 2.5, the game of production of a public good. This game has also been investigated in Chapter 8, 1; and the payoffs are as shown in Table 10.2. Table 10.2 corresponds to game 2.5, except that, following Shapley and Shubik, we treat the payoffs as nontransferable and

*Table 10.2 Game 2.5 revised: a game of production of a public good*

| Payoffs: A,B,C | | C | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Produce B | | | | | | Don't B | | | | | |
| | | Produce | | | Don't | | | Produce | | | Don't | | |
| A | Produce | 8 | 8 | 8 | 6 | 9 | 6 | 6 | 6 | 9 | 4 | 7 | 7 |
| | Don't | 9 | 6 | 6 | 7 | 7 | 4 | 7 | 4 | 7 | 5 | 5 | 5 |

noncomparable. (This will not change the results in any qualitative way but is considered more sound in terms of neoclassical economic thinking on utility theory and certainly is more general.)

Coalition values will not be meaningful since the payoffs are not transferable, but we may ask what payoffs an agent can expect if a particular coalition is formed. Accordingly, consider the payoff to the singleton coalition $\{c\}$. If $\{a, b\}$ choose "produce, produce" and produce two units of the public good they will increase their own payoffs from 5,5 to 6,6. However, in so doing, they would increase the payoff of the singleton coalition $\{c\}$ from 5 to 9. This increase in the payoff to $\{c\}$ is a positive externality to $\{c\}$. Therefore, according to the assurance principle, $\{a,b\}$ will not do it. Instead, they will direct to $c$ an ultimatum: join in and share the cost of the public good or no public good will be produced. Therefore, the value of the singleton coalition $\{c\}$ is not 7 but 5 – the smallest that it is in the power of $\{ab\}$ to make it. It follows that 8, which $c$ obtains from the grand coalition, is an improvement for $\{c\}$; the grand coalition forms and the public good is produced. In this case the major cooperative solutions coincide.

What about the payoffs to a two-person coalition such as $\{a, b\}$? If they choose "produce, produce," then they can assure themselves of 6,6; while if they do not they will receive 5,5; and there is nothing the singleton coalition $\{a\}$ can do to reduce them below 6,6. Therefore (according to the assurance principle) they will produce the public good and gain payoffs of 6,6.[1]

Here is an alternative scenario. Knowing that he will face the unified opposition of $\{a, b\}$, $\{c\}$ sees his opportunity simply to enjoy the public good for free, gaining a total payoff of 9. He does not find the threat by $\{a, b\}$ to be credible, since it is not a Nash equilibrium. So $c$ becomes a "holdout." This is not as asymmetrical as it may seem. The point is that any of the three might find himself as a holdout: the possibility is symmetrical although the realization is unsymmetrical. This means that the grand coalition payout of 8 will not be sufficient to persuade anyone to remain

*Table 10.3 Game 10.2: a game of pollution*

| Payoffs: A,B,C | | C | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Dump on A B | | | | | | Dump on B B | | | | | |
| | | Dump on A | | | Dump on C | | | Dump on A | | | Dump on C | | |
| A | On B | 5 | 9 | 10 | 9 | 9 | 9 | 9 | 5 | 10 | 10 | 5 | 9 |
| | On C | 5 | 10 | 9 | 9 | 10 | 5 | 9 | 9 | 9 | 10 | 9 | 5 |

in the grand coalition and the grand coalition will not be stable. By adopting the assurance principle, Shapley and Shubik have simply assumed the possibility of such holdout behavior away. They show that, on that basis, positive externalities do not prevent the formation of an efficient grand coalition. But if (as common sense suggests) holdouts are no less common forms of human action than threats, positive externalities may prevent the formation of an efficient grand coalition.

Now we consider a case with negative externalities. Each of the three agents begins the game with wealth of 10 and a bucket of garbage that he may dump in the backyard of either of his neighbors. (For technical reasons we can ignore the possibility of dumping the garbage in his own backyard.) If he has one bucket of garbage dumped in his backyard, the victim loses well-being equivalent to one unit of wealth, reducing his payoff to 9. However, the impact of garbage pollution is nonlinear: if he has two buckets dumped in his backyard, the victim's well-being is decreased by 5, leaving a payoff of 5. (This reflects an idea that even if pollution cannot be reduced, as it cannot in this game, its worst effects may be prevented by dispersing it as widely as possible.) The payoff table for Game 10.2 is Table 10.3.

Once again, what payoffs can be expected by a player who joins a particular coalition? First, a singleton coalition will face a unified opposition that will conspire to reduce the singleton's payoff to 5. Second, a two-person coalition can agree to dump on the nonmember, and depending on the nonmember's decision, each of them will do no worse than 9. Therefore, 9,9 is the security outcome for a two-person coalition. The grand coalition can realize 9, 9, 9 by adopting a cycle of dumping, such as $a$ on $b$, $b$ on $c$ and $c$ on $a$. This avoids the nonlinear impact of two buckets dumped in one backyard.

However, this is not a stable arrangement. For example, $\{a,b\}$, with an agreement that both will dump on $c$, assures each of a payoff of at least 9 and a chance at 10. But this in turn is not stable either. Suppose the payoff

to $\{a,b\}$ is 9, 10. Then $\{a,c\}$, with the same agreement, makes $c$ better off with an assured payoff no less than 9 and $a$ no worse off and standing a chance of getting the better payoff of 10. (Conversely if the payoff to $\{a,b\}$ is 10, 9, $c$ would approach $b$ for the realignment.) This sort of reasoning supports the conclusion of Shubik and Shapley that the core of the game with negative externalities may be null. Notice that we have a dominance cycle, as $\{a,b\}$ is dominated by $\{a,c\}$ as we have seen, but again, $\{a,c\}$ will be dominated in the same way by either $\{a,b\}$ or $\{b,c\}$, and so on. However, notice that after the second realignment $b$ (or $a$ if he were the original gainer) is worse off. He might anticipate that, and be cautious enough to decline the opportunity to join with $a$ in "ganging up" on $c$. Thus, a more "farsighted" solution concept than the core might point to a different conclusion.

The difficulty in the case of positive externalities derives in part from the assurance principle, but in part from the representation of the game in coalition function form itself. When there are externalities, including what Scitovsky (1954) called "pecuniary externalities," the value of one coalition depends on the other coalitions that form, as in these cases. This suggests the use of the partition function form as the alternative that could allow for such things as holdout behavior. The partition function assigns a value to each coalition imbedded in a particular partition. How are these values to be determined?

## 10.3 THE DERIVATION OF COALITION VALUES FROM THE UNDERLYING GAME IN STRATEGIC NORMAL FORM

Von Neumann and Morgenstern conjectured that (as against their assurance principle) the opposition to a coalition might be motivated more by the prospect of increasing their own gain than by that of minimizing the payoff of the opponent. That might lead the game theorist to assign those values instead by assuming Nash equilibrium play between the opposed coalitions. The Nash equilibrium was not yet known, so von Neumann and Morgenstern had no such alternative available. But subsequent work has not usually considered this option either.

Telser may be unique in considering the Nash equilibrium as an alternative to the security value in determining the values of coalitions (1978, p. 12). He does note the inconsistency of the assumptions determining the values of a coalition and its residual (pp. 13, 22–3; and see note 1) but asserts that consistency is not needed, since the value only represents a (maximal) feasible threat, and the results of threats by opposite sides may

indeed not be consistent. As against the Nash equilibrium, he writes (p. 13) ". . . if there are $N > 2$ firms with the freedom to form any coalition they please, then the Cournot-Nash point becomes ambiguous. It depends on how the nonmembers of S sort themselves into a set of coalitions forming a partition of $\overline{S}$." He concludes (p. 14) that the security value is the best expression of the value of a coalition. "Admittedly there are objections, but these are less weighty than the objections to the alternatives." However, the one objection he gives to the Nash value alternative is not applicable once we have adopted the partition function as the representation of a cooperative game. Each way in which "the nonmembers of S sort themselves into a set of coalitions forming a partition of $\overline{S}$" corresponds to a different imbedded coalition; given the partition there is no ambiguity about the Nash value.[2] Indeed, the dependency of the coalition value on the partition of the agents not in the coalition is exactly the structure the partition function is meant to express. Conversely, one of the advantages of taking the partition function as our expression of a cooperative game is that we can rely on the Nash equilibrium to determine the values of imbedded coalitions and thus base those values on consistent decisions.

In the remaining chapters of the book the discussion of coalitions will be based on the matched assumptions of the representation of the game in partition function form and the determination of imbedded coalition values by Nash or correlated equilibrium play among the coalitions in a particular partition or coalition structure.

## 10.4 RATIONALITY

The contrast has been noted again and again between contingent strategies underlying cooperative game theory and the behavior strategy often applied in non-cooperative game theory; and so also has the contrast between cooperative and non-cooperative game theory in general. In this section the argument will be made that these differences are at base different conceptions of rationality. The issue is *not* whether people are rational, irrational or partly irrational. Rather, the issue is what it means to say that people are rational, either wholly or partly.

### 10.4.1 Weakness of Will and Rationality

Non-cooperative game theory has been greatly influenced by Selten, and his 1975 paper is a key paper for our purposes here. It is interesting to contrast the assumptions of this paper with those of Selten (1964). There, we recall, Selten had acknowledged in an afterword that his (cooperative) model in

the 1964 paper required the assumption that players could commit themselves to any contingent pure strategy, acknowledging Thomas Schelling (1960) as the source of his change of mind. In 1975 Selten's assumptions are the reverse of those he made in 1964, but his terminology is also inconsistent in a way that may obscure the difference. Selten defines a pure strategy not as a plan assigning probability 1 to one of all possible contingency plans, as von Neumann and Morgenstern did and as Selten did in 1964, but as the assignment of probability 1 to a particular behavior strategy choice at a particular information set. Selten now assumes what Schelling (1980) called weakness of will.

But this will make no difference for rational behavior as Selten now conceives it. Selten limits his subject matter to games with perfect recall. He writes (CGT, 1997, p. 320) "Since game theory is concerned with the behavior of absolutely rational decision makers whose capabilities of reasoning and memorizing are unlimited, a game, where the players are individuals rather than teams, must have perfect recall." He then excludes consideration of teams, and he justifies this by limiting his scope to "strictly non-cooperative games." This means Kuhn's multiple-agent games are excluded. For multiple agents to be joined together as a single player would require a cooperative agreement among them, and this is excluded by assumption. (In particular an example as in Chapter 9, Section 9.1, above is excluded). But this, taken with (Selten, 1975, p. 328) "Player 2's choices should not be guided by his payoff expectations in the whole game but by his conditional payoff expectations," tells us that for rational behavior as Selten now conceives it, there can be no commitment whatever. By assumption, only behavior strategies are relevant to rational behavior.

This concept of rationality has become predominant in economics as well as non-cooperative game theory, and it is appropriate now to expand on the different concepts of rationality in those fields and in most cooperative game theory.

In economics, the issue of weakness of will and commitment arises in the context of intertemporal inconsistency of rational choice. We adopt the neoclassical convention of expressing time preference by a discount rate. Most economic literature assumes that this discount rate per unit time is the same regardless of the delay before the payment is made. This assumption of a uniform rate of time preference has no basis in empirical observation, but is made in order to reconcile the theory of rational choice, as it is understood in modern economics, with the assumption of time preference. The difficulty is that a nonconstant rate of discount can result in what are called intertemporal inconsistencies in decision-making. What this means is that a rational, maximizing decision-maker would make one decision at one point of time, but at a later point of time would rationally prefer

the alternative he has initially, rationally rejected. (There has been some recent research on alternatives to constant rates of time preference, such as hyperbolic discounting, but it has been directed to a different issue.)

### 10.4.2 Intertemporal Inconsistency

Let us illustrate this point by an example. Suppose that the decision-maker discounts any prospect delayed by more than six months at 18 percent, but that his rate of discount for prospects delayed six months or less is zero. Now the decision-maker must choose at $t_0$ between two alternatives. Alternative A1 is a payment of \$5000 at $t_0 + 1$ year. Alternative A2 is a payment of \$10 000 at $t_0 + 5$ years, but A2 has a cancellation clause: at any time during the first year, for a cancellation fee of \$100, the decision-maker can cancel his decision for A2 and receive the payment of \$5000 at $t_0 + 1$ year.

At $t_0$, the discounted present values are:

Alternative A1 \$4237
Alternative A2 \$4371

Accordingly, the decision-maker chooses alternative A2. However, at $t_1 = t_0 + 6$ months and one day, the payoff for alternative A1 is less than six months away, and so is not discounted, and is valued at \$5000. To obtain this payment, however, the decision maker must pay the cancellation fee of \$100. The net values discounted to $t_1$ are:

Alternative A1 \$4900
Alternative A2 \$4748

Therefore, the rational decision-maker reverses his decision.

This is a one-person game. Suppose we express these decisions as plans of action for the successive stages, like the pure strategies as understood by von Neumann and Morgenstern. The decision-maker has three pure strategies:

(1) Choose A1
(2) Choose A2, then do not cancel
(3) Choose A2, then cancel

The payoffs of these strategies, discounted to $t_0$, are:

(1) \$4237
(2) \$4371
(3) \$4127

Why, then, does our rational decision-maker not simply choose strategy 2 and stick with it? Suppose that the decision-maker has a weak will, in Schelling's sense, and knows that he does. Then he can anticipate that if he chooses A2, he will indeed cancel it after six months and in fact carry out strategy 3. Because of his weakness of will, strategy 2 simply is not available to him. That being so, in the spirit of Ulysses and the Sirens, (note Elster, 1977) the rational but weak-willed decision-maker will choose strategy 1 and alternative A1.

This is not to say that intertemporal inconsistency does not exist. No doubt a strong-willed decision-maker, having chosen strategy 2, will feel some subjective tension in the nature of regret or temptation during the time interval $t_1$ to $t_2 = t_0 +$ one year. Does rationality require him to act on the temptation? Well – perhaps it does.

### 10.4.3 Weakness of Will in a Game in Extensive Form

Weakness of will may also be a factor in interactive decisions. Consider Game 10.3 in extensive form, shown as Figure 10.2. All decisions are close enough together in time that there is no need to discount payments to present value.

First we note that the perfect equilibrium for this game is for decision-maker $a$ to choose alternative A1 for a payoff of \$4237. However, when

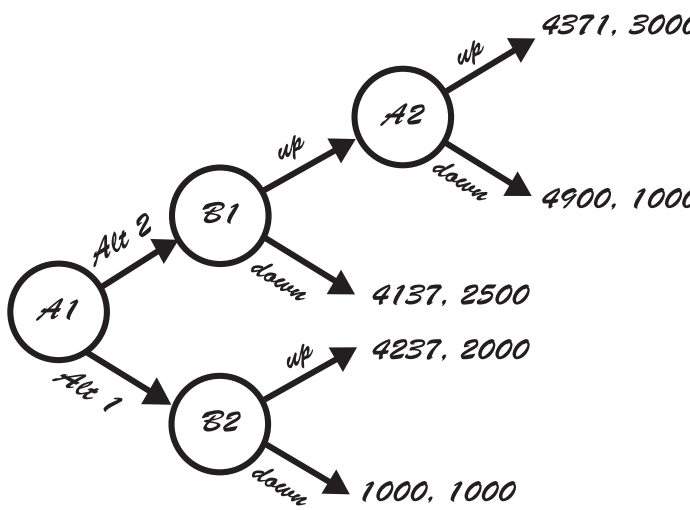


*Figure 10.2 Game 10.3: two-person game*

we express this game in terms of contingent strategies, we have, for decision-maker *a*:

(1′) Choose A1
(2′) Choose A2, then, if *b* chooses up, choose up
(3′) Choose A2, then, if *b* chooses up, choose down

and for decision-maker *b*:

(4′) If *a* chooses A2 then choose up
(5′) If *a* chooses A2 then choose down

If decision-maker *a* chooses strategy 3′, then decision-maker *b*'s best response is strategy 5′, while if decision-maker *a* chooses strategy 2′, *and b knows this with certainty,* then *b*'s best response is strategy 4′. Taking this into account, the payoffs to be expected from these strategies would seem to be:

(1′) $4237
(2′) $4371
(3′) $4137
(4′) $3000
(5′) $2000

This being so, we ask again, why does decision-maker *a* not simply choose strategy 2′? There are two possibilities: (i) Decision-maker *b* believes that decision-maker *a* has a weak will, and will not carry out strategy 2 but, having arrived at decision point A2, will choose down. Decision-maker *b* therefore chooses strategy 5′; and this is known to agent *a*, who then chooses strategy 1′ as his best response to strategy 5′. Thus, it seems, the subgame perfect equilibrium can be necessary because of the belief that *a* has a weak will. (ii) The second possibility is that *b* believes *a* is dishonest and opportunistic and will choose "down" at decision point A2 regardless of any protestations to the contrary. Thus, the subgame perfect equilibrium can be necessary because of the belief that A is dishonest. But suppose that decision-maker A has a strong will, that is, a capability to choose strategy 2′ and stick to it despite the temptation to choose "down" at decision point A2; and suppose that this is known to decision-maker *b*. Suppose decision-maker *a* also is honest, and this, too, is known to decision-maker *b*. Thus, decision-maker *a* needs only announce "on my honor, I am choosing strategy 2′," and then *b*'s rational decision is for strategy 4′, and the cooperative solution A1, up, up, results.

### 10.4.4 Perfect and Ideal Rationality

Notice that, so far as *a*'s decisions are concerned, decision sequence 1′, 2′, 3′, with its payoffs, is identical to 1, 2, 3 in the intertemporal inconsistency example. The major difference is that it is decision-maker *b*'s belief that *a* has a weak will, rather than *a*'s belief that *a*'s will is weak, that puts A it 2, up, up, out of *a*'s reach. Supposing *b* to be rational, what basis might he have for that belief? One possibility is that we *define* rationality as maximization *constrained by weakness of will*. Then we need only apply common knowledge of rationality to induce *b*'s belief in *a*'s weakness of will. I submit that this is indeed the concept of rationality in non-cooperative game theory and in neoclassical economics. In what follows choices that maximize payoffs subject to the constraint of weakness of will be called *perfectly rational* choices, not because their outcomes are perfect (as the example shows) but because it is rationality in this sense that defines subgame perfect equilibrium.

But common knowledge of *perfect* rationality is not the only possibility, and we need to consider others. First consider the possibility that *b* believes *a* is dishonest. Then *b* will not believe any assertions by *a* that he will choose "up" at decision point A2 and accordingly *b* chooses strategy 5′. But (i) *a*'s honesty is of concern to *b* only if *b* believes *a* has strong will. If *b* believes *a* has a weak will then *b*'s decision will not be affected by the further knowledge that *a* is honest or dishonest. (ii) *a* can benefit by acting dishonestly only if *b* believes both that *a* has strong will and that *a* is honest. (iii) Accordingly, we must consider a 4-step game in which *a*'s decision whether to act honestly or dishonestly is the first stage. If *a* has a strong will he can commit himself to one or the other and carry out the commitment. (iv) However, if *b* believes *a* has chosen to act honestly, then *a*'s best response is dishonesty. (v) Therefore, this first stage requires a mixed-strategy solution. (vi) Since *b* is rational, he will be aware of this and will accordingly estimate the payoffs of strategies 4′ and 5′ as expected values reflecting the optimal mixed strategy for *a*, which is to act honestly with probability 2/5. Thus, *b*'s belief that *a* will be dishonest with probability 1 is either irrelevant or irrational.

Common sense suggests that rationality, strength of will and honesty are distinct traits and that rational individuals may exist in positive numbers whose will is strong and weak; and that within each category some are honest and some are crooked. These conditions are also relative and a typical person is more likely to act in an honest and strong-willed way in some circumstances than others. Suppose *b* believes that a very large proportion of all human beings have weak wills, but has no way to know which type *a* is. In that case, once again, he would estimate the payoffs of his choices as expected values, using the probabilities based

on the frequency of weakness of will in the population and such other evidence as he may have. To fail to do so would be irrational or at best boundedly rational! Perfect rationality is naive on this score, and in what follows decisions based on maximization with estimates of the probability that other agents have strong wills and are honest will be called "sophisticated rationality" to distinguish them from the rationality expressed by subgame perfect equilibrium. (For the concept of sophisticated rationality and some evidence that there are multiple types of decision-makers in a real human population, see Stahl and Wilson, 1995.)

It seems that $b$'s behavior, as assumed in subgame perfect equilibrium theory, can be rational only if $b$ believes that weakness of will is a common trait of all human beings. This in turn can be considered a rational belief only if (i) it is true, or (ii) $b$'s experience has been so idiosyncratic that it seems to $b$ that the belief is true, although $b$ is mistaken. We can eliminate (ii) as inappropriate to be the basis of a general theory, and conclude that for subgame perfect equilibrium theory, universal weakness of will is a necessary assumption. If both weakness of will and perfect rationality are common human characteristics, then there is little point in distinguishing between them. But the results of such an identification can be rather peculiar. The results of the example of intertemporal inconsistency and of the two-person game from Figure 10.2 can both be stated in the following way. (i) Define rationality as perfect rationality. (ii) Suppose decision-maker $a$ in fact adopts strategy 2 (or 2′) and carries it out. (iii) As a result of this choice, decision-maker $a$ is better off. (iv) Decision-maker $a$ has acted irrationally. Stated in just that way, perfect rationality is not a very intuitively appealing concept of rationality.

How would von Neumann and Morgenstern have treated the game in Figure 10.2? In the first instance they would have expected the two players to form a coalition around strategies 2′, 4′, since the total value generated by that pair, 7371, dominates all other strategy pairs. This will be no difficulty if both have strong wills and are honest.

In Game 10.3, the non-cooperative equilibrium is also the assurance value for both players. Unlike Game 6.8, for example, Game 10.3 has no threat strategies. For a game like Game 6.8, Von Neumann and Morgenstern seem to envision a negotiating process along the following lines: agent $a$ says "if you adopt strategy D2 I will adopt strategy P, leaving you with 5 rather than 7." This is a threat designed to increase $b$'s bargaining power, and for von Neumann and Morgenstern (and Nash to the contrary) *all feasible threats are credible.* But none of this makes sense unless each agent believes the other has strength of will enough to carry out his threats, even when they are irrational in the sense of perfect rationality. For von Neumann and Morgenstern, a rational agent maximizes his

expected utility on the assumption that all agents maximize and all have strong wills. Strength of will is here considered an aspect of rationality. In what follows, rationality in this sense will be called "ideal rationality."

Notice that the assumption that all feasible threats are credible is central to the definition of the coalition function in von Neumann and Morgenstern, and as most cooperative game theory is based on the characteristic function, we may conclude that the assumption of ideal rationality is characteristic of cooperative game theory. Thus, it is appropriate to distinguish between cooperative and non-cooperative game theory by noting that while non-cooperative game theory assumes perfect rationality, cooperative game theory assumes ideal rationality.

### 10.4.5 Bounded Rationality

We now have three concepts of rationality: perfect, ideal, and sophisticated. Ideal rationality is not perfect, perfect rationality is not ideal, and neither is sophisticated (if some agents in the actual world have strong wills and others do not). Moreover, there is reason to believe that none is very descriptive of actual human behavior. Computation of a perfect, ideal or sophisticated rational solution to a problem may require a great deal of cognitive effort, and cognitive effort is a very scarce resource for real human beings. Acting according to a rule of thumb that may not be "optimal" from any point of view is "boundedly rational." Can "boundedly" rational decisions be ideal, perfect, or sophisticated?

The answer is yes. Suppose that in fact the population comprises individuals both with strong and weak wills, and this is a known fact. Then only sophisticated rationality can be defended as consistent with rational belief. Suppose then that an agent faces a threat, and will attempt to judge the credibility of the threat. Suppose also that the individual believes, on evidence and experience, that most people (though not all) have weak wills. Then computing a perfect equilibrium for the game will be easier than computing a sophisticated solution, since the sophisticated solution requires us to know the perfect solution anyway (and in a particular case, such as Figure 10.2, the perfect solution may be very easy indeed). Then the rule that "only subgame perfect threats are credible" is boundedly rational. Suppose in addition (as seems very plausible) that most people are better able to exercise a strong will in some circumstances than in others, and that in particular, the agents are situated in a culture that values personal honor highly and regards oath-breaking as dishonorable. In such a society, we suppose, the probability that an oath will be carried out is very high. Then the ideally rational solution may be boundedly rational, in case oaths have been sworn. Finally, consider the game of "running the red light." A

stoplight is a correlated equilibrium solution to an anticoordination game: its perfectly rational solution is to obey the light. However, if an individual is ideally rational, he may commit himself to running the stoplight, and carry the strategy out despite the temptation to stop at the last second. This could maximize his utility if the light has just changed and his intention is made very clear by speeding up. Having observed that about one in three drivers will do this (in Philadelphia) the sophisticated solution of delaying one's start into a green light (without trying to see whether the driver coming the other way has speeded up or not) is boundedly rational, although it is neither ideal nor perfect.

### 10.4.6 Perfect Rationality and the Manipulation of Elections

Let us return to Gibbard's theorem on manipulation of voting. As we have observed, it belongs to non-cooperative game theory. Indeed Gibbard remarks (1973, p. 593) "If a system does make outcomes a function of preferences, it is in virtue of individual integrity, ignorance, or stupidity . . .." An interesting example arises from the US presidential election of 2000. In that election, about 3 percent of the popular vote was cast for the Green Party candidate, Ralph Nader, and it is quite likely that if those votes had gone to Al Gore, George Bush would never have become president. Some Democrats criticized the Green Party voters very bitterly, with the implication that they acted dishonorably by failing to cast a strategic vote for Gore. Now, one could argue (consistently with Gibbard's model) that the "honest" vote in an American presidential election is a vote for whichever of the "realistic" candidates one prefers. But this example may be better understood in the light of cooperative game theory.

In elections prior to 2000, Democrats and Greens had constituted an informal de facto coalition. (American politics does not allow formal coalitions, with the partial exception of New York State.) In the period before the 2000 election (in effect) the Greens demanded a bigger payoff from the coalition, threatening that if their demand were not met, they would vote for a third-party candidate, taking whatever risk of a Republican (or worse) victory that might follow. The Democrats, perhaps assuming that the Greens would act with perfect rationality – or that however strong their will might be the Greens were bluffing – declined the demand. The Greens would " . . . execute the threat[,] . . . not . . . something [the Greens] would want to do, just of itself" (Nash, 1953, p. 130). In so doing the Greens acted with ideal, but not perfect, rationality. The Democrats' failure was a failure of sophisticated rationality: they failed to take adequate account of the mixture of perfectly and ideally (and sophisticatedly) rational agents in their interactive decision problem.

It seems clear that Gibbard's model presupposes perfect rationality. In a world of ideally rational (and honest) agents, it would be quite possible for a grand coalition to arrive at a "social contract" to vote honestly, and come to a general agreement as to what that might mean. "Individual integrity" would then assure that collective decisions would depend only on honestly expressed individual preferences. (Perhaps we should not assume too hastily that this would be a good thing.) In the absence of such a social contract, though, it is not clear that nonmanipulated voting has any meaning at all in a cooperative approach to game theory. Is not every political coalition an attempt to manipulate the outcome? In a real world in which there are at least a few ideally rational agents, some of whom are honest sometimes, the only real rationality is sophisticated rationality. But this hardly favors a case for nonmanipulated voting! What does seem clear, nevertheless, is that models based on perfect (or ideal) rationality alone are likely to mislead us. The very concept of a coalition, after all, derives from (European) electoral politics, and to try to discuss political choice in non-cooperative terms that exclude most coalitions seems odd, if very American.

### 10.4.7 Coalition Formation

Suppose that the population includes individuals both with strong and weak wills, and that at least some of those with strong wills are honest. How then will coalitions form? First, there will be no mutually beneficial coalitions comprising only the weak-willed. Such coalitions would accomplish nothing that would not be accomplished by a non-cooperative equilibrium. The typical coalition, then, will include at least a subset of strong-willed individuals who adopt threat strategies that encourage the others to keep their agreements and correlate their strategies so as to increase the value of the coalition. These strong-willed individuals may be known to the others as leaders, but more probably as officious busybodies, nosy parkers, or snitches. It may be that the officious busybodies, nosy parkers and snitches will form a grand coalition and formalize some of their threat strategies as institutions such as property rights and enforcement of contracts. If, as seems likely, people are better able to act with strength of will in some circumstances than in others, we are likely to see cooperative arrangements more often in some social circumstances than others, for example among people of a common religious faith, and to see a good deal of non-cooperative interaction among the coalitions that do occur. If a part of the population are both strong-willed and dishonest, they may be able to form some coalitions for their dishonest purposes, by means of committed threat strategies, even though dishonesty breeds

distrust and distrust is an obstacle to cooperation, as in Game 10.3. For coalitions of this kind, the "irrationality" of gang vendettas would be seen as an expression of ideal (though not perfect) rationality. We recall criminal gang leaders with nicknames like "Bugsy" or "Banannas," nicknames that express their lack of (perfect) rationality.

The rationality of Selten 1964 is ideal rationality. "Ideal rationality" links rationality to strength of will. It seems that "ideal rationality" is characteristic of cooperative game theory and is the substantive difference that distinguishes cooperative game theory from non-cooperative game theory. "Perfect rationality" links rationality to weakness of will. The example of intertemporal inconsistency shows that "perfect rationality" characterizes neoclassical economics as well as non-cooperative game theory. The third concept of rationality is "sophisticated rationality," which is consistent with the belief that the population includes both types with strong and with weak wills. This belief leads toward a world very much like the world we seem to live in, a world not susceptible to analysis in terms either of perfect or of ideal rationality.

It does seem likely that both cooperative and non-cooperative game theory are mistaken in their extreme views. On the one hand, commitment does occur in human interactions, and on the other hand, it is not easy nor altogether predictable. It seems more plausible to say that real human beings can make and carry out commitments – that is, some people can, sometimes, and under some circumstances! We may then suggest some circumstances that favor successful commitment:

(1) The existence of a contract enforced by a third party such as the state or a private bondholder
(2) Repeated interaction with other parties, over a long term
(3) Patience
(4) Strong relevant social norms of promise keeping or honor
(5) A high trust environment
(6) An agreement consistent with motives of equity or reciprocity
(7) A large differential between the payoffs attainable by cooperative action and those that result from non-cooperation

No doubt other such circumstances can be offered.

## 10.5 SUMMARY

Among the simplifying assumptions of both non-cooperative and cooperative game theory, there are a number that call for reconsideration.

For non-cooperative game theory, the identification of "strategies" with behavior strategies is oversimple, and the identification of rationality with weakness of will is both counterintuitive and counterempirical. One could put it this way: for non-cooperative game theory, people are always opportunistic but never spiteful. Conversely, for cooperative game theory, people are never opportunistic but always spiteful. It seems fairly certain that real people are sometimes opportunistic and sometimes spiteful, sometimes both and sometimes neither, and may act in all these ways with great calculation and using all available information. The representation of the game in coalition form also rests on an assumption, the assurance principle, which rules out free-rider behavior, and consequently prevents any satisfactory discussion of externalities. An alternative, to adopt the partition function and base the values of imbedded coalitions on non-cooperative equilibria among the coalitions, will be explored in the second part of this book. Finally, the assumption implicit in cooperative game theory, that agents always have strong wills, seems no more satisfactory than the assumption in non-cooperative game theory that agents always have weak wills. The second part of the book will explore an idealization of the great complexity of rational behavior by real human beings, assuming that agents act cooperatively (with ideal rationality) toward their partners in coalitions but that the coalitions act non-cooperatively (with perfect rationality) toward other coalitions.

## NOTES

1. The inconsistency of the previous two paragraphs has been noted in a previous chapter, in the context of transferable utility, and will be further considered in the next section.
2. The necessary qualification is that the Nash equilibrium need not be unique even for a given partition, a difficulty we will deal with in a later chapter. In any case this is not the ambiguity with which Telser is concerned, as it does not arise from uncertainty about the exact partition formed.

# PART II

# Encapsulated cooperation

# 11. Coalition formation and stability

> you let the farmers alone . . . all they got to do is gang up efficiently among themselves . . . but they never can stay ganged up they run out on each other. (Archy the cockroach in Marquis,1950)

We have seen enough to suspect, and perhaps become persuaded, that neither cooperative nor non-cooperative game theory can alone supply a satisfactory foundation for public policy studies. This part of the book will outline a model of coalition formation that draws on elements of both cooperative and non-cooperative game theory. The theory will proceed from the following foundational assumptions: (1) Since externalities may occur, the cooperative game analysis will represent the game in partition function form. (2) The cooperative game analysis will also allow for nonaggregative games including games of "imperfect recall," so that the game in partition function form may not be superadditive. (3) In general, coalition structures other than the grand coalition may be stable. (4) Cooperative relations take place only within coalitions, so that there are no side payments between coalitions and the interdependent strategy decisions of different coalitions are non-cooperative. (5) Accordingly, the partition functions are determined as the payoffs to non-cooperative equilibria in the play between coalitions. (6) Individual agents decide non-cooperatively to affiliate themselves into coalitions to choose joint strategies and receive or make side payments for the coalitional play. This might be described as a model of *encapsulated cooperation.*[1]

This chapter will outline a cooperative analysis of games in partition function form, focusing on assumptions (1)–(3). The presentation in this chapter will be relatively intuitive and the arguments made by examples. A more rigorously mathematical treatment of some topics is reserved for Chapter 13. Some redundancy will be unavoidable as a result. The reader who is less interested in mathematical details should be able to skip Chapter 13 without loss of continuity. The themes of this chapter are continued in Chapter 12, along with assumption (6), and Chapter 14 will address assumptions (4) and (5) and apply the analysis to several relatively simple, illustrative examples. Chapters 15 and 16 will sketch a political economy drawing on these ideas.

## 11.1 FURTHUR TERMINOLOGY FOR PARTITION FUNCTIONS

A game in partition function form (we recall) consists of a set of players N (comprising n players) and a coalition value function. A partition is a set of sets of agents, that is, a set of coalitions such that each agent is a member of just one of the coalitions. The coalitions in turn are the members or elements of the partition. The value function assigns a value to a pair comprising the coalition C and a partition $\mathcal{P}$ of which the coalition C is one member. The value function or *partition function* will often be written as v($\mathcal{P}$, C).

A useful real-world illustration may be found in the 2005–2008 German government. This is based on a coalition of the two biggest parties, the social-democratic SDP and the religious conservative CDU/CSU. Neither of the traditional coalitions, the red-green coalition of SDP with the Greens nor the black-gold coalition of CDU/CSU and FDP had a majority. However, the CDU/CSU and FDP came closest, with more parliamentary seats (284) than the SDP with the Greens or than any other party. Thus, a minority government of the CDU/CSU and FDP might have been considered in some circumstances. But this would be possible only on the condition that the SDP and the Greens continued their unwillingness to join in any coalition with the Left party – since the SDP, Greens and Left together had a majority with 326 votes. The value of a black-gold coalition depended very much on whether the SDP and Greens would coalesce with the Left. The black-red coalition of the SDP and CDU/CSU, with an absolute majority, did not depend on how the other parties associated themselves and therefore promised more stability. Wherever a minority government is possible, in a parliamentary scheme, it will be maintained only on the condition that the other parties do not unite against it – its value depends on the other coalitions that are formed.

There seems to be no simple formula to compute the number of partitions for a set of cardinality N. It is not difficult to write an algorithm to generate all possible partitions for a set of cardinality N, however. The computation makes it clear that this number increases very rapidly with N. Table 11.1 shows the computed number of partitions for N = 1, . . . , 10.

### 11.1.1 Partitions and Imputations

Consider a partition that has only one coalition, the grand coalition of all agents, N. Conventionally we will call this *partition* $\mathcal{G}$. (It is important to distinguish the coalition itself from the partition of which it is a member,

*Table 11.1 Partitions of a set of N*

| N | Number of partitions |
|---|---|
| 1 | 1 |
| 2 | 2 |
| 3 | 5 |
| 4 | 15 |
| 5 | 52 |
| 6 | 203 |
| 7 | 877 |
| 8 | 4140 |
| 9 | 21 147 |
| 10 | 115 975 |

even when the distinction is trivial: to fail to do so is to invite confusion.) The unique partition that consists of N singleton partitions will be called the *fine partition, $\mathcal{F}$.*

In general we allow for externalities, but there may be some important games in which there are no externalities. If $\Gamma = \{N,v(\mathcal{P}, C)\}$ is such a game then the value $v(\mathcal{P}, C) = v(\mathcal{Q}, C)$ for any two partitions $\mathcal{P}$ and $\mathcal{Q}$ each of which has C as a member. Such a game will be called a *proper game*. The game in coalition function form $\Gamma^{\dagger} = N,v(C)$, where $v(C) = v(\mathcal{P}, C) = v(\mathcal{Q}, C)$, will be called the *game proper to* $\Gamma$. For a proper game we may apply the well-understood theory of games in coalition function form. For the most part we will be concerned with improper games, though.

The value of each coalition is a mass of "transferable utility" that may be distributed among the members of the coalition. The payments to the n players in the game can be denoted by **x**. Conventionally, **x** is written in boldface to denote that it is a vector and comprises $\{x_1, x_2, \ldots, x_n\}$ where $x_i$ is the net payment to player *i*. We write $x_S$ to denote the total of all payments to members of set S. The vector **x** is called an imputation and if $x_S = v(\mathcal{P}, S)$ for every S that is a coalition in $\mathcal{P}$, then **x** is *admissible for P*. That is, the condition of admissibility is that each coalition spends only its own value; "every tub sits on its own bottom." This definition excludes side payments from one coalition to another. For some purposes, $\mathbf{x}_S < v(\mathcal{P},S)$ would be sufficient, but the definition adopted here requires that to be admissible, an imputation must be efficient in the sense that each coalition distributes its entire value among its members. For examples, refer to the public goods production game, Game 2.5. There, line 2, 6,6,9 and 1,11,9 are both admissible imputations, but 7,7,7, 7,7,9, and 5,5,9 are not admissible.

### 11.1.2 Refinements of Partitions

Let $\mathcal{P}$ and $\mathcal{Q}$ be two partitions of N and suppose they are related in that $\mathcal{Q}$ is formed by subdividing the coalitions that are members of $\mathcal{P}$. That is, for every coalition B that is a member of $\mathcal{Q}$ there is a coalition C, a member of $\mathcal{P}$, and every individual member of B is also a member of C, or in the terms of set theory, B is a subset[2] of C. Then $\mathcal{Q}$ is a refinement of $\mathcal{P}$. Trivially, every partition is a refinement of itself, but when $\mathcal{P}$ and $\mathcal{Q}$ are different we can say that $\mathcal{Q}$ is a *proper refinement* of $\mathcal{P}$. We can observe in passing that every partition other than $\mathcal{G}$ is a proper refinement of $\mathcal{G}$. The fine partition, $\mathcal{F}$, is a proper refinement of every other partition. In Game 2.5, partition 2 is a refinement of partition 1, and partition 5 is a refinement of partition 2, but partitions 3 and 4 are not, because the two-person coalitions in those partitions overlap the two-person coalition in partition 2.

A *particulate* refinement of $\mathcal{P}$ with respect to S is one that subdivides only S, leaving the other coalitions in $\mathcal{P}$ unchanged. Note that every partition is a particulate refinement of the grand coalition $\mathcal{G}$ with respect to N. A granular refinement of $\mathcal{P}$ with respect to S is one that retains S, but allows any further refinement of N\S. Note that if $\mathcal{P}$ is the partition $\mathcal{G}$, comprising grand coalition, it has no proper granular refinement. These concepts are used to adapt the definition of superadditivity to games in partition function form in Chapter 13, Section 13.2. An example is given in Chapter 14, 14.4, that applies this adaptation.

### 11.1.3 Efficiency

For a game of this sort *efficiency* is an attribute of a partition together with an admissible imputation. The simpler concept of efficiency is that a partition is efficient if it generates the maximum total value, summed over all coalitions. Efficiency in this sense will be called a-efficiency, since Aumann and Dreze (1974) propose such a criterion of efficiency for a coalition structure. Economists customarily associate efficiency with the following condition on imputations: an imputation is efficient if there is no other imputation that can make at least one agent better off without making any agent worse off. This is Pareto efficiency. We may say that a partition $\mathcal{P}$ is p-efficient if there are Pareto-efficient imputations that are admissible for $\mathcal{P}$. If a partition is a-efficient then it is p-efficient, but the converse is not necessarily so. (See Chapter 13 for more detail and proofs.) In Game 2.5, only the grand coalition is efficient in either sense.

### 11.1.4 Candidate Solution

For our purposes, a *candidate solution* to a game in partition function form will be a pair comprising a partition $\mathcal{P}$ and an imputation $\mathbf{x}$ admissible for $\mathcal{P}$. Now, suppose $\mathcal{P}$ is a partition and S is a set of players in the game who *do not* form one of the coalitions in $\mathcal{P}$. However, the group S jointly considers withdrawing from the coalitions they now participate in and forming a coalition among themselves. This proposal is called a *deviation* from $\mathcal{P}$. If they do so and benefit by doing so then we say that $\{\mathcal{P},\mathbf{x}\}$ *is disrupted by* $S$.[3]

Suppose that the coalitions in $\mathcal{Q}$ comprise S together with other coalitions each formed by all of the members of a particular coalition in $\mathcal{P}$ who have not joined S. Then $\mathcal{Q}$ will be said to be the residual partition of $\mathcal{P}$ with respect to S and will be written as $\mathcal{P}'_S$. Again, consider Game 2.5 as an example. In Game 2.5, if $\mathcal{P}$ is the grand coalition, line 1, then for example, $\mathcal{P}'_{\{a,b\}}$ is line 2 and $\mathcal{P}'_{\{b\}}$ is line 3 (and not line 5!) If a group of two or more who do not comprise a coalition in the current partition consider deviating from the current partition and forming a new coalition, the *immediate* result would be the residual partition $\mathcal{P}'_S$.

## 11.2 STABILITY

An objective of these chapters will be to address the problem of "endogenous coalitions," that is, to determine what partitions, and thus what coalitions, are most likely to be observed. This will be treated as a question of *relative stability.* That is, of two partitions, the one that is relatively more stable is the one that is more likely to be observed. One threat to stability is captured (in the theory of games in coalition function form) by the theory of the core. The threat to stability is that a particular group, not a coalition in the current game, will "gang up . . . among themselves" and precipitate the formation of a new partition. Thus, if one of the two partitions can support some imputations that are in the core of the game, and the other cannot, then the first is judged to be more stable and therefore more likely to be observed in actual cases that approximate the game model. But what if there are no candidate solutions in the core? Then we may legitimately, if with due caution, rely on some less demanding criterion of stability. In general, if stability is *less likely* to be observed according to one of two criteria, that criterion will be said to be the *more strict* stability criterion of the two. Accordingly, we may begin with the core.

### 11.2.1 Coalition Formation and Relative Stability

The concept of the core for coalition function games is based on the idea that no group can be denied the value that they can generate if they withdraw from other associations and form a coalition among themselves. If no group can improve its total payoff by seceding in this way, then the existing situation is supposed to be stable. For present purposes, this will depend on the partition as a whole, since the value of a coalition depends on the partition as a whole. The core of a game is the set of all candidate solutions that are stable in this sense. If any set S of players in the game can impose a different partition $\mathcal{Q}$ and corresponding admissible imputation $\mathbf{y}$, then they can expect at least $\mathbf{y}_S$ in $\mathcal{P}$ as the price of their acquiescence, and if $\mathbf{x}_S < \mathbf{y}_S$, the candidate solution fails and is not an element of the core. If no partition $\mathcal{P}$ and imputation $\mathbf{x}$ meet this test, then the core is null. These ideas will need clarification appropriate for partition function games. One approach (and a discussion of some of the difficulties) is set out in Koczy (2007). This subsection will borrow some ideas from Koczy's work, while the following subsection will review it more systematically and critically.

Koczy gives a numerical example of a symmetrical four-person game that will serve to illustrate the difficulties. For Koczy's game the players are N = $\{a_1, a_2, a_3, a_4\}$ and the partition function is as given as Table 11.2. Since the game is symmetrical, we need consider separately only five families of partitions. Thus, for example, $\{i,j,k\}$, $\{l\}$ refers to $\{a_1, a_2, a_3\}$,$\{a_4\}$ or $\{a_4, a_2, a_3\}$,$\{a_1\}$ or any other of the four possible partitions into coalitions of 3 and 1 members, and so on. Koczy gives no "story" or application to motivate this example, but it might be a case in which decentralization can be advantageous to those who decentralize but creates negative externalities for others. Now consider line 1, the grand coalition, $\mathcal{G}$. The grand coalition yields a value of 8. However this value is distributed there will be two members who are paid a total of four or less; let $i, j$ denote those two. Then suppose S = $\{i, j\}$ deviate from the grand coalition. The residual partition $\mathcal{P}'_S$ is line four and S can divide 6 among

*Table 11.2 Game 11.1: Koczy's game*

| | Partition | Payoffs |
|---|---|---|
| 1 | $\{i,j,k,l\}$ | 8 |
| 2 | $\{i,j,k\},\{l\}$ | 6,1 |
| 3 | $\{i,j\},\{k\},\{l\}$ | 0,4,4 |
| 4 | $\{i,j\},\{k,l\}$ | 6,6 |
| 5 | $\{i\},\{j\},\{k\},\{l\}$ | 1,1,1,1 |

themselves. Accordingly, we might suggest that candidate solutions based on $\mathcal{G}$ are not stable. But this is a naive view (as we shall see) and accordingly a core defined in this way will be called the *naive core*.

To make this a little more precise, define the *naive excess* for S with respect to $\{\mathcal{P},\mathbf{x}\}$ as $e^n(\mathcal{P},\mathbf{x},S) = v(P'_S,S) - \mathbf{x}_S$. That is, for any set of players S and candidate solution $\mathcal{P}$, $\mathbf{x}$, the naïve excess is the difference between the value that S could obtain in $\mathcal{P}'_S$ and their total payment in the candidate solution. The excess can be negative and we will identify the naive core as the set of all candidate solutions for which, for all sets S that are not members of the partition $\mathcal{P}$, $e^n(\mathcal{P},\mathbf{x},S) \leq 0$. Put in more ordinary terms, the naive core consists of all candidate solutions in which no group can complain that they could do better if they were to withdraw and cooperate only among themselves (supposing the remainder do not reorganize themselves).

But now return to the deviation proposed by $\{i,j\}$. The residual partition $\mathcal{P}'_S$ is no more stable than $\mathcal{P}$ was, since $k,l$ can improve their payoffs by dissolving their coalition and playing as singletons, leading to partition 3. Since $\{i,j\}$ would then be worse off, this may lead them to hesitate before deviating from the grand coalition. Thus, perhaps the grand coalition will be stable after all. On this basis, even though $\mathcal{G}$ is not a member of the naive core for Koczy's game, it might be stable in a more foresightful sense.

We suggested that $\{i,j\}$ might hesitate to deviate from $\mathcal{G}$ because a further deviation by $k$ or $l$ would lead to partition 3, leaving $\{i, j\}$ worse off than they were at line 1. This hesitation would represent a measure of foresight, but perhaps not complete foresight, because 3 might not be stable either. From 3, $\{i,j\}$ would want to dissolve their own partnership, leading to line 5, and leaving $j$ and $k$ worse off than they were at line 4. Should we suppose that, with that thought, $k$ and $l$ would hesitate to deviate from line 4, rendering line 4 stable? And if so, could it be that $\{i,j\}$ would anticipate that, and accordingly feel confident that they could deviate from $\mathcal{G}$ and profit by it? In that case, we might conclude that $\mathcal{G}$ is unstable after all. Indeed, if $i$ and $j$ attribute the same foresight to $k$ and $l$ that they themselves practice, then they would rationally identify partition 4 as the successor of partition 1, given their deviation from the grand coalition.

Attitudes of optimism or pessimism might also influence the judgment of stability, a point stressed by Koczy. Suppose that $\{i,j\}$ simply ask themselves what the value of their separate coalition would be. Since it occurs in two partitions, the answer could be either 0 or 6. If they are optimistic, then they will focus on the payoff of 6 and disrupt the grand partition; accordingly we regard $\mathcal{G}$ as unstable. (As we have seen, extensive foresight could give some grounds for this optimism.) But if they are pessimistic,

they focus instead on the payoff of zero and do not deviate; in that case we regard $\mathcal{G}$ as stable.

These examples illustrate that (1) the naive core is indeed naive, since it allows for no foresight, (2) if we do allow for foresight, we may find a different set of stable partitions, since foresight may dissuade some coalitions from disrupting an existing partition even though, without foresight, it would seem profitable for them to do so, (3) depending on the degree of foresight we allow and how it enters the process, we may find a family of more or less stable sets of candidate partitions, and (4) if we admit given attitudes of optimism and pessimism, a greater degree of optimism will lead to a smaller set of stable outcomes.

### 11.2.2 Koczy's Recursive Core

Koczy deals with these difficulties by means of a recursive definition of the core. Koczy writes (p. 42) "In CFF games [that is, games in coalition function form] the value of a deviation does not depend on the reaction of the remaining residual players . . .." However in a game in partition function form, "After a deviation we may expect widespread reshuffling of residual players . . . *Given a deviation, the residual players face the problem of solving another, smaller PFF game.* We call this a *residual game*. Deviating coalitions must expect a residual core outcome to form" (italics in the original). Koczy then proceeds by induction, (1) defining the core of the trivial one-person game, and (2) given the definition of the core of an N − 1 person game, defining the core for an N-person game.

Following a deviation, as noted above, the residual players face a smaller game in partition function form among themselves, the residual game. In case the core of the residual game is unique and not null, this resolves any ambiguity about the value of the deviating coalition S. The residual players will adopt that partition and it determines the value of the deviation. However, for cases in which the core of the residual game includes two or more partitions (with at least one imputation corresponding to each) and in which the core is null, Koczy defines two core concepts, an optimistic and a pessimistic core.

(a) *Pessimistic core*: (1) Suppose the core of the residual game is non-null and non-unique. Then $\mathcal{P}$ is disrupted by S if $v(\mathcal{Q},S)>\mathbf{x}_S$ for every $\mathcal{Q}$ in the core of the residual game; (2) Suppose that the core of the residual game is null. Then $\mathcal{P}$ is disrupted by S if $v(\mathcal{Q},S)>\mathbf{x}_S$ for every $\mathcal{Q}$ a refinement of $\mathcal{P}$ granular with respect to S.

(b) *Optimistic core*: (1) Suppose the core of the residual game is non-null. Then $\mathcal{P}$ is disrupted by S if $v(\mathcal{Q},S)>\mathbf{x}_S$ for at least one $\mathcal{Q}$ in the core

*Table 11.3 Game 11.2*

| | Partition | Payoffs |
|---|---|---|
| 1 | $\{a,b,c\}$ | 15 |
| 2 | $\{a,b\},\{c\}$ | 12,2 |
| 3 | $\{a,c\},\{b\}$ | 12,2 |
| 4 | $\{b,c\},\{a\}$ | 12,2 |
| 5 | $\{a\},\{b\},\{c\}$ | 1,1,1 |

of the residual game; (2) Suppose that the core of the residual game is null. Then $\mathcal{P}$ is disrupted by S if $v(\mathcal{Q},S)>\mathbf{x}_S$ for at least one $\mathcal{Q}$ a refinement of $\mathcal{P}$ granular with respect to S. In any case if $\mathcal{P}$ is disrupted then the candidate solution $\{\mathcal{P}, \mathbf{x}\}$ fails and is excluded from the core.

The pessimistic core is pessimistic in the sense that a partition and imputation are disrupted only if the deviators can be certain that they will profit by doing so, while the optimistic core is optimistic in that the disruption takes place whenever there is any possibility that the deviators will profit. Perhaps one could multiply cases by applying some Bayesian or similar weighting based on the likelihood of a profitable outcome.

Notice, in passing, how Koczy's approach excludes dominance cycles. As an example consider Game 11.2, shown in partition function form as Table 11.3. (This might be a game with an indivisible technology and in which a coalition of two agents can gain market power at the expense of the third, with some cost in inefficiency.) As a candidate solution consider partition 1 with payoffs 5, 5, 5. Agents $a$ and $b$ may consider a deviation to partition 2, with payoffs 6, 6. This is profitable in itself, so we may conclude that 1, 5, 5, 5 is not in the naive core, and indeed the naive core for this game is empty. But partition 2 is not stable either. It will be disrupted if $c$ approaches a and offers to form $\{a, c\}$ paying 7,5 in partition 3. We might say that $c$ has raided the membership of the deviating coalition by recruiting a. But this in turn is disrupted by $\{b, c\}$ paying 5, 7 in partition 4, as $b$ raids the membership of $\{a, c\}$, and this in turn will be disrupted by $\{a, b\}$ paying 5,7 or by $\{a, b, c\}$ paying 2.5,5.5,7.5. We see that this game displays dominance cycles. With Koczy's approach, however, membership raiding is not allowed. Once agents $a$ and $b$ deviate to partition 2, the residual set is the singleton $\{c\}$, and since this is the only possible organization of a one-person set, partition 2 with 6,6,2 is in the core of Koczy's residual game, so there is no further reorganization. The deviation to partition 3 is a reorganization of the original game, not the residual game, and so is not considered as a possibility in Koczy's schema.

Taking this as a dynamic schema, how might it be justified? Koczy suggests that the deviation involves a commitment on the part of the deviating group, such as $\{a, b\}$ in the example, which will not be broken by subsequent reorganizations of the residual group. Perhaps we might envision it this way: coalitions are based on long-term contracts, which expire at different times. At random, in each period, one of the contracts expires, but may be renewed. However, each contract has a clause that allows renegotiation of all contracts if any of the groups currently in existence is broken. Then the contract governing coalition C expires, some of the members of C decline to continue it, and this releases members of other groups from their contracts as well. The new coalition S, comprising some members of C and some others, then forms with a long-term contract. The residual members are free to negotiate their new organization, which does not release the members of S because it does not break an existing group. The new organization then is based on a new set of long-term contracts and continues in force until one of the new coalitions expires and is not renewed. In what follows this will be called a *residual contract dynamics*.

Koczy allows "partitional" deviations, that is, simultaneous formation of two or more new coalitions that may disrupt an existing partition. Using this assumption, Koczy proves that a recursive core allocation must be Pareto-optimal, with the following reasoning: "Assume . . . there exists an outcome $(\mathbf{x}, \mathcal{P}) \in C(N,v)$ such that there is another outcome $(\mathbf{y}, \mathcal{P}')$ with $\mathbf{y} > \mathbf{x}$. Consider the – profitable – deviation by coalition N forming partition $\mathcal{P}'$. . . . Contradiction."

Koczy illustrates this point with the 4-person game, Game 11.1, above. We note that in Game 11.1, only partition 4 can support a Pareto optimum. However, Koczy remarks that partition 4 cannot be reached from the grand coalition $G$ by "coalitional deviation." This seems to be his reasoning: Let $G$ with payments 2,2,2,2 be the candidate solution. Suppose that a two-person coalition $\{i, j\}$ defects from it. The residual game is a two-person game among $k$, $l$ with two partitions, $\{k\},\{l\}$ for 4, 4 and $\{k, l\}$ for 6. Clearly the first dominates the second and so is the core of the two-person residual game; consequently the value of the deviation is 0 and $G$ cannot be dominated via that coalition. Nor can it be dominated via any other. Permitting partitional deviations, though, Koczy allows $\{i,j\}$ and $\{k,l\}$ to be formed simultaneously. Presumably his meaning is that in such a case the residual partition is null, so there can be no further reorganization and the value of the deviation is 6 for each coalition.

Note how this approach blurs the definition of a coalition! What is more important, one of the pragmatic objectives of this book is to discuss cases of persistent inefficiency. On the basis of Koczy's proof, his concept of the recursive core cannot be helpful in that discussion. Moreover, it

*Table 11.4 Game 11.3*

| | Partition | Payoffs |
|---|---|---|
| 1 | $\mathcal{G} = \{a,b,c\}$ | 6 |
| 2 | $\{a,b\},\{c\}$ | 3,0 |
| 3 | $\{a,c\},\{b\}$ | 1,1 |
| 4 | $\{a\},\{b,c\}$ | 1,1 |
| 5 | $\mathcal{F} = \{a\},\{b\},\{c\}$ | 0,0,8 |

is not clear how partitional deviations are implied by the residual core approach. The smallest games in which coalitional deviation is meaningful are 3-person games. Consider Game 11.3, Table 11.4, which does not have the symmetry of Koczy's example.

The only a-efficient partition in Game 11.3 is line 5, but it is not stable, since $\{a,b\}$ can benefit by deviating to impose partition 2. The residual, $\{c\}$, is in the core of the residual game since it is trivial, the singleton being the only possible organization for the residual. Let the candidate solution be $\mathcal{G}$, with $\mathbf{x} = \{2, 2, 2\}$. Secession by $\{c\}$ will leave $\{a,b\}$ in the core of their residual game at partition 2, so its value will be 0. A secession by $\{a\}$ or $\{b\}$ will not leave the two-person residual in its core, since $\{a,c\}$ or $\{b,c\}$ will in turn be disrupted by the unilateral secession of $\{c\}$, leaving $\{a\}$ or $\{b\}$ respectively with zero. Therefore deviations by $\{a\}$ and $\{b\}$ do not disrupt $\{\mathcal{P},\mathbf{x}\}$. No two-person deviation can disrupt $\mathcal{G}$ either. However, if we "Consider the – profitable – deviation by coalition N forming partition $\mathcal{P}'$" where $\mathcal{P}'$ is partition 5, and the deviation is supported by side payments of at least 1.5 to $a$ and $b$, then this "partitional" deviation can lead to the efficient partition. Then $\mathcal{F}$ is the unique partition in the (Koczy) core. This game raises the further question whether intercoalition side payments are permitted, on which Koczy seems to make no commitment. If intercoalition side payments are permitted, then partition 5 is uniquely p-efficient, but otherwise the grand coalition supports a family of Pareto-efficient imputations.

It seems, however, that the recursive core, partitional deviations, and intercoalition side payments are all independent assumptions and that these games could be solved with equal consistency assuming that intercoalition side payments and (more centrally to the point) partitional deviations are excluded. It seems that we cannot avoid making a series of arbitrary assumptions about the circumstances in which agents will or will not disrupt a partition by pursuing opportunities for profitable deviations. If we are determined to have a unique stability analysis, this seems to be a problem. What this book proposes instead is that for a particular game,

we focus on the most demanding stability concept that generates a non-null stable set for that particular game.

The analysis for this chapter, however, will differ from Koczy's in three important ways. First, it will be assumed that if no subset of N\S has a motive to disrupt $\mathcal{P}'_S$, then the value of the deviation by S is determined by $\mathcal{P}'_S$. Second, it is assumed that there is *no* cooperation between different coalitions: thus no side payments from one coalition to another and no coordinated deviations from an existing partition. Finally, rational foresight will be more completely allowed for.

### 11.2.3 A Successor Function

In subsection 11.2.1, discussing Koczy's game, we followed the consequences to be expected in the case of a two-person deviation from the grand coalition. The immediate result would be partition 4, but this would create a situation in which $\{\acute{k},\acute{l}\}$ could benefit by a further shift to partition 3. From partition 3, $\{\acute{\imath},\acute{\jmath}\}$ could in principle benefit by dissolving their partnership; but according to the residual contract dynamics they are not free to do so. Thus, partition 3, not partition 4, is the successor of partition 1 in the case of a two-person deviation.

As this example suggests, we will posit a *successor function*, $\mathcal{Q} = \mathcal{R}(\mathcal{P},S)$, where a deviation by S from partition $\mathcal{P}$ predictably gives rise to partition $\mathcal{Q}$. We will not attempt to identify $\mathcal{R}$ uniquely, nor, indeed, would it be useful to do so. As has already been argued, assumptions about the permissibility of partitional deviations and membership raiding must unavoidably be made, and the purpose of $\mathcal{R}$ is to capture those assumptions, whatever they may be. However, $\mathcal{R}$ must also be consistent with the rationality of the agents, and thus with a degree of foresight consistent with their rationality. For many games this consideration will determine $\mathcal{R}$, as it does in Game 11.1.

Once again we can find an illustrative example in the German parliament after the elections of 2005. Since no party and none of the traditional coalitions that had governed the country at one time or another in the previous 25 years had a majority, there was a period of what we might call recontracting, in which various coalitions (and corresponding partitions) were discussed. One that was discussed was the "traffic-light coalition" of the Greens, the Social-Democrats and the Free Democrats. However, the FDP firmly refused to participate in a traffic-light coalition. Let us consider $\mathcal{P}$ as the partition comprising the traffic-light coalition as the governing coalition and the rest of the parties acting independently in opposition, and let S be the singleton coalition comprising the FDP. The deviation of S from $\mathcal{P}$ directly produced, as $\mathcal{P}'_S$, a minority government of the SDP and

the Greens, a residual coalition with a plurality of 273 votes. However, this would not be likely to be a stable partition, and indeed it was not given much consideration. The residual after the deviation of S included both the CDU/CSU and the SDP, which differed in their parliamentary seats by only one seat, and as they played their residual game they formed the "grand coalition" (in parliamentary politics not a coalition of the whole but of the two largest, usually opposed parties). Denote as $\mathcal{Q}$ the partition with a coalition of the CDU/CSU and SDP governing and the rest acting independently in opposition. There is good reason to believe that the FDP expected this result, so that (in their view) $\mathcal{R}(\mathcal{P},\mathrm{S}) = \mathcal{Q}$.

The assurance principle underlying most traditional cooperative game theory can be accommodated as a special case in this schema. For the assurance principle, the successor $\mathcal{R}(\mathcal{P},\mathrm{S})$ is $\mathcal{Q}$ that yields the smallest value $\mathrm{v}(\mathcal{Q},\mathrm{S})$ over all partitions granular with respect to S. This was Thrall and Lucas's (1963) procedure for deriving a coalition function from a partition function and could be characterized as a globally pessimistic successor function.

*In general*, however, rational foresight may not determine the successor function uniquely. Suppose S deviates from $\mathcal{P}$; then the nonmembers of $\mathcal{P}$ have to play their residual game: they play it, again, subject to residual contract dynamics. Thus a group among the residual members forms a coalition $^2$S, and those who are in neither S nor $^2$S have yet another residual game to play. Coalitions $^3$S and $^4$S are formed in the same way, and this process eventually terminates either because all agents are committed to one of the new coalitions or because none of the remainder have anything to gain by further reorganization. (In a majority game such as the German parliament this occurs when a majority coalition has been formed, if not sooner.) At the end, each of $^2$S, $^3$S, $^4$S is better off than they could have been in the partition from which they deviated, and in that sense they have rational foresight. But there may be more than one such sequence if, for example, at step 3 there could be more than one group that could have formed a deviation $^3$S with the expectation of being better off at the end. A partition that results from such sequence will be called a *rational successor* of $\{\mathcal{P}, \mathrm{S}\}$. If there is no $^2$S that meets the criteria for a sequence, in that no group $^2$S within $\mathcal{P}'_{\mathrm{S}}$ can anticipate being better off as a result of deviating from $\mathcal{P}'_{\mathrm{S}}$, then $\mathcal{P}'_{\mathrm{S}}$ is the rational successor of $\{\mathcal{P},\mathrm{S}\}$. Thus, there will always be at least one rational successor to $\{\mathcal{P},\mathrm{S}\}$, but in general there may be more than one.

However, we can set limits to the successor functions that can be consistent with rational foresight by computing an optimistic successor function $\mathcal{R}^-(\mathcal{P},\mathrm{S})$ and a pessimistic successor function $\mathcal{R}^+(\mathcal{P},\mathrm{S})$. The optimistic successor function is optimistic in that it selects among the rational

successors to $\mathcal{P}$,S the one that yields the largest value v($\mathcal{P}$,S), while the pessimistic successor function selects the one that yields the smallest value v($\mathcal{P}$,S). If these disagree, then the rational successor function might have to be determined by some further test, such as a Bayesian judgment of the relative likelihood of the different successors. In case the optimistic and pessimistic successors agree, as they will do for some important games, there is no ambiguity about the value of a potential deviating group S. Even if there is more than one rational successor, all yield the same value for S, and this will be sufficient for our purposes.

In a particular case it may be that $\mathcal{R}(\mathcal{P},S) = \mathcal{P}'_S$. In case $\mathcal{R}(\mathcal{P},S) = \mathcal{P}'_S$ for every $S \notin \mathcal{P}$, the partition $\mathcal{P}$ will be said to be *brief*. If, for a particular game $\Gamma = \{N,v(\mathcal{P},S)\}$, every partition is brief, then the game $\Gamma$ is said to be brief. Every proper game is brief.

Alternatively, if people are *in fact* boundedly rational and shortsighted, and if the system is so complex that they make no effort to anticipate what the reorganizations of the residual might be, then we might identify $\mathcal{R}(\mathcal{P},S) = \mathcal{P}'_S$ in general. The point is that $\mathcal{R}$ should be consistent with our assumptions about agent rationality, whatever they may be. For our purposes it is also necessary that $\mathcal{Q} = \mathcal{R}(\mathcal{P},S)$ is granular with respect to S. In that case the value of S is defined as v($\mathcal{Q}$,S) where $\mathcal{Q} = \mathcal{R}(\mathcal{P},\mathbf{x},S)$. However, if $\mathcal{Q}$ is not granular with respect to S then $S \notin \mathcal{Q}$ so that v($\mathcal{Q}$,S) is undefined. The condition that $\mathcal{Q} = \mathcal{R}(\mathcal{P},S)$ is granular with respect to S is fulfilled if we adopt a residual contract dynamics.

### 11.2.4 The Core with a Successor Function

Accordingly, taking the successor function as a given datum, define the excess for set S, given candidate solution $\{\mathcal{P},\mathbf{x}\}$, as $e(\mathcal{P},\mathbf{x},S) = v(\mathcal{Q},S) - x_S$, where $\mathcal{Q} = \mathcal{R}(\mathcal{P},S)$. In case the members of N\S have no reason to disrupt $\mathcal{P}'_S$, then $\mathcal{Q} = \mathcal{P}'_S$ and the excess according to the successor function is the same as the naive excess. For a brief game this will always be true. In general, however, it may not be.

Following the outline of the Aumann-Dreze core for coalition structures, we then define the core as comprising $\{\mathcal{P},\mathbf{x}\} \ni \forall S \subset N,\ S \notin \mathcal{P}$ $e(\mathcal{P},\mathbf{x},S) \leq 0$. That is, in ordinary language, a candidate solution will be in the core if there is no deviation that will disrupt it in the sense that the deviating group are better off in the successor partition. The core will be denoted by $\Xi$ and a candidate solution in the core will be said to be $\xi$-stable. This extended core conception is a generalization of the concept of the core for games in coalition function form, given the assumptions we have made about $\mathcal{R}$, and in particular that $\mathcal{Q} = \mathcal{R}(\mathcal{P},\mathbf{x},S)$ is granular with respect to S. A formal proof will be given in Chapter 13, but it

is worthwhile to sketch it here. For a proper game, the value $v(\mathcal{Q},S)$ is independent of the partition $\mathcal{Q}$, provided that $\mathcal{Q}$ includes S as an element. Thus, for any $\mathcal{R}$ in the family considered here, and for the game $\Gamma^{\dagger} = \{N, v(S)\}$; $\Gamma^{\dagger}$ proper to $\Gamma = \{N, v(\mathcal{P},S)\}$; $v(S) = v(\mathcal{Q},S)$ and so the excess for $\Gamma^{\dagger}$, $v(S) - x_S$ is identical with $e(\mathcal{P},\mathbf{x},S)$. It follows that the core of $\Gamma$ for $\mathcal{P}$ is identical to the Aumann-Dreze core for coalition structure $\mathcal{P}$ for $\Gamma^{\dagger}$. In ordinary language, when the coalition function is applicable, the core as defined here gives the same results.

In general the excesses, $e(\mathcal{P},\mathbf{x},S)$, may differ for an optimistic, a pessimistic, and a rational successor function. Denote the excess for an optimistic successor function as $e^{-}(\mathcal{P},\mathbf{x},S)$, and the excess for a pessimistic successor function as $e^{+}(\mathcal{P},\mathbf{x},S)$. Clearly $e^{-}(\mathcal{P},\mathbf{x},S) \geq e^{+}(\mathcal{P},\mathbf{x},S)$. The core as determined for a pessimistic and an optimistic successor function may differ both from one another and from a rational successor function. Let the core as determined from a rational successor function be denoted by $\Xi = \{\mathcal{P},\mathbf{x}\}$. Then the core as determined from an optimistic successor function is denoted by $\Xi^{-}$ and the core as determined from a pessimistic successor function is denoted by $\Xi^{+}$. Then we have $\Xi^{-} \subset \Xi \subset \Xi^{+}$. (Details and proofs are in Chapter 13.) The optimistic core is the smallest – that is why the optimistic constructs are symbolized by the minus sign, as in $e^{-}(\mathcal{P},\mathbf{x},S)$. We will also sometimes speak of candidate solutions as $\xi^{+}$, $\xi^{-}$-stable or unstable, that is, pessimistically or optimistically stable or unstable.

If the agents in a group S, contemplating a deviation from $\mathcal{P}$, are optimistic about the value that will be realized by their deviation, then they will be more likely to disrupt $\mathcal{P}$. This means that $\mathcal{P}$ is less likely to be stable, and the set of stable partitions and imputations smaller, if agents are optimistic than otherwise. Conversely, the set of stable partitions and imputations will be larger in case the agents comprising S are pessimistic about the value that will be realized by their deviation. In that sense, the pessimistic core is a less strict criterion of stability than the optimistic core.

In what follows, therefore, our procedure will be first to determine the optimistic core of a game in partition function form. If the optimistic core is not null, then further attention will be limited to partitions in the optimistic core. If, however, it is null, then we will proceed to analysis of the pessimistic core.

### 11.2.5 Example

To illustrate these proposals consider Game 14.1, the NIMBY game. The game is based on the idea that a facility is to be built that will inconvenience one of the three players *a*, *b*, and *c* but will also supply a public good

*Table 11.5 The partition function for the NIMBY game*

| | Partition | Payoffs |
|---|---|---|
| 1 | $\{a,b,c\}$ | 14 |
| 2 | $\{a,b\},\{c\}$ | 8,6 |
| 3 | $\{a,c\},\{b\}$ | 8,6 |
| 4 | $\{b,c\},\{a\}$ | 8,6 |
| 5 | $\{a\},\{b\},\{c\}$ | 3,3,3 |

beneficial to them all. Chapter 14, Section 14.2.1 gives more detail. The partition function is shown as Table 11.5.

First we need to determine a successor function. Consider deviations of the form (m, $\{i,j\}$), where m denotes any partition and $i$, $j$ any two agents. Since the residual, $\{k\}$, cannot reorganize, $\{\{i,j\}, \{k\}\}$ is the unique successor. Consider deviations of the form ($\mathcal{G}$, $\{i\}$). This leads immediately to $\{\{i\}, \{j,k\}\}$, and since the residual, $\{j,k\}$, can reorganize only by decomposing and will lose by such a reorganization (supposing both are paid at least 3 in coalition $\{j,k\}$), the successor is $\{\{i\}, \{j,k\}\}$. Consider deviations of the form (m, $\{i,j,k\}$), where m = 2, 3, 4, 5. Since the residual is null no further reorganization is possible and $\mathcal{G}$ is the successor. Consider deviations of the form ($\{\{i,j\}, \{k\}\}$, $\{i\}$). Then $\mathcal{P}'_S = \mathcal{F}$, but the residual, $\{j\},\{k\}$, can profitably coalesce so that the successor is not $\mathcal{F}$ but $\{\{i\}, \{j,k\}\}$. (Note that it follows that this game is NOT brief.)

We see that the successor is always unique so that, in this case, we need not consider optimistic and pessimistic cases. Is $\mathcal{G}$ $\xi$-stable? Again consider a deviation of the form ($\mathcal{G}$, $\{i\}$). The excess is $6 - x_i$ and it follows that for an imputation $\mathbf{x}$ to be $\xi$-stable, it must have each $x_i \geq 6$, $x_N \geq 18$, and this is not admissible; so $\mathcal{G}$ is not $\xi$-stable. Consider a deviation of the form ($\{\{i,j\}, \{k\}\}$, $\{i\}$). Since the successor is not $\mathcal{F}$ but $\{\{j,k\}, \{i\}\}$, again, the excess is $6 - x_i$, so that a stable imputation for each coalition of the form $\{i,j\}$ must yield $x_{\{i,j\}} \geq 12$, which is not admissible, and it follows that partitions of the form $\{\{i,j\}, \{k\}\}$ support no candidate solutions in the core. Consider a deviation of the form ($\mathcal{F}$, N). Its excess is identically $5 > 0$, so the fine partition also is not stable. The conclusion is that the core for NIMBY is null.

The last paragraph should raise some questions. The deviation of a singleton from the coalition to produce the public good in partition $\{\{i,j\}, \{k\}\}$ above leads to a disruption that is cyclical, in that it returns a partition of the same form, though with a different assortment of agents among coalitions. In plainer terms, we have just reasoned that $\{i\}$ would abandon his coalition with $j$ to become a holdout, in the confidence that $k$ would abandon his holdout position and join with $j$ in a new coalition to produce

the public good. But is this a reasonable judgment for $\acute{\imath}$ to make? It is, if for some reason $\acute{\imath}$'s decision not to join in producing the public good is a commitment, while the pre-existing position of $\acute{k}$ not to produce the public good is not a commitment. This is an artifact of the dynamic assumptions of the previous section, and in this instance, not a particularly plausible one. Accordingly, we will explore an alternative, less limiting dynamic schema in the following chapter.

## 11.3 SUMMARY

This chapter began the construction of a model of the role of coalitions in interdependent decisions. We begin with the game expressed in partition function form. A candidate solution for such a game comprises a partition and an imputation that is admissible for that partition. However, such an arrangement may prove to be unstable, in that a group not organized within the current partition may be able to withdraw and organize, precipitating a new partition. To judge the benefits of doing so, they will have to consider the reactions of the residual group, who may reorganize themselves. This consideration leads to a successor function, which describes a partition that the deviating group can expect to see if they do indeed disrupt the existing partition. Using the successor function, we may extend the concepts of the core and the nucleolus, although in some intractable cases these extensions need not be unique. We can always derive limits for the core under assumptions of optimism and pessimism, and for more tractable cases they will agree, resolving any ambiguity.

## NOTES

1. This phrase borrows, and seemingly reverses, a phrase from Etzioni (1988), who wrote about encapsulated competition. But what we think of as market competition will indeed be imbedded in the cooperative game, as it is "encapsulated" in non-individualist institutions for Etzioni, so the opposition between the two views may not be as diametric as the grammatical reversal suggests.
2. For the purposes of this book, a subset may or may not be a proper subset and the conventional notation $B \subset C$ will be read as "B is a subset of C," not "B is a proper subset of C" as might be done in some recent work in set theory.
3. This terminology seems new in the present book but will be a little more direct than the more conventional phrase. The conventional expression would be along these lines: suppose that by deviating and forming a coalition resulting in partition $\mathcal{Q}$, group S can gain $y_S > x_S$, where $\mathbf{y}$ is a payoff vector admissible for $\mathcal{Q}$. The more conventional term in game theory would be to say that $\mathcal{Q}$ dominates $\mathcal{P}$ via $\mathbf{y}$.

# 12. Bargaining, weak dynamics, and consensus

As with the core for coalition function games, the partition function core may contain more than one partition or, even if only one partition, the core may contain many imputations admissible for that partition. This is well known as a shortcoming of the core, when the core is considered alone as a solution concept; but it is hardly surprising in a stability concept. Stability is a property that may be possessed by a family of states of a system. Nevertheless, given that a partition is stable, we naturally ask how the benefits of that coalition will be distributed among the members. When there is a range of stable imputations, the *specific* answer to that question is a matter of bargaining.

## 12.1 BARGAINING

In the tradition of game theory, the earliest and best-known discussion of bargaining is that of Nash, for two-person games. Shapley's value theory has been interpreted as a bargaining theory for n-person games, and it has the advantage that it can always be computed and is always unique. These advantages are shared by Schmeidler's nucleolus, and the nucleolus has the further advantage that it is within the core, whenever the core is non-null. Thus the nucleolus can be put to work as a "core-assignment algorithm" – that is, it may be used to determine which of the imputations in the core of a game is hypothetically most likely to occur. For the purposes of this book we adopt the nucleolus in this role.

To compute the nucleolus for a particular coalition structure (partition) $\mathcal{P}$ of a game, $\Gamma$, we first establish an ordering over all imputations $\mathbf{x}$ that are admissible for $\mathcal{P}$. This has two main stages, and the first stage is to order all sets of agents, $S \subset N$. There is a different ordering of the sets for every admissible imputation $\mathbf{x}$. We compute the excess, $e(\mathcal{P},S,\mathbf{x})$, for each set S. The set that ranks first for $\mathbf{x}$ is the set with the largest excess $e(\mathcal{P},S,\mathbf{x})$, the set that ranks second the one with the second largest excess, and so on. The excess can be thought of as a measure of the discontent with $\mathbf{x}$ felt by group S. Thus, the set that ranks first is the group most discontented with $\mathbf{x}$, and so on.

*Table 12.1 Game 12.1: an unsymmetrical proper game in coalition function form*

| Coalition | Value | Excess with imputation **x** | Excess with imputation **y** |
|---|---|---|---|
| $\{a,b,c\}$ | 12 | | |
| $\{a,b\}$ | 9 | 0 | 0 |
| $\{a,c\}$ | 6 | −1 | −1.5 |
| $\{b,c\}$ | 6 | −2 | −1.5 |
| $\{a\}$ | 3 | −1 | −1.5 |
| $\{b\}$ | 3 | −2 | −1.5 |
| $\{c\}$ | 3 | 0 | 0 |

Suppose the compensation committee of coalition C are considering two admissible imputations, **x** and **y**, and want to choose the one that will reduce the discontent of the most discontented group that may be affected by the choice. They first consider the group that ranks first for **x**, S, and the group that ranks first for **y**, T. If $e(\mathcal{P},S,\mathbf{x}) < e(\mathcal{P},T,\mathbf{y})$, then **x** is considered less objectionable than **y**, and ranks more highly than **y** in the ranking of imputations. If $e(\mathcal{P},S,\mathbf{x}) > e(\mathcal{P},T,\mathbf{y})$, the converse is true. If $e(\mathcal{P},S,\mathbf{x}) = e(\mathcal{P},T,\mathbf{y})$, then the choice between **x** and **y** cannot affect the discontent of the most discontented groups, so the committee then considers the groups that rank second in their discontent at **x** and **y**, computes their excesses, and if the excesses differ, ranks **x** and **y** according to their excesses. If the second most discontented groups are equally discontented with **x** and **y**, then the committee considers the third most discontented, and so on until a ranking of **x** and **y** is arrived at. This may be called the Schmeidler (1969) ordering.

To illustrate it consider Game 12.1 (see Table 12.1). This is a proper three-person game in which the benefits of working together can be secured by $\{a,b\}$ and agent $c$ adds nothing but its value as a singleton. (Thus, agent $c$ is a "dummy player.") We compare imputation **x** = 4,5,3 and **y** = 4.5,4.5,3. For all sets with **x**, $\{a,b\}$ and $\{c\}$ have the largest excesses, namely zero. However, considering the excesses for **y**, we find the same: **x** and **y** are tied with respect to $\{a,b\}$ and $\{c\}$. (This occurs because both imputations give $\{a,b\}$ and $\{c\}$ exactly their coalition values). Accordingly we check the second largest excesses by coalitions. For **x** this is a two-way tie at –1, while for **y** it is a four-way tie at –1.5. Since the second maximum, –1.5, is the smaller, **y** is the more acceptable imputation. (Indeed **y** is the nucleolus for this game.)

The Schmeidler ordering has some nice mathematical properties. In

particular, it always has a unique maximum. The imputation that corresponds to the maximal value is the *nucleolus* for coalition structure $\mathcal{P}$. The nucleolus for coalition structure $\mathcal{P}$ will sometimes be written as nuc($\mathcal{P}$).

Two qualifications should be stated.

First, the successor function is taken as given. The excess and therefore the ordering and the nucleolus may be different if the excesses are computed according to the pessimistic, optimistic, or rational successor function. Accordingly it will be necessary sometimes to speak of an optimistic or a pessimistic nucleolus. The optimistic nucleolus will sometimes be written as $\text{nuc}^{-}(\mathcal{P})$, and the pessimistic nucleolus as $\text{nuc}^{+}(\mathcal{P})$.

Second, there is a prima facie problem with the discussion of the nucleolus above. The compensation committee for coalition C can control only the part of imputation **x** that corresponds to the payments to their own members, that is, $\mathbf{x}_C$. The account above can be taken just as it stands only if C is the grand coalition of all elements of N. Nevertheless this discussion has followed Aumann and Dreze (1974) in defining the nucleolus for other partitions in the same way except that the constraint is imposed that **x** be admissible for $\mathcal{P}$. Here is an alternative. Suppose there are two or more coalitions $C_1$, $C_2$, . . .. The compensation committee for $C_1$ chooses $\mathbf{x}_C$ treating $\mathbf{x}_{N\setminus C}$ as given. To do otherwise would be to enter into some cooperative arrangement with another coalition, and by assumption there are no such intercoalition cooperative arrangements. Thus, coalition $C_1$ chooses its best response, in terms of the Schmeidler ordering, to the payment schedules of the other coalitions. Therefore, the imputation **x** would be a non-cooperative equilibrium comprising the mutual best responses of each coalition to the compensation schedules chosen by the others, a Nash equilibrium. The coalition structure nucleolus, as defined here, is such a Nash equilibrium.[1]

We will think of the nucleolus as a bargaining outcome, along the following lines. First, given $\mathcal{Q} = \mathcal{R}(\mathcal{P},S)$, $v(\mathcal{Q},S)$ is the threat point for group S. The excess is specifically an excess of the threat payoff for S over the payoff they receive given **x**. If this excess is positive, then the threat by S to deviate from $\mathcal{P}$ is credible in a non-cooperative sense. If the excess is negative, then the threat is not credible in a non-cooperative sense, but cooperative solutions are distinguished from non-cooperative solutions in that commitments (including threats) are credible that would not be credible in a non-cooperative approach. Thus, even if all computed excesses are negative (that is, $\{\mathcal{P},\mathbf{x}\}$ is in the core in the relevant sense), the group with the largest excess may reasonably be supposed to be the one most likely to defect. Thus, in either case, by choosing an imputation that reduces the greatest excess, the compensation committee reduces the likelihood that their coalition will be disrupted.

Since the core (whether optimistic, pessimistic, or rational) is defined by the condition that excesses for all S are negative, it is clear that the nucleolus must be in the core if any other imputation is. To be exact, whenever there is any $\mathbf{x}$ such that $\{\mathcal{P},\mathbf{x}\}$ is in the core, then $\{\mathcal{P},\text{nuc}(\mathcal{P})\}$ is in the core. The qualification for this point is that if the core is optimistic, then the nucleolus is $\text{nuc}^-(\mathcal{P})$, the optimistic nucleolus; and if the core is pessimistic, then the nucleolus is $\text{nuc}^+(\mathcal{P})$.

The nucleolus can be computed by linear programming, although the details can be quite complex. In any case, the intimate connection between the nucleolus and the core, via the excess function, seems to make it the natural choice for a "core assignment algorithm," to settle the payments in a core solution that allows more than one payment schedule for one or more partitions.

## 12.2 WEAK DYNAMICS

Here is an alternative to the contractual scheme proposed in the previous chapter. Again, coalitions are governed by long-term contracts, and from time to time the contract binding a particular coalition expires. Again, all contracts have cancellation clauses that release their members in case the expiring contract is not renewed. However, what follows is a period of "recontracting;" that is, no new contracts are formed until every agent has a coalition and payoff that is the best that the player can achieve. With recontracting we may have membership-raiding, in that members of a deviating group S may be approached by members of the residual to propose yet another deviation. Then a cautious group will not deviate unless the deviation leads to a new partition that is itself stable – in particular they will want to avoid precipitating a cycle such as we have observed in Game 11.2, since there is no foreseeable end to the recontracting in such a case. This might be considered as a longer-run schema, as agents look forward to developments that may occur when the contracts they commit themselves to have again expired.

For this process stability will be defined for a deviation $\{\mathcal{P},\text{S}\}$, $\text{S} \notin \mathcal{P}$, with the nucleolus $\text{nuc}^+(\mathcal{P}) = \mathbf{v}$. Then $\{\mathcal{P},\text{S}\}$ is $\omega$-unstable if S can do better in the partition $\mathcal{Q} = \mathcal{R}^+(\mathcal{P},\text{S})$ and, moreover, if $\mathcal{Q}$ is itself unstable in that some group $\text{T} \notin \mathcal{Q}$ can profitably disrupt $\mathcal{Q}$, bringing about a further transition to partition $\mathcal{U}$, then S still does better in partition $\mathcal{U}$ than they had done in partition $\mathcal{P}$. Since the outcome $\mathcal{U}$ of such a process may not have S as one of its member coalitions, however, $\text{v}(\mathcal{U},\text{S})$ may not be defined. To judge whether $\mathcal{U}$ will be more profitable for S than $\mathcal{P}$, it will be necessary to refer to particular imputations. Drawing from the previous section, the

payoff to any group given any partition will correspond to the nucleolus for that partition. Given $\mathcal{P}$, for example, an imputation $\mathbf{x} \neq \mathbf{v}$ will not be relevant to stability in this sense, in that bargaining within the coalitions of $\mathcal{P}$ will in any case cause a shift from $\mathbf{x}$ to $\mathbf{v}$ prior to any deviation, and we suppose that the agents will foresee this. To reiterate, then, S will deviate from $\mathcal{P}$ only if the pessimistic nucleolus of $\mathcal{Q}$ pays S better than that of $\mathcal{P}$ (this is condition v.1.1. in the next chapter) and either there is no T that is better paid in the nucleolus of $\mathcal{U}$ than in that of $\mathcal{Q}$ (condition v.1.2.a.) or S is better paid in the nucleolus of $\mathcal{U}$ than in that of $\mathcal{P}$ (condition v.1.2.b.)

Now, suppose we have a cycle; that is, $\mathcal{Q} = \mathcal{R}^+(\mathcal{P},S)$, $e^+(\mathcal{P},S,\mathbf{v})>0$, $\mathbf{y} = \text{nuc}^+(\mathcal{Q})$, and there is a group T, not a coalition in $\mathcal{Q}$, with $\mathcal{P} = \mathcal{R}^+(\mathcal{Q},T)$ and $e^+(\mathcal{Q},T,\mathbf{y}) > 0$. Then S is no better after both deviations than before – S will receive $\mathbf{v}_S$ in either case. Therefore, S will not disrupt $\mathcal{P}$ for no *ultimate* gain.

Accordingly, we will say that $\{\mathcal{P}, S\}$ is ω-stable if the conditions for ω-instability are not met, and that a candidate solution $\{\mathcal{P},\mathbf{x}\}$ is ω-stable if $\{\mathcal{P},S\}$ is ω-stable for every $S \notin \mathcal{P}$. While the ω-stability of $\{\mathcal{P},\mathbf{x}\}$ does not depend on $\mathbf{x}$, we are interested in the properties of candidate solutions, and we may say that $\{\mathcal{P},\mathbf{x}\}$ is stable *in the sense that* bargaining within the coalitions will lead to an imputation $\mathbf{v}$ that is stable.

In this spirit we may define two further stable sets of candidate solutions. Let $\Theta$ be the set of all ω-stable $\{\mathcal{P},\mathbf{x}\}$. Clearly $\Xi^+ \subset \Theta$ and if $\{\mathcal{P},\mathbf{x}\} \in \Theta$ for some $\mathbf{x}$ admissible for $\mathcal{P}$, it will follow that $\{\mathcal{P},\mathbf{v}\} \in \Theta$. (This is lemma v. 3 in the next chapter.)

Finally, we define the set $\Omega$ as the set of all $\{\mathcal{P},\mathbf{x}\}$ with the property that for all $S \notin \mathcal{P}$, either $\{\mathcal{P},S\}$ is ω-stable or $\mathcal{Q} = \mathcal{R}^+(\mathcal{P},S)$ is ω-unstable. This again reflects the idea that a group will not deviate to a partition that is unstable in the same sense as the one disrupted. Evidently $\Theta \subset \Omega$. Thus we have $\Xi^- \subset \Xi^+ \subset \Theta \subset \Omega$. Moreover, $\Omega$ can never be null. These assertions are proved in Chapter 13 and justify the claim that ω-stability is a weaker stability condition than core-stability.

We now have the following outline to analyze a game in for the purposes of this book. First, the partition function is determined by the analysis of non-cooperative play among the coalitions in each respective partition. Second, the relatively stable set of partitions is determined. If there are candidate solutions $\{\mathcal{P},\mathbf{x}\}$ in $\Xi^-$, then the relatively stable set are the partitions that correspond to candidate solutions $\Xi^-$. If there are no candidate solutions $\{\mathcal{P},\mathbf{x}\}$ in $\Xi^-$, but there are candidate solutions $\{\mathcal{P},\mathbf{x}\}$ in $\Xi^+$, then the relatively stable set are the partitions that correspond to candidate solutions $\Xi^+$. If there are no candidate solutions $\{\mathcal{P},\mathbf{x}\}$ in $\Xi^+$, but there are some in $\Theta$, then "relatively stable partitions" will be partitions in $\Theta$. Finally, if $\Theta$ is null,[2] "relatively stable partitions" will be partitions that

support some imputations in Ω. Since Ω is never null, we will always find some stable partitions at this stage. For each such partition, the payoffs the agents can expect correspond to the appropriately computed nuc($\mathcal{P}$). If either of the core concepts is not null, we will say in what follows that the solution is highly stable, while if the pessimistic core is null and the solution is in Θ or Ω, we may say instead that it is relatively stable.

As an example, again consider Game 14.1, NIMBY. As we saw in the previous chapter, the core for this game is null. Therefore we explore ω-stability for NIMBY. Indeed the set of ω-stable solutions for NIMBY is quite large. Partitions 1–4 all support candidate solutions in Θ. Here is the reasoning: We will first establish that partitions 2, 3, and 4 support ω-stable imputations. Consider a deviation of the form $\{\{i,j\},\{k\};N\}$. Since the excess for this deviation is identically zero, it is not profitable and so does not disrupt $\{i,j\},\{k\}$. Note that a deviation of the form $\{i\}$ is succeeded by $\{\{i\}, \{j,k\}\}$. We know that for the nucleolus 4,4,6, $e^+(\{\{i,j\}, \{k\}\},\{i\},\mathbf{x})>0$, satisfying condition v.1.1. in Chapter 13; however, equally, a deviation $\{j\}$ from $\{\{i\}, \{j,k\}\}$ also yields $e(\{\{i\}, \{j,k\}\},\{j\},\mathbf{z})>0$ with $\mathbf{z} = \text{nuc}^+\{\{i\},\{j, k\}\} = 6, 4, 4$, so that condition v.1.2 is violated. A similar argument applies to deviations of the form $\{\{i,j\},\{k\};\{j,k\}\}$. Therefore, a partition of the form $\{i,j\}, \{k\}$ cannot be ω-unstable. That is, any deviation from $\{i,j\}, \{k\}$ that is profitable leads to a successor that is unstable in the same sense, so with ω dynamics no deviations occur, and partitions of this form are ω-stable.

For $\mathcal{G}$, the nucleolus is 4.67, 4.67, 4.67. As we have seen, any one-person deviation will be profitable, satisfying condition v.1.1. Let the deviation be of the form $\{\mathcal{G}, \{k\}\}$ so that the successor is $\mathcal{Q} = \{\{i,j\},\{k\}\}$. Then $\mathbf{y} = 4,4,6$. A singleton deviation $T = \{j\}$ will be succeeded by $\mathcal{U}=\{\{i,k\},\{j\}\}$ with $\text{nuc}^+(\mathcal{U}) = \{4,6,4\}$; $\mathbf{z}_S = 4<\mathbf{v}_S=4.67$. Thus conditions v.1.2 cannot in general be satisfied and it follows that $\mathcal{G}$ is ω-stable.

## 12.3 THE CONSENSUS GAME

This chapter and the previous one have been concerned with cooperative game theory in a broad sense. However, the solutions can with equal validity be thought of as solutions of a non-cooperative game, and for some purposes (as we will see in Chapter 16) this is useful. Accordingly, this section outlines a non-cooperative game that yields the cooperative solutions discussed here. (But this is not an *implementation* of the cooperative game, as the enforceability of the coalition agreements is assumed, not demonstrated.)

We suppose, then, that the cooperative play described in the chapters

so far is imbedded in a non-cooperative game in which agents choose the coalitions with which they will associate themselves.[3] This is a rational action approach, so we suppose that the agents will anticipate the payoffs resulting from coalitional play, the cooperative stability or instability of a partition, and the distribution of payoffs according to the nucleolus. Thus, relatively unstable partitions can be ignored. This may be the end of the story if a relatively stable partition is unique. Suppose, however, that there are two or more partitions that support candidate solutions in the relatively stable set. Normalize the payoffs so that all are positive. Consider the following non-cooperative game: for each agent, the strategy is to designate a *partition* from the relatively stable set. If all designate the same partition, then the payoffs to individual agents are the nucleolus for the partition selected; otherwise the payoffs are zero. Thus, if the agents all arrive at a consensus as to the coalition structure of their society, they receive the nucleolus payoff for that coalition structure; otherwise nothing. This is the *consensus game.*

Here is a prima facie objection: it cannot be right to say that the strategy for an individual is the partition he chooses to participate in. The individual can control only his own coalition, at most. He cannot control the coalitions into which other people group themselves. Thus, to select a particular *partition* as the best response to the strategies adopted by others is to do what is beyond the powers of the individual. However, for the consensus game, it is not the best-response property of Nash equilibria that is of interest to us. Rather it is the consistent conjectures property: recall (Chapter 4, Section 4.4) that every Nash equilibrium in pure strategies is a consistent conjectures equilibrium. In order to choose a *coalition* with which to associate himself, the agent must form a conjecture as to the coalitions into which other agents will sort themselves and so the *partition* that will result from his decision. The key conclusion is that any stable partition, with payments according to its nucleolus, will be a consistent conjectures equilibrium in the non-cooperative process by which rational agents sort themselves out into coalitions. At equilibrium, all make the same conjecture about the coalition structure of their society, and as a result of their action based on that conjecture, the conjecture proves to be true. (McCain, 1992).

As an illustrative example, we may turn again to NIMBY. Because the game is highly symmetrical, the nucleolus will in every case correspond to equal payments within each coalition. Thus the consensus game in this case is characterized by the payoffs in Table 12.2. Each line of Table 12.2 describes a relatively stable cooperative solution to the NIMBY game. Since these solutions are not highly stable but only relatively stable, the difficulty of arriving at a cooperative solution in a NIMBY problem in actual practice is predictable.

*Table 12.2 The consensus game for NIMBY*

| Partition | A | B | C |
|---|---|---|---|
| 1 ($G$) | 4.67 | 4.67 | 4.67 |
| 2 | 4 | 4 | 6 |
| 3 | 4 | 6 | 4 |
| 4 | 6 | 4 | 4 |

Another important qualification is that the model predicts that within any public-good producing coalition, compensation will be made to the individual who actually produces the public good so that the net benefits are equal within the coalition. Compensation seems rarely to be proposed in public policy where NIMBY problems arise. But in the absence of compensation, that is, of side payments, no cooperative solution can occur and the predicted outcome is the non-cooperative equilibrium, in which the public good is not produced.

## 12.4 SUMMARY

For this chapter, the determination of payoffs within an imbedded coalition is modeled by the nucleolus, and this leads on to some dynamic stability concepts that are less stringent than the core. Taking the stable set as the set of candidate solutions that is stable under the most stringent rule, and assigning the payoffs as the nucleolus computed for a corresponding successor function, we resolve the game to a limited number of stable imputations to individuals. This set of imputations also defines a non-cooperative game, the consensus game, that corresponds to the cooperative game.

## NOTES

1. The proof of this proposition is quite easy once the proposition has been expressed formally. However, the formal expression is a bit digressive, so the details are reserved for Chapter 13, Section 13.4.3.
2. It is a conjecture that $\Theta$ may be null.
3. This is not a new idea in the literature of game theory. Nash's (1953) demand game and indeed a conjecture in von Neumann and Morgenstern have suggested this approach.

# 13. Formal aspects of games in partition function form

This chapter will review some of the concepts of games in partition function form that have played a role in the previous chapters. The purpose of this chapter is to give a formal statement of some of the concepts and results that have been used. The reader who is not interested in the mathematical aspects should be able to skip this chapter without difficulty in following the argument in the remainder of the book.

## 13.1 FUNDAMENTALS

Let N be an index set of agents in a game, $a_i \varepsilon N$, $i = 1, \ldots, n$. A partition $\mathcal{P}$ is a set of subsets $\{S_i\}$ where $S_i \neq S_j \Rightarrow S_i \cap S_j = \varnothing$ and $N = \bigcup_{S_i \in \mathcal{P}} S_i$. Let $\Pi_N$ be the set of all partitions of N and $\mathcal{P} \in \Pi_N$. $\mathcal{P} = \{C_1, C_2, \ldots, C_r, \varnothing\}$. $|C_i|$ will denote the number of members in $C_i$ and $|\mathcal{P}|$ will denote the number of nonempty coalitions in $\mathcal{P}$. A pair $\{\mathcal{P}, C_i\}$ with $C_i \in \mathcal{P}$ is called an embedded coalition. A coalition value function $v(\mathcal{P}, C_i)$ assigns a real number value to coalition $C_i$ where $C_i \in \mathcal{P}$. $\Gamma = \{N, v(\mathcal{P}, C_i)\}$ comprises a game in partition function form. For $a_j \varepsilon N$, $C_P(a_j) = C_i \in \mathcal{P} \ni a_j \varepsilon C_i$.

A game in partition function form is *proper* if $v(\mathcal{P}, S) = v(\mathcal{Q}, S)\ \forall \mathcal{P}, \mathcal{Q}$, $S \ni \mathcal{P} \in \Pi_N$, $\mathcal{Q} \in \Pi_N$, $\mathcal{P} \neq \mathcal{Q}$, $S \subset N$, $S \in \mathcal{P}$, $S \in \mathcal{Q}$. Other games in partition function form are *improper*. For a proper game $\Gamma$, we may define a game in coalition function form by $N, v^{\dagger}(S) = v(\mathcal{P},S)$ for some $\mathcal{P} \ni S \in \mathcal{P}$. The game in coalition form generated in this way will be said to be the game in coalition function form *proper to* $\Gamma$.

If $\mathcal{P}$ is a partition and $\mathcal{Q} = \{B_1, \ldots, B_s, \varnothing\}$ is a partition and $\forall i = 1, \ldots, s, \exists k \in \{1, \ldots, r\} \ni B_i \subset C_k \in \mathcal{P}$, then $\mathcal{Q}$ is said to be a *refinement* of $\mathcal{P}$. *Remark*: Formally, each partition is a refinement of itself. If $\mathcal{P} \neq \mathcal{Q}$ then $\mathcal{Q}$ is said to be a proper refinement.

The *fine* partition is $\mathcal{F} = \{\{a_1\}, \{a_2\}, \ldots, \{a_n\}\}$. Trivial lemma: $\forall \mathcal{P} \in \Pi_N$, $\mathcal{F}$ is a refinement of $\mathcal{P}$.

In the partition $\mathcal{G} = \{N\}$, N, the only member set, is the grand coalition. Trivial lemma: any partition $\mathcal{P}$ is a refinement of $\mathcal{G}$.

For any $\mathcal{P} \in \Pi_N$, $\mathcal{P}'(j) = \{C_1, C_2, \ldots, C_{\mathcal{P}}(a_j) \backslash a_j, \{a_j\}, \ldots, C_r, \varnothing\}$. Trivial

lemma: $\mathcal{P}'(j)$ is a refinement of $\mathcal{P}$. We will refer to $\mathcal{P}'(j)$ as the first refinement of $\mathcal{P}$ with respect to $j$. For any $\mathcal{P} \in \Pi_N$ and $S \notin \mathcal{P}$, $\mathcal{P}'_S = \{C | \exists B \in \mathcal{P} \ni C = B \backslash S\} \cup \{S\}$. $\mathcal{P}'_S$ will be called the residual partition of $\mathcal{P}$ with respect to S. Trivial lemma: If and only if $S \subset B \in \mathcal{P}$, $\mathcal{P}'_S$ is a refinement of $\mathcal{P}$. *Remarks*: In that case $\mathcal{P}'_S$ can be referred to as the residual refinement of $\mathcal{P}$ with respect to S. If the group S forms a new coalition then $\mathcal{P}'_S$ is the *immediate* result. If a single individual $j$ considers withdrawing to go it alone as a singleton coalition, this would lead to the first refinement $\mathcal{P}'(j)$ as the *immediate* result.

For $\mathcal{P} \in \Pi_N$, $S \in \mathcal{P}$, a refinement $\mathcal{Q}$ is said to be *granular* with respect to S, $\mathcal{P}$, iff $\forall B \in \mathcal{Q}$, either $B = S$ or $\exists C \in \mathcal{P}$, $C \neq S$, $\ni B \subset C$. Trivial lemma: $\forall C \in \mathcal{P}$, $C \neq S$, $\mathcal{Q}$ granular with respect to S, $\mathcal{P}$, $\exists \mathcal{B} = \{B_j\} \subset \mathcal{Q}$, $\ni C = \bigcup_{B_j \in \mathcal{B}} B_j$. *Remark*: The trivial lemma in this paragraph tells us that each coalition in $\mathcal{P}$ can be constructed of unions of sets in $\mathcal{Q}$.

For $\mathcal{P} \in \Pi_N$, $S \in \mathcal{P}$, a refinement $\mathcal{Q}$ is said to be *particulate* with respect to S, $\mathcal{P}$, iff $\forall B \in \mathcal{Q}$, either $B \subset S$ or $\exists C \in \mathcal{P}$, $C \neq S$, $\ni B = C$.

As with games in coalition function form we will be interested in stable imputations for games in partition function form. For a game $\Gamma = \{N, v(\mathcal{P}, C_i)\}$, an imputation $\mathbf{x} = \{x_1, \ldots, x_N\}$ is a vector of payments to the N players, and $x_S$ is the sum of payments to the members of set $S \subset N$. For partition $\mathcal{P}$ an imputation $\mathbf{x}$ is *admissible* iff $\forall S \in \mathcal{P}$, $\sum_{i \in S} x_i = x_S = v(\mathcal{P}, S)$. *Remark*: The admissibility condition excludes side payments from one coalition to another.

## 13.2 SUPERADDITIVITY

Intuitively, a game is superadditive if the value of a merged coalition is no less than the sum of the values of the coalitions merged to create it. Let $\Gamma = \{N, v(\mathcal{P}, C_i)\}$. Suppose that, for any $\mathcal{P} \in \Pi_N$, $\forall S \in \mathcal{P}$ if $\mathcal{Q}$ is a refinement of $\mathcal{P}$ and is particulate with respect to S, then for all $C \in \mathcal{P}$, $v(\mathcal{P}, C) \geq \sum_{B \in \mathcal{Q}\, B \subset C} v(\mathcal{Q}, B)$. Then $\Gamma$ is superadditive.

*Remark*: Using the concept of particularity excludes the following kind of possibility. Consider a four-person game with negative externalities. The players are $a$, $b$, $c$, $d$. For the fine partition the coalition values are 2, 2, 2, 2. If the first two singleton coalitions merge, the value of $\{a, b\}$ is 5 so long as $c$ and $d$ continue as singletons, but if $\{c, d\}$ is formed, the value of $\{a, b\}$ then is 3. Nevertheless, this might be a superadditive game. Argument A applied to this example would be as follows: $\{a, b\}$ can do no worse than $\{a\}$ and $\{b\}$ separately, *when* $\{c\}$ *and* $\{d\}$ *continue as singletons,* because $a$ and $b$ as a coalition can adopt the same strategies they played against $c$ and $d$ as singletons and thus *obtain the same total payoff.* However, if there are negative

externalities from $\{c, d\}$ to $\{a, b\}$, the same strategies may not result in the same payoff after $\{c, d\}$ has formed. However, $\{a, b\}\{c, d\}$ is *not* particulate with respect to $\{a, b\}\{c\}\{d\}$ and the coalition $\{a, b\}$. This definition can be supported by the "plausible" argument for superadditivity, while without the restriction to particulate partitions, that would not be so.

If the game is not superadditive then the grand coalition may not be efficient. We might define an efficient partition following Aumann and Dreze (1974). First, if $\Gamma = \{N, v(\mathcal{P}, C_i)\}$ is a game in partition function form, let $\Gamma^* = \{N, v^*(\mathcal{P}, C_i)\}$ be the *superadditive cover* of $\Gamma$. The superadditive cover is defined as follows: Let $\mathcal{H}$ be the set of all refinements of $\mathcal{P}$ that are particulate with respect to S. (Recall, trivially $\mathcal{P} \in \mathcal{H}$.) Then

$$v^*(\mathcal{P}, C_i) = \underset{\mathcal{Q} \in \mathcal{H}}{MAX} \sum_{\substack{B \in \mathcal{Q} \\ B \subset C_i}} v(\mathcal{Q}, B).$$

That is, the value of an embedded coalition in the superadditive cover is the maximum over all refinements of the partition of the sum of the values of the subsets of that coalition in the original game, given that the nonmembers of that coalition do not reorganize themselves into another partition. Note that the superadditive cover is itself superadditive. A partition $\mathcal{Q}$ is a-efficient if $\sum_{B \in \mathcal{Q}} v(B) = v^*(N)$. That is, an a-efficient partition generates the maximum total value, summed over all its coalitions, that the game admits of.

The more customary definition of efficiency in economics is Pareto optimality. Let us say that $\mathcal{P}$ is p-efficient if $\exists \mathbf{x}$, an imputation admissible for $\mathcal{P}$, $\ni \forall \mathcal{Q} \in \Pi_N$, and $\forall$ imputations $\mathbf{y}$ admissible for $\mathcal{Q}$, $y_i > x_i \Rightarrow \exists j \in \{1, \ldots, n\} \ni y_j < x_j$. *Remark*: That is, relative to $\mathbf{x}$, any imputation admissible for any partition that makes one player better off will make some other player worse off than his payoff at $\mathbf{x}$. Then a partition is considered p-efficient if it supports at least one Pareto-optimal imputation.

*Theorem ii. b.1*: If $\Gamma$ is superadditive then the grand coalition $\mathcal{G}$ is p-efficient. *Proof*: Suppose the contrary. Then for any $\mathbf{x}$ admissible for the grand coalition, $\exists \mathcal{Q} \in \Pi_N$, $\exists \mathbf{y}$ admissible for $\mathcal{Q}$, with $y_j \geq x_j$ for all players $j$, and moreover $\exists i \ni y_i > x_i$. Therefore

$$\sum_N x_i = v(\{N\}, N\} < \sum_N y_i = \sum_{S \in \mathcal{Q}} \sum_{i \in S} y_i = \sum_{S \in \mathcal{Q}} v(\mathcal{Q}, S);$$

this, however, contradicts superadditivity.

*Theorem ii. b.2*: If partition $\mathcal{P}$ is a-efficient then it is p-efficient. *Proof*: Again, suppose the contrary, that $\mathcal{P}$ is not p-efficient. That means that

for any **x** admissible for $\mathcal{P}$, there exist a partition $\mathcal{Q}$ and an imputation **y** admissible for $\mathcal{Q}$ such that $x_j \leq y_j$, $\forall j \in \{1, \ldots, n\}$ and $\exists i \ni x_i < y_i$. Let $i \in C \in \mathcal{Q}$. Then $x_C < y_C$. For $B \neq C$, $B \in \mathcal{Q}$, $x_B \leq y_B$. It follows that

$$\sum_{S \in P} v(P, S) = \sum_{S \in P} x_S < \sum_{B \in Q} x_B \leq MAX \sum_{\substack{C \in R \\ R \in \Pi_N}} x_c.$$

It follows that $\mathcal{P}$ is not a-efficient.

However, the converse is not so. Consider the five-person game in partition function form with $\mathcal{P} = \{a, b, c\}\{d, e\}$ yielding values 30, 20 and $\mathcal{Q} = \{a, b, c\}\{d\}\{e\}$ yielding values 30,30,1. All other imbedded coalitions have values of zero. $\mathcal{P}$ supports the imputation 10, 10, 10, 10, 10. Any other imputation admissible for $\mathcal{P}$ will make some player worse off, and any imputation admissible for $\mathcal{Q}$ will make E worse off. Therefore $\mathcal{P}$ is p-efficient. However, the superadditive cover of this game assigns 30, 31 to $\mathcal{P}$ and 61 to the grand coalition, so in this game only $\mathcal{Q}$ is a-efficient.

## 13.3 STABILITY

One of our objectives is to draw conclusions about the partitions more or less likely to form, that is, the problem of endogenous coalitions. In this section we are concerned with the stability of a partition and an associated imputation in the face of tendencies of groups to seek profit by forming new coalitions. Accordingly, A *candidate solution* to $\Gamma$ is a partition $\mathcal{P}$ and an imputation **x** that is admissible for $\mathcal{P}$. A *deviation from* a candidate solution ($\mathcal{P}$, **x**) is a set of players $S \notin \mathcal{P}$. The candidate solution is *disrupted* if the group withdraws from the coalitions its members participate with (in the partition $\mathcal{P}$) and precipitate a new partition by forming a new coalition.

### 13.3.1 Naive Stability

Here is a relatively simple adaptation of the idea of the core to the game in partition function form. Let $\mathcal{P}$, **x** be a candidate solution and S a deviation. If $v(\mathcal{P}'_S, S) - \mathbf{x}_S > 0$, then $\mathcal{P}$, **x** is (naively) *disrupted* by S. Equivalently, a disruption of a candidate solution ($\mathcal{P}$, **x**) is a set of players $S \notin \mathcal{P}$ and an imputation **y** admissible for $\mathcal{Q} = \mathcal{P}'_S$ such that

iii.a.1. $\forall i \in S, y_i \geq x_i$
iii.a.2. $\exists i \in S \ni y_i > x_i$

If a disruption of $\mathcal{P}$, $\mathbf{x}$ exists, then we will say that $\mathcal{P}$, $\mathbf{x}$ is *naively unstable*, while otherwise it is *stable*. The set of (naively) stable candidate solutions will be called the naive core or n-core.

*Remark*. It is sufficient that $v(\mathcal{P}'_S, S) > \mathbf{x}_S$, since a disrupting imputation can then easily be constructed. Note that if $\mathbf{x}$ is admissible for $\mathcal{P}$ and $S \in \mathcal{P}$ then $v(\mathcal{P}'_S, S) - \mathbf{x}_S$ must identically be zero.

*Remarks*: The naive core comprises a set of partitions and admissible imputations that are stable in the sense that no group (thinking naively) has any incentive to destroy the partition by organizing among themselves. That is, let $\mathcal{P} \in \Pi_N$, $\mathbf{x}$ an imputation admissible for $\mathcal{P}$, $S \subset N$, $S \notin \mathcal{P}$, $v(\mathcal{P}'_S, S) > \mathbf{x}_S$. Then S have an incentive to disrupt $\mathcal{P}$, and so $\mathcal{P}$, $\mathbf{x}$ is not in the naive core. However, the shift of S to form $\mathcal{P}'_S$ might lead yet another group $T \subset N\backslash S$, $T \notin \mathcal{P}'_S$ to form a coalition, transforming the partition from $\mathcal{P}'_S$ to $\mathcal{Q}$, with $v(\mathcal{Q}, S) < \mathbf{x}_S$. In that case, at least some of group S will unavoidably be worse off, and if the members of S have sufficient foresight they will restrain themselves from forming a separate coalition and so disrupting $\mathcal{P}$; thus $\mathcal{P}$, $\mathbf{x}$ is stable even though it is not in the naive core. Alternatively, it might be the case that $v(\mathcal{P}'_S, S) < \mathbf{x}_S$, but $\mathcal{P}'_S$ itself is unstable and likely to lead to a reorganization of N\S that would leave the members of S better off. Then $\mathcal{P}$, $\mathbf{x}$ is unstable even though it is in the naive core.

### 13.3.2 A Successor Function

The naive approach supplies a value for a deviation, but the value may not be consistent with farsighted rationality. The difficulty, as we have seen, is that partition $\mathcal{P}'_S$ may not actually follow if the deviation by S occurs, and in deciding whether or not to carry out their deviation, the group S will have to form a judgment as to what the consequences of their move will be, given the opportunities for a residual group to reorganize themselves. To capture this judgment, we might postulate a successor function,[1] $\mathcal{R}$, such that for $\mathcal{P}$, $S \notin \mathcal{P}$, the deviation by S results in the formation of partition $\mathcal{Q} = \mathcal{R}(\mathcal{P}, S)$. However, some care needs to be taken to assure that this function is consistent with the structure of the game itself and with our assumptions of agent rationality, whatever they may be.

Consider $\mathcal{P}$, $S \notin \mathcal{P}$, and a sequence of embedded coalitions ${}^1S \in {}^1\mathcal{Q}$, ${}^2S \in {}^2\mathcal{Q}, \ldots, {}^mS \in {}^m\mathcal{Q}$, and denoting S as ${}^1S$, ${}^1\mathcal{Q} = P'_S$, ${}^{k+1}S \subset N\backslash(\bigcup_{i=1}^{k} {}^iS)$, and $N\backslash(\bigcup_{i=1}^{m} {}^iS) = \varnothing$. This is a sequence of refinements of $\mathcal{P}$ granular with respect to S, and every such sequence of refinements will correspond to one or more sequences ${}^kS$. Note that $m \leq \nu = |N\backslash S| + 1$, since even if ${}^iS$ is a singleton for all $i > 1$, $\nu$ adjustments will exhaust the residual set N\S;

with larger sets ${}^{i}$S the number of steps will be still fewer. However, not all such sequences will occur, if adjustments are rational. Suppose that $\forall\, {}^{k}\mathbf{z}$ admissible for ${}^{k}$Q,

iii.b.1. $\forall k > 1\ k < m$, v$({}^{m}\mathcal{Q}, {}^{k}S) > {}^{k-1}$z$_{k\mathrm{S}}$. That is, each set ${}^{k}$S deviates from ${}^{k-1}\mathcal{Q}$ in the expectation of ultimately doing better than it would do at ${}^{k-1}\mathcal{Q}$.

iii.b.2. Let U = N\($\bigcup_{i=0}^{m} {}^{i}S$). Either U = $\varnothing$ or $\forall$T $\subset$ $\mathcal{U}$, $\forall$k > m, v$({}^{m}\mathcal{Q}$, T) > ${}^{k}\mathbf{z}_{\mathrm{T}}$.

That is, no subset of the remaining residual group has any motive to disrupt ${}^{m}\mathcal{Q}$. Then ${}^{m}\mathcal{Q}$ is a $\xi$-*stable outcome* for $\mathcal{P}$, S. Let $\mathcal{M}(\mathcal{P}, \mathrm{S})$ denote the set of all $\xi$-stable outcomes for $\mathcal{P}$,S.

Now suppose that m = 1, that is, no group in the residual set has any motive to disrupt $\mathcal{P}'_{\mathrm{S}}$. Then conditions 1 and 2 are trivially satisfied. It follows that $\mathcal{M}(\mathcal{P}, \mathrm{S})$ cannot be null. On the other hand, there may be more than one $\xi$-stable outcome, that is, more than one sequence satisfying iii.b.1,2, as each step i might allow of two or more sets ${}^{i+1}$S. (An example will be found in the public good production game at Chapter 14, Section 14.3.) It may be that some of the stable outcomes can be ruled out by further rational considerations. In general, a rational agent would form a judgment, by means of some Bayesian or other procedure, as to which of the stable outcomes is most likely. With sufficient information of this kind we might define a rational successor function $\mathcal{R}(\mathcal{P}, \mathrm{S}) = \mathcal{Q}$, with $\mathcal{Q}$ the (expected) successor[2] in case of a deviation of S from $\mathcal{P}$, followed by all rational adjustments by the residual group. In general, however, we will not be able to characterize a rational successor function, as it is likely to depend on information not available in the description of the game but dependent on the particular circumstances of the application, time and place. Following some aspects of Koczy's (2007) work, however, we may set some limits to the range of rational successor functions by considering optimistic and pessimistic cases.

### 13.3.3 A Hypothetical Extension of the Core

For now, take the successor function as given, and let $\mathcal{R}(\mathcal{P}, \mathrm{S}) = \mathcal{Q}$. We define the excess for $\mathcal{P}$, S, and $\mathbf{x}$ as e$(\mathcal{P}, \mathrm{S}, \mathbf{x})$ = v$(\mathcal{Q}, \mathrm{S}) - \mathbf{x}_{\mathrm{S}}$. This construct will play a key role in the remainder of the chapter.

*Remark*: Schmeidler defined the excess in his paper that introduced the concept of the nucleolus, and Aumann and Dreze showed that the excess could be used in the derivation of the core and other cooperative solution concepts, not including the Shapley value. However, most derivations of

the core do not use the excess function. Schmeidler defined the excess function for games in coalition function form, so that they might be ambiguous for partition function games. The successor function resolves this ambiguity, at the cost that there may be a family of excess measures if the rational successor function is not uniquely identified.

We will then define the core as the set of candidate solutions $\mathcal{P}$, $\mathbf{x}$ $\ni \forall S \notin \mathcal{P}, e(\mathcal{P}, S, \mathbf{x}) \leq 0$.

*Lemma iii. c. 1*: if $\Gamma = N, v(\mathcal{P}, S)$ is a proper game and $\Gamma^{\dagger} = N, v(S)$ is the game in coalition function form proper to it, and $\mathcal{P}$,$\mathbf{x}$ is an element of the core of $\Gamma$, then x is an element of the coalition structure core of $\Gamma^{\dagger}$ for coalition structure $\mathcal{P}$. *Proof*: For any S, $e(\mathcal{P}, S, \mathbf{x}) \leq 0$. Further, $v(S) = v(\mathcal{Q}, S)$ for any $\mathcal{Q}$ such that $S \in \mathcal{Q}$ and therefore for $\mathcal{Q} = \mathcal{R}(\mathcal{P},S)$ in particular. Thus $e(S, \mathbf{x}) = v(S) - \mathbf{x}_S = v(\mathcal{Q},S) - \mathbf{x}_S = e(\mathcal{P}, S, \mathbf{x}) \leq 0$, so that $\mathbf{x}$ is in the coalition structure core.

*Remark*: Transparent as this lemma is, it is important in that it establishes that the core as defined here is a generalization of the coalition structure core, which in turn generalizes the core to games that may not be superadditive.

### 13.3.4 Optimism and Pessimism

We can set some limits to the range of possible successors and core-like stable sets. (The influence of Koczy and of Aumann and Maschler (1964) will be evident here). Define

iii.d.1. $\mathcal{R}^{-}(\mathcal{P}, S) = \mathcal{Q}^{-} = \underset{\mathcal{Q} \in \mathcal{M}(\mathcal{P},S)}{\operatorname{argmax}} v(\mathcal{Q}, S)$

iii.d.2. $e^{-}(\mathcal{P}, S, \mathbf{x}) = \underset{\mathcal{Q} \in \mathcal{M}(\mathcal{P},S)}{\max} v(\mathcal{Q}, S) - \mathbf{x}_S$

iii.d.3. $\mathcal{R}^{+}(\mathcal{P}, S) = \mathcal{Q}^{+} = \underset{\mathcal{Q} \in \mathcal{M}(\mathcal{P},S)}{\operatorname{argmin}} v(\mathcal{Q}, S)$

iii.d.4. $e^{+}(\mathcal{P}, S, \mathbf{x}) = \underset{\mathcal{Q} \in \mathcal{M}(\mathcal{P},S)}{\min} v(\mathcal{Q}, S) - \mathbf{x}_S$

Here, iii.d.1, 2 characterize an optimistic perspective, selecting the stable outcome that leaves the deviating group S with the best value and iii.d.3, 4 characterize a pessimistic perspective, selecting the outcome that leaves S with the least value. Neither optimism nor pessimism is generally rational. If we define

iii.d.5. $e(\mathcal{P}, S, \mathbf{x}) = v(\mathcal{R}(\mathcal{P}, S), S) - \mathbf{x}_S$

for $\mathcal{R}$ an (unknown) rational successor function, then we will have $e^+(\mathcal{P}, S, \mathbf{x}) \leq e(\mathcal{P}, S, \mathbf{x}) \leq e^-(\mathcal{P}, S, \mathbf{x})$. Note that the minimum is indicated by the plus sign + and the maximum by the minus sign $^-$ because the optimistic perspective leads to a smaller core. This perspective is optimistic in the sense that a potential deviator group is optimistic about the results of their deviation and so are more likely to disrupt the existing partition than they would be in a pessimistic perspective. Put otherwise, by looking to the maximum of the payoff to a deviation, the optimistic perspective minimizes the set of stable partitions.

Now denote the *optimistic core* by $\Xi^-$ and $\mathcal{P}, \mathbf{x} \in \Xi^-$ iff $\forall S \notin \mathcal{P}$, $e^-(\mathcal{P}, S, \mathbf{x}) \leq 0$. Denote the *pessimistic core* by $\Xi^+$ and $\mathcal{P}, \mathbf{x} \in \Xi^+$ iff $\forall S \notin \mathcal{P}, e^+(\mathcal{P}, S, \mathbf{x}) \leq 0$. Denote the *rational core* by $\Xi^+$ and $\mathcal{P}, \mathbf{x} \in \Xi$ iff $\forall S \notin \mathcal{P}, e(\mathcal{P}, S, \mathbf{x}) \leq 0$.

*Theorem iii.d.1*: $\Xi^- \subset \Xi \subset \Xi^+$. This simple theorem follows from the fact that

$$\max_{Q \in \mathcal{M}(\mathcal{P},S)} v(\mathcal{P}, S) - \mathbf{x}_S \geq v(\mathcal{R}(\mathcal{P}, S), S) - \mathbf{x}_S \geq \min_{Q \in \mathcal{M}(\mathcal{P},S)} v(\mathcal{P}, S) - \mathbf{x}_S$$

for a rational successor function[3] $\mathcal{R}$.

Note that the lemma of the previous section establishes that a core based on an arbitrary successor function is an extension of the core for games in coalition function form. This lemma applies in particular to both the optimistic and pessimistic cores: each is an extension of the core for games in coalition function form.

## 13.4 NUCLEOLUS

Among the cooperative solution concepts, many share the shortcomings of the von Neumann-Morgenstern solution set: they may be null, and if they are not null, also not unique. Two exceptions are the Shapley value and the nucleolus. The nucleolus has the further property that if the core is not null, the nucleolus is an element of it (and is also an element of the kernel, a cooperative solution concept that we will not consider here). Aumann and Dreze defined the nucleolus for a coalition structure (in a game in coalition function form), so, for a proper game, the nucleolus can be defined for each partition. In defining the nucleolus, Schmeidler first defines the excess functions and then defines an ordering in terms of the excess functions.

### 13.4.1 Ordering

Schmeidler's ordering is adapted as follows. In all that follows we take the successor function as given and define the excess function accordingly. For each $\mathbf{x}$ and a given $\mathcal{P}$ we define an index, $I_x(S)$ over the set of all subsets of N as follows:

iv.b.1. $I_x(S) = 1$ iff $\forall T \subset N$, $e(\mathcal{P}, S, \mathbf{x}) \geq e(\mathcal{P}, T, \mathbf{x})$ (The first-ranked set for $\mathbf{x}$ is the set whose excess is largest.)

iv.b.2. $I_x(S) < I_x(T)$ iff $e(\mathcal{P}, S, \mathbf{x}) > e(\mathcal{P}, T, \mathbf{x})$ (A set with a greater excess is ranked before one with a lesser excess.)

iv.b.3. $\exists U \subset N, I_{\mathbf{x}}(U) \ni \forall T \subset N, I_{\mathbf{x}}(T) \leq I_{\mathbf{x}}(U) = M$ (There is a last-ranked index.)

iv.b.4. if $n < M$, then $\exists S \subset N \ni I_{\mathbf{x}}(S) = n$. (Every rank from $1 - M$ corresponds to at least one set).

Note that for a tie, $e(\mathcal{P}, S, \mathbf{x}) = e(\mathcal{P}, T, \mathbf{x}) \Leftrightarrow I_x(S) = I_x(T)$, from b. If $k = I_x(S)$ it will sometimes be convenient to write S as $S_k^{\mathbf{x}}$.

*Lemma iv. b.1*: For $\mathbf{x}$, $\mathbf{y}$ admissible for $\mathcal{P}$, $\mathbf{x} \neq \mathbf{y} \Rightarrow \exists S \notin \mathcal{P} \ni e(\mathcal{P}, S, \mathbf{x}) \neq e(\mathcal{P}, S, \mathbf{y})$. *Proof*: $\mathbf{x} \neq \mathbf{y} \Rightarrow \exists i \ni x_i \neq y_i$. Let $T = C_p(i)$. T cannot be a singleton, since if it were, $x_i = v(\mathcal{P}, T) = y_i$. (It follows also that $\mathcal{P} \neq \mathcal{F}$.) Therefore, $S = \{a_i\} \notin \mathcal{P}$ and $e(\mathcal{P}, S, \mathbf{x}) = v(\mathcal{R}(\mathcal{P}, \{a_i\}) - x_i \neq v(\mathcal{R}(\mathcal{P}, \{a_i\}) - y_i = e(\mathcal{P}, S, \mathbf{y})$.

Now given $\mathbf{x}$ and $\mathbf{y}$ admissible for $\mathcal{P}$, suppose that

iv.b.5. $\exists k \ni \forall i < k, e(\mathcal{P}, S_i^x, \mathbf{x}) = e(\mathcal{P}, S_i^y, \mathbf{y})$

iv.b.6. $\exists S \subset N \ni I_y(S) \geq k, e(\mathcal{P}, S_k^x, \mathbf{x}) < e(\mathcal{P}, S, \mathbf{y})$

then $\mathbf{x} P \mathbf{y}$, and conversely. Note that if S* is maximal over all sets with $I_y(S) \geq k$, that is, $S^* = S_y^*$, then it will follow also that $e(\mathcal{P}, S_k^x, \mathbf{x}) < e(\mathcal{P}, S^*, \mathbf{y})$.

*Lemma iv. b.2*: $\sim[(\mathbf{x} P \mathbf{y})$ and $(\mathbf{y} P \mathbf{x})]$. *Proof*: Suppose the contrary and let k be such that either $k = 1$ or $\forall i < k, e(\mathcal{P}, S_i^x, \mathbf{x}) = e(\mathcal{P}, S_i^y, \mathbf{y})$. Then we have both $e(\mathcal{P}, S_k^x, \mathbf{x}) > e(\mathcal{P}, S_k^y, \mathbf{y})$ and $e(\mathcal{P}, S_k^x, \mathbf{x}) < e(\mathcal{P}, S_k^y, \mathbf{y})$, a contradiction.

*Theorem iv. b.3*: $\mathbf{x} P \mathbf{y}$ and $\mathbf{y} P \mathbf{z} \Rightarrow \mathbf{x} P \mathbf{z}$. *Proof*: $\mathbf{x} P \mathbf{y} \Rightarrow \exists k \ni \forall i < k, e(\mathcal{P}, S_i^x, \mathbf{x}) = e(\mathcal{P}, S_i^y, \mathbf{y})$ and $e(\mathcal{P}, S_k^x, \mathbf{x}) < e(\mathcal{P}, S_k^y, \mathbf{y})$; $\mathbf{y} P \mathbf{z} \Rightarrow \exists l \ni \forall i < l, e(\mathcal{P}, S_i^y, \mathbf{y}) = e(\mathcal{P}, S_i^z, \mathbf{z})$ and $e(\mathcal{P}, S_l^y, \mathbf{y}) < e(\mathcal{P}, S_l^z, \mathbf{z})$.

(1) Suppose $k = l$. Then we have $e(\mathcal{P}, S_i^x, \mathbf{x}) = e(\mathcal{P}, S_i^y, \mathbf{y}) = e(\mathcal{P}, S_i^z, \mathbf{z})$ $\forall i < k$, $e(\mathcal{P}, S_k^x, \mathbf{x}) > e(\mathcal{P}, S_k^y, \mathbf{y}) > e(\mathcal{P}, S_k^z, \mathbf{z})$, therefore $\mathbf{x} \text{Þ} \mathbf{z}$.
(2) Suppose $k < l$. Then we have $e(\mathcal{P}, S_i^x, \mathbf{x}) = e(\mathcal{P}, S_i^y, \mathbf{y}) = e(\mathcal{P}, S_i^z, \mathbf{z})$ $\forall i < k$, $e(\mathcal{P}, S_k^x, \mathbf{x}) > e(\mathcal{P}, S_k^y, \mathbf{y}) = e(\mathcal{P}, S_k^x, \mathbf{z})$, therefore $\mathbf{x} \text{Þ} \mathbf{z}$.
(3) Suppose $k > l$. Then we have $e(\mathcal{P}, S_i^x, \mathbf{x}) = e(\mathcal{P}, S_i^y, \mathbf{y}) = e(\mathcal{P}, S_i^z, \mathbf{z})$ $\forall i < l$, $e(\mathcal{P}, S_k^x, \mathbf{x}) = e(\mathcal{P}, S_k^y, \mathbf{y}) > e(\mathcal{P}, S_k^z, \mathbf{z})$, therefore $\mathbf{x} \text{Þ} \mathbf{z}$.

*Theorem iv. b. 4*: Suppose $\mathbf{x} \neq \mathbf{y}$. Then either $\mathbf{x} \text{Þ} \mathbf{y}$ or $\mathbf{y} \text{Þ} \mathbf{x}$. *Proof*: The contrary supposition means that for all i we have $e(\mathcal{P}, S_i^x, \mathbf{x}) = e(\mathcal{P}, S_i^y, \mathbf{y})$.

Clearly $\mathbf{x} \neq \mathbf{y} \Rightarrow \exists j \ni x_j > y_j$. Consider the singleton $\{j\} = B$; $x_j > y_j \Rightarrow e(\mathcal{P}, B, x) < e(\mathcal{P}, B, y)$. Suppose $I_x(B) = m$, $I_y(B) = l$.

(1) If $m = l$, then we immediately have $\mathbf{x} \text{Þ} \mathbf{z}$, contradiction.
(2) Suppose $m < l$. By the definition of $I_x(S)$, $e(\mathcal{P}, S_l^x, \mathbf{x}) < e(\mathcal{P}, B, \mathbf{x}) < e(\mathcal{P}, B, \mathbf{y})$, and with the equality for all $i<l$, we have $\mathbf{x} \text{Þ} \mathbf{y}$, contradiction.
(3) Therefore $m > l$.
(4) Consider $T = C_P(j)$. Further, consider $C = T\backslash B$. By admissibility, if $x_j > y_j$, $x_C < y_C$. Let $m^* = I_x(C)$, $l^* = I_y(C)$. By an argument similar to the above we must have $l^* > m^*$.
(5) Suppose $m^* \geq m$. Therefore $e(\mathcal{P}, B, \mathbf{x}) > e(\mathcal{P}, C, \mathbf{x}) > e(\mathcal{P}, C, \mathbf{y})$. Letting S take the value C and k take the value m, we have $\mathbf{y} \text{Þ} \mathbf{x}$, contradiction.
(6) Suppose $m^* < m$. Therefore $e(\mathcal{P}, C, \mathbf{x}) > e(\mathcal{P}, B, \mathbf{x}) > e(\mathcal{P}, B, \mathbf{y})$. Letting S take the value B and k take the value l, we have $\mathbf{x} \text{Þ} \mathbf{y}$, contradiction.

Since the set of admissible $\mathbf{x}$ is a closed and compact set and the ordering Þ is complete, a maximal element exists and is unique. The nucleolus for $\mathcal{P}$, $\text{nuc}(\mathcal{P})$, is that maximal element.

For a proper game, as noted before, the excess function as defined here reduces to Schmeidler's excess function for a game in coalition function form. Thus the ordering and the nucleolus also will do so. Accordingly,

*Lemma iv.b.5*: If $\Gamma$ is proper and $\Gamma^\dagger$ is the game in coalition function form proper to $\Gamma$, let $\text{nuc}^\dagger(\mathcal{P})$ be the nucleolus for $\Gamma^\dagger$ given $\mathcal{P}$. Then $\text{nuc}(\mathcal{P}) = \text{nuc}^\dagger(\mathcal{P})$.

In all the foregoing, the excess functions have been based on an arbitrary successor function $\mathcal{R}$. If we compute the excess functions consistently with an optimistic successor function, denote the excesses by $e^-(\mathcal{P}, S, \mathbf{x})$, and the

corresponding nucleolus is $\text{nuc}^-(\mathcal{P})$. If we compute the excess functions consistently with a pessimistic successor function, denote the excesses by $e^+(\mathcal{P}, S, \mathbf{x})$, and the corresponding nucleolus is $\text{nuc}^+(\mathcal{P})$.

### 13.4.2 Decentralized Decisions

In a partition with two or more non-null coalitions, the payoff vector will be the joint result of decisions within the separate coalitions – a problem of interactive decisions! It will be helpful to note that the nucleolus can be understood as a Nash equilibrium of an appropriately specified game among the coalitions. For a payoff vector $\mathbf{x}$ let the vector $\mathbf{x}_S$ denote the vector of payoffs to members of group S, and $\mathbf{x}_{N\backslash S}$ denote the payoffs to the rest. Conversely, we will interpret $\{\mathbf{x}_S, \mathbf{x}_{N\backslash S}\}$ as referring to the payoff vector $\mathbf{x}$, with the order appropriately permuted. Let an order $\succ_S$ over $\mathbf{x}_S$ be induced such that $\{\mathbf{x}_S, \mathbf{x}_{N\backslash S}\} \succ \{\mathbf{y}_S, \mathbf{x}_{N\backslash S}\} \Rightarrow \mathbf{x}_S \succ_S \mathbf{y}$. Now consider the induced game in which the strategy set for $C \in \mathcal{P}$ is the set of $\mathbf{x}_C$ and the payoff is the ranking of $\mathbf{x}_C$ according to $\succ_C$. Let $\mathbf{v} = \text{nuc}(\mathcal{P})$. Clearly, any shift from $\mathbf{v}_C$ will reduce the payoff for C; so $\mathbf{v}$ is a Nash equilibrium of the induced game. Conceptually, this will serve to assure us that there is no conflict between the global order defined and the determination of the payoffs by the different coalitions; computationally, we see that there is nothing to be gained by computing the nucleolus separately on a coalition-by-coalition basis.

### 13.4.3 Synthesis

It remains to consider the relationship between the core and the nucleolus. In the theory of games in coalition function form, we may consider the nucleolus in general as a core assignment algorithm; that is, given that the core of a game is non-null, the nucleolus is in the core and may be considered as the unique assignment of values to be predicted within the core. This result can readily be extended for games in partition function form, with the approach proposed in this chapter.

*Theorem iv.d.1*: If $\mathbf{x} = \text{nuc}(\mathcal{P})$ and $\mathbf{y}$, $\mathcal{P}$ is an element of the core, then $\mathbf{x}$, $\mathcal{P}$ is an element of the core. *Proof*: $\mathbf{x} = \text{nuc}(\mathcal{P}) \Rightarrow \mathbf{x} \succ \mathbf{y}$.

Therefore, $\exists\, k, S \ni e(\mathcal{P}, S_k^x, \mathbf{x}) < e(\mathcal{P}, S, \mathbf{y})$. Since $\mathbf{y}$, $\mathcal{P}$ is an element of the core, $e(\mathcal{P}, S, \mathbf{y}) \leq 0$. It follows that $e(\mathcal{P}, S_k^x, \mathbf{x}) < 0$. Now, consider $i < k$. For such a case $e(\mathcal{P}, S_i^x, \mathbf{x}) < e(\mathcal{P}, S_i^y, \mathbf{y}) \leq 0$. Consider $i > k$. For such a case $e(\mathcal{P}, S_i^x, \mathbf{x}) < e(\mathcal{P}, S_k^x, \mathbf{x}) < e(\mathcal{P}, S_k^y, \mathbf{y}) \leq 0$. We have that for every $S \notin \mathcal{P}$, $e(\mathcal{P}, S, \mathbf{x}) \leq 0$, that is, $\mathbf{x}$, $\mathcal{P}$ is an element of the core.

Therefore, the nucleolus is available as a core assignment algorithm for the core concept developed here for games in partition function form. For each of the core concepts, $\Xi^-$, $\Xi$, and $\Xi^+$ there will be a distinct excess function and thus a distinct Schmeidler ordering and nucleolus, but when the excess and nucleolus are computed according to the corresponding successor function, and if the core is not null, the nucleolus will be a member of the core.

## 13.5 A NON-CORE DYNAMIC PROCESS

A pair $\{\mathcal{P}, S\}$ with $S \notin \mathcal{P}$ is ω-unstable if $\mathbf{v} = \text{nuc}^+(\mathcal{P})$ and *both*

v.1.1. $e^+(\mathcal{P}, S, \mathbf{v}) > 0$
v.1.2. Given $\mathcal{Q} = \mathcal{R}^+(\mathcal{P}, S)$, $\mathbf{y} = \text{nuc}^+(\mathcal{Q})$ $\forall T \subset N$, $T \notin \mathcal{Q}$, *either*
  v.1.2.a. $e^+(\mathcal{Q},T, \mathbf{y}) \leq 0$ *or*
  v.1.2.b. for $\mathcal{U} = \mathcal{R}^+(\mathcal{Q},T)$, $\mathbf{z} = \text{nuc}^+(\mathcal{U})$, $z_S > v_S$

We may say that if $\{\mathcal{P}, S\}$ is ω-unstable then a candidate solution $\{\mathcal{P}, \mathbf{x}\}$ is ω-disrupted by S. These conditions reflect the ideas that a group (1) anticipates that bargaining within a newly formed coalition will lead to payments according to the nucleolus, and (2) will not deviate from $\mathcal{P}$, the status quo, unless the deviation and the bargaining lead to an outcome that is itself stable against further, similarly motivated deviations; unless the subsequent deviations leave the original deviant group S better off than they were before the deviation. Consider the particular case of a cycle, that is, $\mathcal{U} = \mathcal{P}$. Then $\mathbf{z} = \mathbf{v}$, so it is not true that $z_S > v_S$ so there will be no deviation from $\mathcal{P}$; $\{\mathcal{P}, S, \mathbf{v}\}$ will not be disrupted if the deviation to S leads to a cycle.

Conversely, $\{\mathcal{P}, S\}$ is ω-stable if it is not ω-unstable; that is, formally, $\{\mathcal{P}, S\}$ is ω-stable if *either*

v.2.1. $e^+(\mathcal{P}, S, \mathbf{v}) \leq 0$ *or*
v.2.2. Given $\mathcal{Q} = \mathcal{R}^+(\mathcal{P}, S)$, $\mathbf{y} = \text{nuc}^+(\mathcal{Q})$ $\exists T \subset N$, $T \notin \mathcal{Q}$, $\ni$ *both*
  v.2.2.a. $e^+(\mathcal{Q}, T, \mathbf{y})>0$ *and*
  v.2.2.b. for $\mathcal{U} = \mathcal{R}^+(\mathcal{Q},T)$, $\mathbf{z} = \text{nuc}^+(\mathcal{U})$, $z_S \leq v_S$

Now we define two stable sets. First Θ:

v. 3. $\{\mathcal{P}, \mathbf{x}\} \in \Theta$ if $\forall S \notin \mathcal{P}$, $\{\mathcal{P}, S\}$ is ω-stable.

Second we define the set Ω: $\{\mathcal{P}, \mathbf{x}\} \in \Omega$ if $\forall S \notin \mathcal{P}$, *either*

v.4.1. $\{\mathcal{P}, S\}$ is ω-stable, *or*
v.4.2. Letting $\mathcal{Q} = \mathcal{R}^+(\mathcal{P}, S), \exists T \ni \{\mathcal{Q}, T\}$ is ω-unstable.

Notice that the membership in $\Theta$ or $\Omega$ does not depend on the value of $\mathbf{x}$. Nevertheless we characterize ω-stability for candidate solutions so that $\Theta$ and $\Omega$ are sets of candidate solutions and are comparable with $\Xi$. Nevertheless we may say without ambiguity simply that a partition $\mathcal{P}$ is ω-stable or ω-unstable, or is an element of $\Theta$ or $\Omega$, and this informal expression will be used in many cases. Clearly and trivially, $\Theta \subset \Omega$. Once again, condition v.4.2. reflects the idea that a group will not deviate to a successor partition that is itself unstable for deviations of the same kind.

*Lemma v.1*: $\Xi^- \subset \Xi^+ \subset \Theta \subset \Omega$. *Proof*: Let $\mathbf{v} = \text{nuc}^+(\mathcal{P})$. $\{P,\mathbf{x}\} \in \Xi^+ \Rightarrow$ by Theorem iv.a.1, that $\{\mathcal{P}, \mathbf{v}\} \in \Xi^+$ and therefore, $\forall S \notin \mathcal{P}$, $e^+(\mathcal{P},S,\mathbf{v}) \leq 0$. Thus, $\{\mathcal{P}, \mathbf{v}\}$ is not ω-disrupted by S; therefore, $\{\mathcal{P}, S\}$ is not ω-unstable. Moreover, by v.4.1, $\{\mathcal{P}, \mathbf{v}\}$ that is ω-stable is an element of $\Omega$. As this is true for all S, $\{\mathcal{P}, \mathbf{x}\} \in \Theta \subset \Omega$. Moreover from theorem iii.d.1, $\Xi^- \subset \Xi^+$.

*Lemma v.2*: $\Omega \neq \varnothing$. *Proof*: Suppose the contrary. Then for any $\{\mathcal{P}, \mathbf{x}\}$ $\exists S$ *both* conditions v.3.1 and v.3.2 must be *false*; so that, letting $\mathbf{y} = \text{nuc}^+(\mathcal{Q})$ $\ni \forall T \notin \mathcal{Q}$, $\{\mathcal{Q}, \mathbf{y}\}$ is *not* ω-disrupted by T. Then consider $\{\mathcal{Q}, \mathbf{y}\}$ as a candidate solution. $\{\mathcal{Q}, \mathbf{y}\} \in \Omega$, contradiction.

*Lemma v.3*: If $\mathbf{v} = \text{nuc}^+(\mathcal{P})$ and for some $\mathbf{x}$ admissible for $\mathcal{P}$, $\{\mathcal{P}, \mathbf{x}\} \in \Theta$, then $\{\mathcal{P}, \mathbf{v}\} \in \Theta$. This follows trivially from the definition of ω-stability.

## 13.6 SUMMARY

In summary, it is true that partition functions are more complex than coalition functions, and there is an overhead cost of notation for dealing with them; and also that the solution concepts lose some of their precision or generality if agents may farsightedly anticipate the consequences of the externalities created by the coalitions they form. However, these difficulties are unavoidable if we are to deal satisfactorily with cases in which externalities are important and reasonable questions about superadditivity can be raised. It is hoped that this chapter contributes toward a satisfactory cooperative game theory applicable to the complex problems of public policy in a complex world.

# NOTES

1. This will recall Luce and Raiffa's (1957) $\psi$-function. However, there are some important differences: in particular the $\psi$-function characterizes sets that will be unable to disrupt a given partition; and is meant in part to capture information on limitations on coalition formation that are in some sense sociological and irrational. However valuable that information might be, it is not an objective of the successor function.
2. If, instead, the rational agent assigns probabilities to the different ${}^{m}\underset{\sim}{Q}$ that might occur, and estimates a mathematical expectation for the value of a new coalition, the discussion that follows would be only slightly complicated.
3. Moreover (in reference to note 2) since the optimistic and pessimistic core set limits for any other rational successor function, however defined, they also set limits for any probability-weighted average of the outcomes of different possible successors.

# 14. Coalitional play

The previous two chapters have focused primarily on a cooperative analysis of games in partition function form, taking the partition functions as givens. This chapter will discuss the determination of partition functions by non-cooperative play among coalitions. Thus we build a link between the cooperative and non-cooperative aspects of interdependent decisions. As Aumann (2003, p. 6) has observed, "those two aspects of game theory are really not two separate disciplines, they are part of the same whole." For the purposes of public policy, though, it is not enough that cooperative and non-cooperative analyses are complementary, as Aumann observes. Rather, we need analyses of given models that are linked, drawing on both cooperative and non-cooperative approaches. This reflects the (often) different roles of cooperative and non-cooperative models in the pragmatic project of public policy, in that it is commonly the non-cooperative models that identify the problems, so that cooperative analysis *of the same examples* is necessary in order to propose solutions. Several examples will be given to illustrate the link and the application of the analysis begun in Chapter 10.

## 14.1 PARTITION FUNCTIONS AND COALITIONAL PLAY

In general, the value of a coalition is determined by the value that it can obtain by its own efforts acting separately. Accordingly (as argued in Chapter 10) the Nash equilibrium, or some appropriate refinement or extension of it, will determine the values of embedded coalitions. That is, the partition function is determined by *coalitional play*.

In the simplest case, the strategies available to a coalition are simply the vectors of pure strategies available to its members. This is the case for aggregative games. To characterize a non-aggregative game we will need the additional information as to the strategy sets available to the various embedded coalitions. Once this information is given the analysis may actually be simpler, if there are relatively few strategies available to each coalition. The next two sections will give a number of examples illustrating non-cooperative play among coalitions in aggregative games, while the

subsequent section will give an example illustrating the similar approach for non-aggregative games. Non-aggregative games need not be superadditive, and we find that the relation of partition functions, indivisibilities, and superadditivity is more complex than it may appear on the surface. We also consider a further complication that may lead to a nonsuperadditive game in partition function form: agents may have preferences to associate themselves, or not to associate themselves, with other particular agents in a coalition. But we will find that this is easily incorporated in a theory of games in partition function form. We will finally consider an example that includes all these complications, drawing on some ideas traditional in economic theory, and then conclude.

## 14.2 SMALL-SCALE EXAMPLES OF AGGREGATIVE GAMES

Three-person games are the smallest games that allow nontrivial coalition structures, so we may rely a good deal on three-person games to illustrate coalitional play. This section considers several such examples, all aggregative games.

### 14.2.1 NIMBY

NIMBY is a three-person by two-strategy game with positive externalities. Consider Table 14.1, which shows Game 14.1, the NIMBY game, in strategic normal form. As one might suppose, NIMBY stands for "Not In My Back Yard," and one purpose of the example is to illustrate the analysis of this important public policy problem. The idea behind the game is that a proposal is made to construct a facility that will provide a public good to the three agents in the game. However, the facility will have to be built at the location of one of the agents and will create sufficient local nuisance

*Table 14.1 Game 14.1: NIMBY*

| Payoffs: *a*,*b*,*c* | | *c* | | | |
|---|---|---|---|---|---|
| | | Accept *b* | | Reject *b* | |
| | | Accept | Reject | Accept | Reject |
| *a* | Accept | 2,2,2 | 2,6,2 | 2,2,6 | 2,6,6 |
| | Reject | 6,2,2 | 6,6,2 | 6,2,6 | 3,3,3 |

*Table 14.2* $\{a,b\}$ *vs* $\{c\}$ *in NIMBY*

| Payoffs: $\{a,b\},\{c\}$ | | $\{c\}$ | |
|---|---|---|---|
| | | Accept | Reject |
| $\{a,b\}$ | Accept, Accept | 4,2 | 4,6 |
| | Accept, Reject | 8,2 | 8,6 |
| | Reject, Accept | 8,2 | 8,6 |
| | Reject, Reject | 12,2 | 6,3 |

so that (in a non-cooperative situation) that agent will be worse off despite enjoying the public good. All agents can accept or reject the facility; if more than one accepts, then redundant facilities are built with the local nuisance but without increasing the supply of the public good.

In this game "reject" is a dominant strategy equilibrium although the payoffs are maximized by any one of the cases in which one accepts and two reject. The symmetry of the game again means we need consider only one of the three coalitional games of two versus one. Table 14.2 shows $\{a, b\}$ vs $\{c\}$. We see that there are two Nash equilibria, but they have identical payoffs and indeed refer to essentially similar outcomes in that one member of the coalition $\{a, b\}$ accepts the facility and the other does not. We can now, without difficulty, construct a partition function for the NIMBY game, and it is shown as Table 11.5. The further analysis of NIMBY has been given as examples in the previous chapters.

### 14.2.2 The VPC Game

We now consider an example with negative externalities: Game 14.2, Table 14.3. This game is meant to express something of environmental political economy as some committed greens conceive it. Readers may judge for themselves whether it is descriptive of the "real world." For this example $a$ and $b$ are consumers who would be better off on the whole (perhaps for health reasons, or perhaps just because they prefer it that way) if the environment were conserved by every agent in the game, but who may nevertheless face a social dilemma in the following way: given that the environment is polluted, each is better off making their own small contribution to the pollution than not, as the addition to the level of pollution already in existence is minor compared to the convenience of exploiting. The third player, player $c$, is a VPC (Villainous Polluting Corporation) whose profits are always greater when there is more exploitation of the environment, whether the profits arise from dumping its own effluent or from selling plastic bags to consumers. At the Nash equilibrium everyone

*Table 14.3 Game 14.2: the VPC game*

| Payoffs: *a*,*b*,*c* | | *c* | | | |
|---|---|---|---|---|---|
| | | Exploit *b* | | Conserve *b* | |
| | | Exploit | Conserve | Exploit | Conserve |
| *a* | Exploit | 2,2,10 | 2,1,7 | 6,6,4 | 6,7,2 |
| | Conserve | 1,2,7 | 1,1,5 | 7,6,2 | 9,9,0 |

*Table 14.4 {a,b} versus {c} in the VPC game*

| Payoffs: {*a*,*b*},{*c*} | | {*c*} | |
|---|---|---|---|
| | | x | c |
| {*a*,*b*} | xx | 4,10 | 12,4 |
| | xc | 3,7 | 13,2 |
| | cx | 3,7 | 13,2 |
| | cc | 2,5 | 18,0 |

*Table 14.5 {a,c} versus {b} in the VPC game*

| Payoffs: {*a*,*c*},{*b*} | | *b* | |
|---|---|---|---|
| | | x | c |
| *a*,*c* | xx | 12,2 | 9,1 |
| | xc | 10,6 | 8,7 |
| | cx | 8,2 | 6,1 |
| | cc | 9,6 | 9,9 |

exploits. The cooperative result at the lower right is not attainable because "exploit" is a dominant strategy for the VPC.

Now suppose {*a*, *b*} form a coalition. Coalitional play for this case is shown in Table 14.4. Their coalition does no good, since the Nash equilibrium once again reproduces the non-cooperative equilibrium xxx – everyone exploits. Now suppose {*a*, *c*} form a coalition. The coalitional play is as shown in Table 14.5. Here we see two Nash equilibria, xxx and ccc. This is, of course, no surprise. Coalitional play is non-cooperative play and will involve all of the difficulties of non-cooperative play.

(We might also find games of coalitional play that have only mixed

*Table 14.6 The partition function for game 14.2*

| | Partition | Payoffs |
|---|---|---|
| 1 | $\{a,b,c\}$ | 18 |
| 2 | $\{a,b\},\{c\}$ | 4,10 |
| 3 | $\{a,c\},\{b\}$ | 10.5,5.5 |
| 4 | $\{b,c\},\{a\}$ | 10.5,5.5 |
| 5 | $\{a\},\{b\},\{c\}$ | 2,2,10 |

strategy equilibria. So long as the equilibrium is unique, this will present no new difficulties. No such example is given here. A case such as the one in this example is further complicated, moreover, by the existence of still more equilibria in mixed strategies; however, nothing will be added to the discussion by the enumeration of these. In what follows, mixed strategies will be ignored.)

How are we to assign the values to imbedded coalitions in a case of this kind, where there are plural nondominated Nash equilibria? In such a case there is a range of correlated strategy equilibria, and the correlated strategy equilibrium that assigns equal probabilities to the nondominated Nash equilibria will have particular salience. Moreover, if we apply the "principle of insufficient reason," equal probabilities apply to the case in which the players are "totally ignorant" as to which of the Nash equilibria will occur. Accordingly, we adopt the expected values from the equiprobable correlated equilibrium, 10.5 for $\{a,c\}$ and 5.5 for $\{b\}$, as the coalition values for this example. The resulting partition function is shown as Table 14.6.

The VPC game is not proper, symmetrical, nor brief. Consider the grand coalition, coalition 1. We find $\mathcal{R}(1, \{a,b\}) = 2$, $\mathcal{R}(1, \{a,c\}) = 3$, $\mathcal{R}(1, \{b,c\}) = 4$, $\mathcal{R}(1, \{c\}) = 2$; but $\mathcal{R}(1,\{a\}) = 5$, since the residual $\{b,c\}$ can benefit by dissolving, and similarly $\mathcal{R}(1,\{b\}) = 5$. Using these observations we find that the grand coalition is $\xi$-stable provided $x_a>2$, $x_b>2$, $x_c>10$, and these inequalities are satisfied for a continuum of imputations, including in particular the nucleolus for the grand coalition, 3.33, 3.33, 11.33. Moreover, all other partitions will be $\xi$-disrupted by a deviation to the grand coalition. We see that an agreement for a grand coalition can be highly stable and attractive, but requires that the VPC be paid the majority of the benefit from the common action, thanks to its strategic, unsymmetrical position in the game.

These simple games illustrate the derivation of the partition function from non-cooperative play among the coalitions in each potential partition, and the difficulties that may arise. They also capture some

fundamental issues that may arise in public policy. As we have seen, the NIMBY game suggests that in public goods provision with a "not in my back yard" motive, compensation payments are crucial for a cooperative agreement, no highly stable agreements exist, and the relatively stable agreements that do exist are likely to include free-riders. The contrast with the VPC game shows that an unsymmetrical game with negative externalities can nevertheless have a solution that is stable in the core sense. Each example must be taken in its own terms.

## 14.3 A FIVE-PERSON GAME OF PUBLIC GOODS PRODUCTION

Throughout the book, games of public goods provision have been used to illustrate concepts of cooperative games, generally three-person games. In this case, it will be helpful to illustrate at least something of the increase in complexity that arises from a larger game. Accordingly, we consider a five-person symmetrical public goods game. Although there are 52 distinct partitions for a five-person game, when the game is symmetrical, we need only consider seven *forms* of partitions. Suppose that agents begin the game with wealth of 5, that it costs 7 to produce a unit of the public good, and that the public benefit per capita is 3. Because of decreasing returns, each agent can produce at most one unit of the public good.

To illustrate the coalitional play in this case we will consider only one form of partition: $\mathcal{P}=\{\{i,j,k\},\{l,m\}\}$. The three-person coalition $\{i,j,k\}$ has four joint strategies: produce none, 1, 2, or 3 units of the public good. The two-person coalition $\{l,m\}$ has three: produce none, 1, or 2 units. The game in strategic normal form is Game 14.3, Table 14.7.

The payoffs for Table 14.7 are computed from the parameters given above. Inspection of the table quickly assures us that for the three-person coalition, production of the maximum of 3 units is a dominant strategy, whereas for the two-person coalition, to produce nothing is

*Table 14.7 Game 14.3, a five-person game of public good production*

| Payoffs: $\{i,j,k\},\{l,m\}$ | | $\{l,m\}$ | | |
|---|---|---|---|---|
| | | 2 | 1 | 0 |
| $\{i,j,k\}$ | 3 | 39,26 | 30,27 | 21,28 |
| | 2 | 37,20 | 28,21 | 19,22 |
| | 1 | 35,14 | 26,15 | 17,16 |
| | 0 | 33,8 | 24,9 | 15,10 |

*Table 14.8 A partition function for game 14.3*

| | | |
|---|---|---|
| 1 | $\{i,j,k,l,m\}$ | 65 |
| 2 | $\{i,j,k,l\}\{m\}$ | 40,17 |
| 3 | $\{i,j,k\}\{l,m\}$ | 21,28 |
| 4 | $\{i,j,k\}\{l\}\{m\}$ | 21,14,14 |
| 5 | $\{i,j\}\{k,l\}\{m\}$ | 10,10,5 |
| 6 | $\{i,j\}\{k\}\{l\}\{m\}$ | 10,5,5,5 |
| 7 | $\{i\}\{j\}\{k\}\{l\}\{m\}$ | 5,5,5,5,5 |

a dominant strategy. The non-cooperative equilibrium is at the upper right corner where the three-person coalition produces 3 units and the two-person coalition produces nothing. Proceeding in the same way we find that maximal production is always a dominant strategy for a coalition of three or more members and nonproduction is always dominant for smaller coalitions; and this leads to the partition function shown as Table 14.8.

Game 14.3 is symmetrical but not proper nor brief. Consider partition 1, $\mathcal{G}$, and note that at least one agent must be paid $x_m \leq 13$. Consider a deviation with S = $\{m\}$. $\mathcal{P}'_S$ is a partition of form 2, but the residuum, $\{i,j,k,l\}$ is unstable to a further singleton deviation leaving a partition of form 4. Thus $\mathcal{R}(1,\{m\})=4$. By contrast $\mathcal{R}(1,\{l,m\}) = \mathcal{P}'_{\{l,m\}} = 3$. Any larger deviation is unprofitable. Nevertheless, the condition for the grand coalition to be stable against one or two person deviations is that $x_N \geq 5*14 = 70$, which is not admissible. Thus $\mathcal{G}$ can support no candidate solution in the core.

As already noted, partition 2 is unstable to a singleton deviation. $\mathcal{R}(2,\{m\}) = 4$, and partition 2 is stable with respect to such a deviation only if $x_{\{i,j,k,l\}} \geq 4*14 = 56$, which is not admissible. Therefore partitions of form 2 also support no candidate solutions in the core.

For a partition of form 3, consider a deviation with S = $\{i,j\}$. Then $\mathcal{P}'_S$ is a partition of form 5, but the residuum, $\{k\}$, $\{l,m\}$ can benefit by merging with the result that $\mathcal{R}(3,\{i,j\}) = 3$. Suppose S = $\{i\}$. Then $\mathcal{P}'_S$ is again of form 5, but $\mathcal{R}(3,\{i\})$ may be of form 4 or 2, depending whether $\{k, l, m\}$ or $\{j, k, l\, m\}$ form at this stage. Here we have an example with a difference between $\mathcal{R}^+(3,\{i\}) = 14$ and $\mathcal{R}^-(3,\{i\}) = 17$. For this discussion we will focus on the pessimistic case. It follows that the condition for a partition of form 3 to be stable against two-person and one-person deviations is that $x_{\{i,j,k\}} = 3*14 = 42$, which is not admissible. Therefore a partition of form 3 cannot support a candidate solution in the core. A partition of form 4 is similar both in that one-person deviations may be

succeeded either by partition 2 or 4, and in that the partition is unstable in a pessimistic sense against one and two-person deviations.

Partitions of forms 5, 6, and 7 are all unstable for any deviations by three or more agents. Clearly, therefore, the pessimistic core is null (and by inference the rational core must also be null).

Thus we proceed to explore ω-stability for the game. Consider a partition of form 3, and let S = N. Then condition v.1.1. from Chapter 13 is satisfied with any **x**, and in particular with the nucleolus imputation 7,7,7,14,14. However, let T = $\{\ell,m\}$. We have already seen that condition v.1.2.a. cannot be satisfied, and moreover since this deviation brings us by a cycle back to form 3, condition v.1.2.b. cannot be satisfied either. Thus, N does not ω-disrupt a partition of form 3. Let S be $\{i, j, k, \ell\}$. Again, condition v.1.1. will be satisfied. Again, letting T = $\{\ell,m\}$, condition v.1.2 cannot be satisfied. Three-person deviations cannot be profitable so long as $x_\ell$ and $x_m$ are both at least 7. Let S = $\{i,j\}$. Then $\mathcal{R}(3,\{i,j\})$ = $\{\{i,j\},\{k,\ell,m\}\}$ and condition v.1.1. is satisfied; but let T = $\{\ell,m\}$ and, again, we have a cycle. Similarly for deviations of the form $\{i\}$. It follows that a partition of form 3 will be ω-stable. Similarly, profitable deviations from a partition of form 4 prove to be cyclical, and so a partition of form 4 will be ω-stable with many imputations including 7,7,7,14,14.

Now consider the partition of form 1, that is, $\mathcal{G}$. As we have seen, it will be ξ-unstable for deviations of one or two agents, but again these are cyclical, so that $\mathcal{G}$ is ω-stable. Again, a partition of form 2 is ξ-unstable to singleton deviations, but such deviations are cyclical, so partitions of this form also are ω-stable.

Now consider $\mathcal{F}$, a partition of form 7. The nucleolus for this unique partition is the only admissible imputation, 5,5,5,5,5. Let S = $\{i,j,k\}$. $\mathcal{R}(\mathcal{P},S)$ will be of form 4. Condition v.1.1 is satisfied for any imputation admissible for $\mathcal{P}$. For deviations that are succeeded by partitions of form 5,6, or 7, condition v.1.2.a. is satisfied, while for deviations that are succeeded by partitions of the form 1, 2, 3, condition v.1.2.b. is satisfied. Thus $\mathcal{F}$ is ω-unstable. Similarly for partitions of the form 6, 7. Therefore Θ comprises candidate solutions with partitions of forms 1,2,3,4.

In the public goods game, as with NIMBY, no solutions are highly stable. For relatively stable solutions, the public good is always produced, but there may be free-riders who profit by being free-riders and solutions with free-riders are no less stable than the grand coalition. When there are free-riders the public good will be produced at less than an efficient level. The benefits and costs (and the privilege of free-rider status) may be unequally distributed in ways that have no functional explanation but are simply the consequences of history or convention. These results can be extended to games with much larger N at the cost of a little algebra. On

the other hand, it should be stressed that this example assumes that public goods production is an aggregative game: and in particular that there are no costs of contracting or enforcement for coalitions to produce public goods. In larger-scale games this assumption may be less credible.

## 14.4 A NON-AGGREGATIVE GAME

Two examples will be given to illustrate coalitional play for a non-aggregative game. In both, indivisibilities play a key role. This section will focus on an example with indivisibilities and negative externalities. Game 14.4 is a five-person symmetrical game of production with overhead costs and externalities that may be abated, but abatement also subject to overhead costs. As before we need consider only seven distinct families of partitions.

Each production coalition may be able to choose between at most two techniques: craft production, which has neither overhead costs nor polluting externalities,[1] and industrial production, which is subject to both, but which has an advantage in greater labor productivity. We will suppose that craft production generates 4 units of product for every member of the coalition but that (because of imperfect recall, especially with regard to effort commitment) craft production is not available to a coalition of three or more agents. For industrial production, gross output per member of the coalition is 10, but it generates a public-good negative externality[2] that reduces the payoff of every agent by 2, including both members and nonmembers of the producing coalition. The coalition has the further choice of abating the externality, but abatement has an additional overhead cost of 7. (As usual, the reader may add an appropriate number of zeros to each number to make the example more "realistic.")

To begin the discussion, suppose (as a best case) that a coalition of $m$ members faces no externalities generated by any other coalition, as, for example, if all other coalitions choose craft production. Then the coalition's payoffs to the three strategies of production are shown by Table 14.9. We see that a singleton coalition can never benefit by choosing industrial production, with its high fixed costs, and a two-person coalition can never benefit by choosing industrial production with abatement of externalities. These are dominated strategies and may be eliminated from consideration. Thus these strategies are for practical purposes not available to these smaller coalitions and the rightmost column lists the strategies available to each coalition according to size.

Admitting this additional information actually simplifies the analysis from this point on. Consider, for example, the fine partition. Since singleton coalitions have only one available strategy, there are no interdependent

*Table 14.9 Game 14.4 payoffs if no other coalition generates externalities*

| N | Craft | No abatement | Abatement | Available strategies |
|---|---|---|---|---|
| 1 | 4 | 1 | −4 | Craft |
| 2 | 8 | 9 | 6 | Craft, no Abatement |
| 3 | | 17 | 16 | Abatement, no abatement |
| 4 | | 25 | 26 | Abatement, no abatement |
| 5 | | 33 | 36 | Abatement, no abatement |

*Table 14.10 Coalitional play in Game 14.4 with a 3 × 2 coalition structure*

| Payoffs: $\{i,j,k\},\{l,m\}$ | | $\{l,m\}$ | |
|---|---|---|---|
| | | Craft | No abatement |
| $\{i,j,k\}$ | No | 17,4 | 11,5 |
| | Abate | 16,8 | 4,9 |

*Table 14.11 Coalitional play in Game 14.4 with a 2 × 2 × 1 coalition structure*

| Payoffs: $\{i,j\},\{k,l\},\{m\}$ | | $kl$ | |
|---|---|---|---|
| | | Craft | No abatement |
| $ij$ | Craft | 8,8,4 | 4,9,2 |
| | No | 9,4,2 | 5,5,0 |

decisions to be made – indeed no decisions – and the payoffs are simply those from craft production. We will find that interdependent decisions occur in only two families of partitions: $\{\{i,j,k\},\{l,m\}\}$, with the three-person coalition choosing between abatement and none, and $\{\{i,j\},\{k,l\},\{m\}\}$, with the two-person coalitions choosing between craft production and industrial production without abatement. Otherwise, a four or five-person coalition will choose pollution-abating industrial production (because the membership is so large that the externality is largely internalized), while a two or three-person coalition facing singletons will choose industrial production without abatement. The two cases of coalitional play are shown by Tables 14.10 and 14.11 and the resulting partition function is shown in Table 14.12. We see that in Tables 14.10 and 14.11, choice of the industrial technology without abatement of pollution is the result of a social dilemma.

*Table 14.12 A partition function for Game 14.4*

| | | |
|---|---|---|
| 1 | $\{i,j,k,l,m\}$ | 36 |
| 2 | $\{i,j,k,l\}\{m\}$ | 26,4 |
| 3 | $\{i,j,k\}\{l,m\}$ | 11,5 |
| 4 | $\{i,j,k\}\{l\}\{m\}$ | 17,2,2 |
| 5 | $\{i,j\}\{k,l\}\{m\}$ | 5,5,0 |
| 6 | $\{i,j\}\{k\}\{l\}\{m\}$ | 9,4,4,4 |
| 7 | $\{i\}\{j\}\{k\}\{l\}\{m\}$ | 4,4,4,4,4 |

Game 14.4 is not proper, brief, aggregative, nor superadditive. In economic terms, it does have "economies of scale." Thus the failure of superadditivity in this case illustrates an important possibility. However, the reconsideration of superadditivity in Chapter 13, relying on the distinction between refinements that are and are not particulate with respect to a particular partition and set, can be illustrated by this game. Consider first partitions of form 3 by comparison with form 7. In 7, the total value of singleton agents $\{l\}$ and $\{m\}$ is 8, whereas in 3 the value of the merged coalition $\{l,m\}$ is 5. This may seem to be a violation of superadditivity, as argument A says "any vector of strategies available to the two coalitions separately is also available to the merged coalition, so that they can do no worse than to adopt the strategies adopted by the two coalitions separately" (Chapter 9), and indeed the strategies of craft production adopted by $\{l\}$ and $\{m\}$ separately are also available to $\{l, m\}$. However, the simultaneous merger of $\{i, j, k\}$ changes the situation, creating negative externalities to $l$ and $m$, and there is no reason to suppose that craft production can generate the payoffs in a partition of form 3 that it does in form 7. Since argument A does not apply, this cannot correctly be thought of as a violation of superadditivity.

Consider instead partitions of form 5 in comparison with form 6. Again, in form 6, the payoffs to singletons $\{k\}$ and $\{l\}$ are each 4, while in form 5, the payoff to $\{k, l\}$ is 5. Since the organization of agents $i$, $j$, $m$ is unchanged, argument A should apply, and so this is a violation of superadditivity. The fact that the organization of agents $i$, $j$, $m$ is unchanged corresponds to the fact that the partition of form 6 is a refinement of 5 which is particulate with respect to S. Consider a partition of form 3 and let S = $\{i,j,k\}$. The refinement $\{\{i\}, \{j\}, \{k\}, \{l,m\}\}$ is particulate with respect to S and is of form 6, so that the total payoffs to $i$, $j$, and $k$ in the refinement is 12, while the payoff to $\{i,j,k\}$ in a partition of form 3 is 11, yet another violation of superadditivity. This result will serve here to illustrate the use of such concepts as

particulate refinements in interpreting such relatively familiar concepts as superadditivity.

Now we explore the ξ-stability of candidate solutions to Game 14.4. First consider $\mathcal{F}$, the partition of form 7. This partition will be disrupted by any deviation to a coalition of two, four or five members, and is ξ-unstable. For a deviation with S = {i,j,k}, $\mathcal{P}'_S$ is of form 4; but the residuum, {k}, {l}, can profit by merging so $\mathcal{R}(\mathcal{P},S)$ is of form 3 and the three-person deviation is not profitable. A partition of form 6 will again be disrupted by deviations to four or five-person coalitions, but not by a three-person deviation for the same reason. In any case it is ξ-unstable. Partitions of forms 3, 4, and 5, are highly unstable, and will also be ξ-disrupted by four and five-person deviations as well as by some one and two-person deviations. Partitions of form 2 are ξ-disrupted by five-person deviations to form the grand coalition. However, $\mathcal{G}$ is ξ-stable with equal payments of 7.2 per member, as no deviation by a smaller group is profitable. The core for this game comprises the grand coalition with a range of payoffs, including the nucleolus payment of 7.2 per member.

The practical conclusion seems to be that, if technologies are as described in this example, consolidation is very likely, very stable, and desirable. This conclusion seems to bear against the application of antitrust policy, which would tend to prohibit the consolidation of an industry as a single grand coalition. The qualification that must be mentioned is that we have not modeled the neighbors or customers of these enterprises, who are likely in fact to bear much of the cost of pollution and also to suffer from higher prices in the case of monopolization. Nevertheless, it does not seem wrong to suggest that the case for antitrust policies against mergers should be qualified to take into account the possibility that competitive conditions would make it more difficult for the firms to adopt feasible means for abatement of negative externalities, while a monopoly could abate the externality at less cost.

## 14.5 COALITIONAL PREFERENCES AND OTHER COSTS OF COALITION

Most work in cooperative game theory has not considered any preferences that agents might have to participate in one coalition and not in another. The principle exception is discussions of "hedonic games," in which only those preferences, and no other aspect of outcomes, play any part in determining coalition formation and payoffs. There is, however, a modest literature in political sociology and political science that addresses these issues. Indeed coalitional preferences may be crucial in reality. As an

example, it appears that the coalition government in Germany after 2005 existed to a considerable extent because of coalitional preferences. First, no one wanted a coalition with the Left party, which includes elements of the former East German Communist Party, and this preference prevented a coalition of the Left with the SDP and the Greens that would have formed a government with a small majority. Second, the Free Democrats preferred not to enter into any coalition with the Social Democrats who – with the Greens – might otherwise have formed a majority. Finally, the Greens preferred not to enter into any coalition with the Right parties, which could otherwise have formed a majority. These constraints left a "grand coalition" of CDU-CSU and Social Democrats as the only possibility for a majority government. The partition function for TU games can capture the implications of coalition preferences rather easily, however.

### 14.5.1 Coalitional Preferences and Transferable Utility

In the spirit of the transferable utility approach, we might express the preferences of individual agents over coalitions by the amount of payoff they would sacrifice or demand to compensate them for entering a coalition with certain other individuals. Suppose, for example, that some individuals prefer not to coalesce with certain other individuals, perhaps because of a difference of skin hue or religious tradition that the culture treats as significant. We could express the intensity of this preference by a numeric penalty that would be deducted from the agent's payoff in case that coalition is formed. Once again consider Game 14.2, and suppose $a$ and $c$ despise one another so much that each will give up a payoff of 5 to avoid any coalition of which the other is a member. Then in effect, the value of any coalition containing both $a$ and $c$ is reduced by 10.

Clearly, in a TU game, there is no need to assign the penalty to particular agents. It makes no difference if $a$ dislikes $c + 7$ and $c$ dislikes $a + 3$. In either case a coalition including $a$, $c$ will need to set aside 10 of the payoffs generated by their joint action to compensate $a$ and $c$ for enduring the company of one another. Rather, the penalty can be deducted from the value assigned to any coalition that includes both $a$ and $c$. With transferable utility, in other words, the hedonic and payoff aspects of coalition formation are additively separable and the partition function that reflects both aspects simultaneously simply sums the two, as shown for this game in Table 14.13. What we notice immediately about this partition function is that it is no longer superadditive. Coalitional preferences provide another reason (along with imperfect recall and indivisibility) why games in partition function form may not be superadditive.

It will be of interest to contrast this game with the unmodified VPC

*Table 14.13 A partition function for Game 14.2 reflecting both strategic coordination and coalitional preferences*

| | Partition | Payoffs |
|---|---|---|
| 1 | $\{a,b,c\}$ | 8 |
| 2 | $\{a,b\},\{c\}$ | 4,10 |
| 3 | $\{a,c\},\{b\}$ | 0.5,5.5 |
| 4 | $\{b,c\},\{a\}$ | 10.5,5.5 |
| 5 | $\{a\},\{b\},\{c\}$ | 2,2,10 |

game. In this modified game, the core comprises lines 2 and 5 with payoffs 2,2,10. The grand coalition is no longer stable, as, for example, a deviation by $\{c\}$ leads to line 2, and $\{a, b\}$ have nothing to gain by dissolving, so that $2 = \mathcal{R}(G, \{c\})$ and $G$ is disrupted. Similarly, line 3 is disrupted by a deviation by$\{c\}$, yielding line 1. If $x_b > 0.5$, then line 4 is disrupted by$\{c\}$, but otherwise it is disrupted by$\{b\}$. For lines 2 and 5, however, deviations leading to other partitions result in losses, and deviations from line 2 to 5, or conversely, make no difference, so none are disruptive. We see that, in this case, coalitional preferences prevent any effective cooperation.

### 14.5.2 "Imperfect Recall," Again

In some previous discussions, we have observed that "imperfect recall" may result in nonsuperadditive game values. In particular, we assumed that some strategies (involving great effort, commitments, or promises to make side payments) cannot be known to some players in the game, and so are not available to coalitions including those agents. But this is a bit absolutist. It might be that the efforts and other strategies could be verified, but only at some cost. This can be captured in a way quite similar to coalitional preferences. For each imbedded coalition, we envision a cost of organizational information equivalent to a certain quantity of payoffs forgone. We then compute the value of the embedded coalition as if the game were aggregative, but deduct the organizational cost just as, in the previous example, we deducted the payoff equivalent of negative coalitional preferences.

Economists in a Marshallean tradition may find it natural to assume that there are costs of coordination in large coalitions. This would be Marshall's case of diseconomies of scale in the firm. This has remained controversial for some economists, but in any case, can be represented for TU games in just the same way. A coalitional cost is thus a flexible and relatively simple way of capturing consequences of imperfect recall,

indivisibilities, and coalitional preferences in a game in coalition function or partition function form, although of course it is more satisfactory if the coalitional costs can themselves be explained, perhaps by an explicit model incorporating imperfect recall, indivisibility, coalitional preferences, and so on. For many purposes, though, we may keep those considerations in the background. For games of transferable utility, the information from a model of imperfect recall, externalities, coalitional preferences, or overhead costs either of production or organization can be captured in a relatively compact and useful form by a partition function, although the partition function will not in general be superadditive. In this sense, the partition function can capture information that the game in normal form does not. For some purposes, the partition function may be the more informative primitive.

## 14.6 A SMITH-CLARK-MARSHALL GAME

A final example, Game 14.6, will illustrate non-aggregative games that illustrate some ideas from the history of economic thought, drawing particularly on ideas from Adam Smith, John Bates Clark, and Alfred Marshall. We begin by modeling production somewhat along the lines of Adam Smith's famous pin factory example (Smith, 1994 [1776], pp. 4–5). Consider a game of four players. We will suppose our four players can choose among ten distinct tasks (strategies): T0, T1, . . ., T9. Any agent who chooses T0 then possesses 5 units of output. In case $i$ chooses T1 and $l$ chooses T2 then $l$ possesses 20 units. In case $i$ chooses T3, $j$ chooses T4, and $l$ chooses T5, then again $l$ possesses 30 units. In case $i$ chooses T6, $j$ chooses T7, $k$ chooses T8, and $l$ chooses T9, then $l$ possesses 45 units. Otherwise all agents possess zero units of output.

Suppose, for a moment, that instead of accruing to $l$, payoffs were to be equally divided among the agents who choose a productive sequence of tasks. Thus, for example, if tasks T6, T7, T8, and T9 were chosen, each player would have 11.25. In that case we would have a coordination game. There would be many solutions, some Pareto-preferable to others. One solution to such a game is to appoint one player as a coordinator – a Clarkian entrepreneur (Clark, 1899).

Instead, though, we have a game in which all players but one either have payoffs of zero or choose T0 for 5. The only Nash equilibrium of a game based on these assumptions will be one in which all agents choose T0, since in any other productive sequence agents $i$, $j$, and $k$ will always be able to improve their outcome by shifting to T0. In order to attain any productive division of labor it will be necessary that a coalition is formed and agents $i$,

$j$, and $k$ are compensated by side payments. Thus, we may naturally treat agent $l$ as the financier and employer of the others. It is not surprising if this agent functions also as Clarkian coordinator.

This schema implies increasing returns to scale in the absence of monitoring costs. However, it is also assumed (1) that with larger total outputs from the four agents, demand price declines in a range from 8 monetary units per unit of output to 6.4. The decreasing price generates a negative (pecuniary) externality from a larger to a smaller or equally-sized coalition. Since marginal cost is zero up to the capacity output and infinity beyond, it will never be profitable even for a monopolist to restrict output in order to raise the price, and the dominant strategy for a coalition will be to adopt the most extensive division of labor it is capable of. (2) Two-person coalitions have their revenues reduced by a 2-unit cost of monitoring effort; and for three-person coalitions the cost is 7 and for four-person coalitions 16. (That is, output is reduced by 2, 7, or 16 units from what it would otherwise be.) This latter is from the idea, traceable to Marshall, that decreasing returns to scale are a consequence of increasing organizational cost for larger business organizations.

The latter assumptions lead to a "long run average cost" curve like the one shown in Figure 14.1, labeled as "with" monitoring costs. Since labor (coalition membership) is the only input in this example, average costs are measured as labor requirements per unit of output. For contrast, the curve "without" shows what the cost curve would be without the monitoring

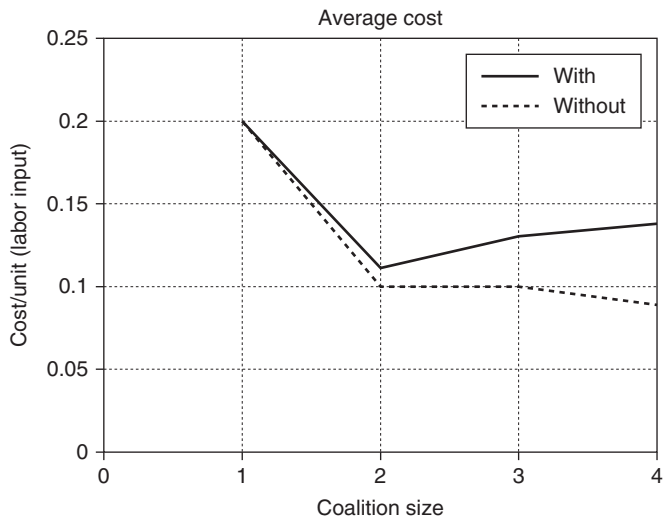


*Figure 14.1* *"Long run average cost curves"*

*Table 14.14 Some data for game 14.6*

| | Partition forms | Net Q | Price | Value |
|---|---|---|---|---|
| 1 | $\{i,j,k,l\}$ | 29 | 7.1 | 205.9 |
| 2 | $\{i,j,k\},\{l\}$ | 28 | 7.2 | 165.6,36 |
| 3 | $\{i,j\},\{k,l\}$ | 36 | 6.4 | 115.2,115.2 |
| 4 | $\{i,j\},\{k\},\{l\}$ | 28 | 7.2 | 129.6,36,36 |
| 5 | $\{i\},\{j\},\{k\},\{l\}$ | 20 | 8.0 | 40,40,40,40 |

costs. As Kaldor (1934) argued, in a Smithian economy, division of labor will lead to increasing returns to scale (if this tendency is not offset by some other consideration). The assumptions altogether lead to Table 14.14, which is, among other things, the partition function for this game.

Now, to explore the $\xi$-stability of this game, consider partition 1. For an admissible imputation, at least two are paid no more than 103. Without loss of generality let those two be $i$, $j$, and let $S = \{i, j\}$. Then $\mathcal{P}'_S$ is of form 3, and the residual, $\{k, l\}$, cannot profitably reorganize so that $\mathcal{R}(1, \{i, j\}) = 3$, and correspondingly $e^+(1, \{i, j\}) \geq 12.2 > 0$. Thus the grand coalition is $\xi$-unstable with any imputation. Consider partitions of form 2. Again, at least two are paid no more than 110.4; letting those two be $i$, $j$, and $S = \{i, j\}$, we will find that $\mathcal{P}'_S$ is of form 4, but the residual, $\{k\}$, $\{l\}$ can profitably merge so that $\mathcal{R}(2,\{i, j\}) = 3$. Thus, again, $e^+(2, \{i, j\}) \geq 4.8 > 0$, and partitions of form 2 are $\xi$-unstable with any imputation. Consider partitions of form 4, and let $S=\{k, l\}$. The residual, $\{i, j\}$, cannot benefit by any reorganization so $\mathcal{R}(4,\{k,l\})=3$ and $e^+(4, \{k, l\}) = 115.2–72 > 0$. It follows that partitions of form 4 are $\xi$-unstable with any imputation. Consider partitions of form 5, and let $S = \{i, j\}$. Since the residual $\{k\}$, $\{l\}$ again can profitably merge, $\mathcal{R}(5,\{k,l\}) = 3$ and $e^+(1, \{i, j\}) = 115.2 - 80 > 0$, so again, partitions of form 5 are $\xi$-unstable with any imputation.

It remains to determine the stability of partitions $\mathcal{P}$ of form 3. (a) Let $S = \{j, k\}$. Then $\mathcal{P}'_S = 4$, but since the residual, $\{i\}$, $\{l\}$ can profitably merge, $\mathcal{R}(3,\{j, k\}) = 3$, and $e^+(3, \{j, k\}) = 115.2 - x_j - x_k$. This will be exactly zero for all $j$, $k$ provided that $x_i = x_j = x_k = x_l$. Therefore, a partition of form 3 will be stable against two-person deviations only where all agents are paid equally. (b) Let $S = \{i, j, k, l\}$. When $\mathcal{P}'_S$ is the grand coalition, $\mathcal{R}(\mathcal{P},S) = \mathcal{P}'_S$. Thus $e^+(3, \{i, j, k, l\}) = 205.9 - 230.4 < 0$, and $\mathcal{P}$ will not be disrupted by a four-person deviation. (c) Let $S = \{i, j, k\}$. Again $\mathcal{R}(\mathcal{P},S)$ is a coalition of form 2. Thus $e^+(3,\{i,j,k\}) = 165.9 - 115.2 - x_k > 0$ only if $x_k < 50.7$. With equal payments of 57.6, a partition of form 3 will not be disrupted by three-person deviations. (d) Let $S = \{l\}$. Then $\mathcal{P}'_S$ is

of form 4, and the residual, $\{\acute{\imath},\acute{\jmath}\},\{\acute{k}\}$ have nothing to gain by reorganization, so $\mathcal{R}(3,\{\acute{l}\}) = 4$. Again, with equal payments of $57.6>36$, partitions of form 3 will not be disrupted by singleton deviations.

The overall conclusion is that the pessimistic core for this game comprises any efficient (form 3) partition with equal payments to all (assumed identical) agents. This is a conclusion that will seem natural and unsurprising to economists in a Marshallean tradition: free entry (as modeled by the core dynamics) results in an efficient organization of the industry. However, a qualification should be mentioned. The result is not robust to what seem to be minor changes in the parameters. These parameters have been carefully chosen to illustrate the Marshallean idea, but it is not at all hard to generate numbers – retaining the "decreasing returns to scale" illustrated by Figure 14.1 – with a null core and others in which larger coalitions may be stable in the same way that the efficient partition is. Thus the example should not be generalized without some caution.

## 14.7 SUMMARY AND RECAPITULATION

The last three chapters have outlined a theory of coalition formation that draws on both cooperative and non-cooperative game theory. Non-cooperative game theory supplies a key diagnosis of a number of problems for public policy. Suppose we begin with an underlying non-cooperative game, such as the public goods production game. For conventional non-cooperative game theory, this is a social dilemma, and the agents play it as individuals and nothing of the public good is produced. But we observe that real rational agents do form coalitions, and we recall (Chapter 10, Section 10.3) that the concept of rationality that underlies the non-cooperative analysis is an incomplete concept of rationality in that rational agents will sometimes form coalitions around commitments that would not be credible in non-cooperative terms.

Accordingly, we suppose that the agents interpose a set of coalitions, a partition, between themselves and the underlying game, so that the game is played non-cooperatively with the coalitions as unitary decision-makers. For this purpose, it is natural to identify what a coalition can "obtain by its own efforts" with a non-cooperative equilibrium among the coalitions: a Nash equilibrium or, in case there are plural Nash equilibria, a symmetrical correlated strategy equilibrium.

The motivation for forming groups can be expressed very briefly by the value assigned to the group, as in a coalition function in the established cooperative game theory reviewed in Chapter 8. However, if we are to take explicit account of externalities, which are crucial for many public policy

issues, we must express this coalitional value as a partition function, in which the value depends not only on the membership of the coalition but also on the organization of the rest of society into coalitions. Accordingly, we derive the partition functions from non-cooperative equilibria in coalitional play, and this chapter has given a number of examples to illustrate this derivation. This determines a partition function.

Within a given partition, some groups may find it rational to initiate reorganizations, which determines the stability of some partitions relative to others. Also, within each coalition, bargaining among the members determines side payments and translates the value of each coalition into a set of net payments to individuals, the nucleolus. The non-cooperative game among the coalitions is imbedded in the cooperative game that determines stability and individual payouts: it is imbedded in the sense that the appropriate analysis is backward induction from the non-cooperative to the cooperative game. In turn the cooperative game is imbedded in a non-cooperative game played by the individual agents that determines their consistent conjecture, that is their common expectation as to the way that their society will be organized into groups. Thus, the cooperative game in this model is encapsulated between two non-cooperative games, and may be described as a model of "encapsulated cooperation."

## NOTES

1. The statement that craft production does not generate polluting externalities is a simplifying *assumption* for this model and not a claim about craft production in the actual world.
2. Baumol and Oates, 1975 "Public Bad" might be a more evocative phrase.

# 15. The government game

This chapter sketches a non-normative theory of the state. For much of history, the vast majority of people have been governed by something we may describe as a state: a body defined on a particular territory that claims a monopoly of predictable violent force among all residents of that territory. Using the game-theoretic approach of this part of the book, such a predictable phenomenon should correspond to a highly stable solution of a game, which in turn would be a stylized description of interactions we may reasonably suppose that most people find themselves engaged in.

While government has often been considered as a grand coalition for the purposes of producing public goods, we will see that this is a somewhat confused conception. We have already seen that the grand coalition is neither highly nor uniquely stable for public goods games in the model of encapsulated cooperation, and that public goods may be provided, though inefficiently, by coalitions that exclude some free-riders. Moreover, there may be many public goods, and no prima facie reason for a unique coalition to produce them all. Indeed, it may be that the assumed identity of the state and the producer of public goods is a source of some of the muddle about the role of the state. Some muddle also arises from the lack of clarity about the purposes of a theory of the state. If the purpose is normative – concerning what the state ought to do – then efficient public goods production is one of a reasonable range of normative goals. But the concept of public goods is too narrow, and the idea we need is not new. It may be found in Hobbes (1968 [1660]).

## 15.1 THE GANG GAME

Any normative public policy presupposes the existence of a state to implement the policy. If the state is anything other than a deus ex machina in this study, then we will need to account for the existence of the state and for the idea that the state might in principle be expected to adopt policies that reflect some normative considerations. It is not clear that the state can be considered a cooperative coalition. As a demographic group, the state comprises residents of a particular territory, and membership in the

state thus is not voluntary. There is a tradition in political philosophy that sees the state as a voluntary association *in principle* and at some level of abstraction, the social contract theory. But even this need not be thought of as a cooperative coalition. As McCain (1992) argued, the Hobbesian variant of social contract theory is consistent with non-cooperative equilibrium (and bounded rationality) among the subjects, and, we recall, presupposes no agreement whatever between the sovereign and the subjects. Instead, however, consider the Gang Game.

The players in this game are a group $n$ of residents in a particular territory. Since $n$ is large, some algebra will be required for the example, but it will be kept as simple as possible. In the absence of any coalitions of two or more players, disorganized violence among the players is general, and property rights are insecure, so the value of each singleton coalition, denoted by $y$, is low. We have Hobbes's "such a war as is of every man against every man," and "the life of man, solitary, poor, nasty, brutish, and short." Without going into formal details, a posture of aggression against others is a dominant strategy in the game among the singleton coalitions.

A coalition in this game takes the form of an agreement (a) not to commit aggression within the group, and (b) to impose a penalty against anyone, whether a member of the group or not, who commits aggression against a member. The penalties are assumed effective enough that the per capita *gross* payoff of each member of the coalition is at least $y + z > y$. However, the coalition will be subject to some degree of imperfect recall, and the enforcement cost that results will be $K$. These enforcement costs reduce the per capita payoff of a coalition below $y + z$. Let the membership of the coalition be denoted by $m \leq n$. The formation of a gang will be profitable if

$$m > \frac{K}{z} \tag{15.1}$$

*The protection provided by such a coalition is not strictly a public good.*[1] By assumption the marginal cost is zero, that is, $K$ is independent of $m$; but it is quite possible for nonmembers to be excluded from protection. Suppose, for example, that $K$ goes to pay the salaries of patrollers who are on the lookout for acts of aggression. Suppose also that in a particular row of houses, number 10 is the home of a member while number 12 is not. The patroller who monitors number 10 can monitor number 12 at no increase of cost, but if he sees that number 12 is burglarized, is not obliged to call for help or try to stop the burglar, and if the burglar is a member of the coalition that pays him, he may be expected to lend a hand in carrying off the loot.

As this example suggests, a coalition may permit its members to commit aggression against nonmembers, while protecting the members against retaliation by the nonmembers. That may explain why this is "the Gang Game." Indeed, a coalition may be able to organize the aggression against nonmembers and exact a tribute from them.

If two or more such coalitions are formed, each will protect its members from aggression or exaction by the other, but the per capita payoffs of all players are reduced by a conflict cost of $h < z$. If only one gang is formed, with a monopoly of anticipated violence, the tribute it can expect from nonmembers is limited by their ability to pay, $y$, and a demand of $y$ will be a dominant strategy for the gang. However, we suppose that the gang will face an overhead cost of $L$ to do so. Thus, exaction will be unprofitable unless

$$m < n - \frac{L}{y} \tag{15.2}$$

We assume also that

$$h < y + z - \frac{K}{n - \dfrac{L}{y}}. \tag{15.3}$$

## 15.2 ANALYSIS: DEFENSIVE COALITIONS

Consider the fine partition $\mathcal{F}$ and a deviation to the grand coalition $\mathcal{G}$. Since this leaves no residual that might reorganize, $\mathcal{G}$ is the only rational successor and the excess is

$$W = n(y + z) - K - ny = nz - K \tag{15.4}$$

which is positive in case

$$z > \frac{K}{n} \tag{15.5}$$

and inequality 15.5. will be assumed from this point on.

Consider a partition of the form $\mathcal{P} = \{\{i, j, \ldots, k\}, \{l\}, \{m\}, \ldots \{n\}\}$, which will be called form 1, with $C = \{i, j, \ldots, k\}$.

*Case 1*: Suppose $m \geq n - (L/y)$ and $m > n - (K/z)$ so that $(n - m) < (K/z)$. The value of C is

$$V_1 = m(y + z) - K \tag{15.6}$$

Consider the fine partition and a deviation to C. There will be no organization of the remainder (nor will it be subject to exaction, since the remainder are too few for this to be profitable), so form 1 is a rational successor.

*Case 2*: Suppose $m \leq n - (K/z)$. Then exaction of tribute from the remainder is profitable and

$$V_1 = z + \frac{ny}{m} - \frac{K + L}{m} \tag{15.7}$$

Consider the fine partition and a deviation to C. This will be profitable in a naive sense if

$$m(y + z) - K - L + (n - m)y > my \tag{15.8}$$

and for a sufficiently large $m$ this clearly will be true. However, the residual can benefit by consolidating as $B = \{\ell, m, \ldots, n\}$, if

$$(n - m)(y + z - h) - K > 0 \tag{15.9}$$

and assumption (15.3) assures that this will be so. Thus a partition of form 2, $\mathcal{Q} = \{\{i, j, \ldots, k\}, \{\ell, m, \ldots, n\}\} = \{C,B\}$, is a rational successor of $\{\mathcal{F},C\}$.

It is not a unique rational successor. For example, a subgroup $B_1 \subset N\backslash C$ might form a defensive coalition, leaving a residual $N\backslash(C \cup B_1)$, from which another defensive coalition $B_2$ could be formed, and so on. This may seem an unlikely sequence, since the unified opposition coalition B will yield the best per-capita payoff to the residual from the deviation $\{\mathcal{F},C\}$, and the members of N\C can presumably anticipate this. (Compare Game 14.3). In any case, however, the sequence will not end until all agents in N\C are organized into defensive coalitions so that there can be no exaction of tribute. Thus the value of C for this deviation is

$$m(y + z - h) - K \tag{15.10}$$

and the deviation will be profitable if

$$m > \frac{K}{z - h} \tag{15.11}$$

Thus, provided

$$n > \frac{K}{z - h} + \frac{L}{y} \text{ or } n > \frac{K}{z} + 1, \tag{15.12}$$

the fine partition will be disrupted by a deviation to a large enough C, a proper subset of N.

## 15.3 STABILITY OF THE GRAND COALITION

Consider a partition of form 2, $\mathcal{Q} = \{\{i, j, \ldots, k\}, \{l, m, \ldots, n\}\} = \{C, B\}$, and consider a deviation to the grand coalition. The excess for this deviation will be identically positive as

$$n(y + z) - K > m(y + z - h) - K + (n - m)(y + z - h) - K$$

$$= n(y + z - h) - 2K \tag{15.13}$$

This gain arises from two sources: first, the cancellation of the conflict cost, and second, the elimination of the duplication of the monitoring cost $K$. Similarly, for a partition of the form $\{B_1, B_2, \ldots, B_r\}$, a deviation to the grand coalition will produce an even greater margin of excess. We see that no partition other than $\mathcal{G}$ is $\xi$-stable.

Is $\mathcal{G}$ $\xi$-stable? Consider a deviation $\{\mathcal{G}, C\}$ with $C = \{i, j, \ldots, k\}$, $|C| = m$. If $m \geq n - (L/y)$, then the per capita value of C is at most $y + z - (K/m)$, whereas the per capita payoff in $\mathcal{G}$ is $y + z - (K/n)$, so the deviation is unprofitable. If $m < n - (L/y)$, then the successor is of form 2, and the per capita value of C is $y + z - h - (K/m)$, so, again, the deviation is unprofitable. Therefore, $\mathcal{G}$ is uniquely $\xi$-stable given assumptions (15.1)–(15.3).

## 15.4 THE GOVERNMENT GAME

The Gang Game is the government game, and the grand coalition in the Gang Game could be called the state. What we see is that within a specific territory, sufficiently compact so that $K$ can reasonably be supposed to be constant or nearly so, a single coalition for mutual protection, a single state, is a highly stable arrangement. *Since it is a cooperative coalition*, the state may find it in the mutual interest of its members to adopt what we think of as the normative role of the state; for example, to assure the efficient production of public goods.

However, we observe that the emergence of a single coalition for mutual protection is not universal. The simplifying assumptions incorporated in the Gang Game, while widely applicable, may not be universally applicable. One simplifying assumption that has been made tacitly is that there are no coalitional preferences. If there were groups within the population

that were strongly averse to coalescing with one another, then we might instead find some partitions of the form {C,B} stable and the grand coalition unstable. We have seen instances of this in Lebanon, Bosnia and elsewhere.

The supposition that $K$ is a constant regardless of the size of the coalition may seem excessive. Common sense suggests that even within a defined territory, $K$ would rise with the size of the coalition, even if less than in proportion. But there are reasons that might lead us to propose that $K$ would instead decline with an increase in the size of the coalition, especially as the coalition approaches a grand coalition. The coalition is, after all, a *commitment* on the part of its members to obey its rules and refrain from internal aggression. Failures of the agents to keep their commitments may be the consequences of weakness of will or of dishonesty. But weakness of will and dishonesty are not universal, as argued in Chapter 9; we need not be quite that cynical. Thus, a substantial part of the compliance with the rules *by members* may be voluntary, and to some extent also the members may monitor one another, reducing $K$ below what it would otherwise be. Indeed it is a commonplace that few laws can succeed without a substantial proportion of voluntary compliance. Thus, as $m$ increases toward $n$, the proportion of agents with some commitment to voluntary compliance increases, and $K$ may well decline as a result. This phenomenon could also contribute to the destabilization of exaction partitions of form 1. As in other matters, an attempt has been made to adopt the simplest assumptions consistent with the objective of the section.

Assumption i.3 assures us that a large majority cannot profit by expelling a very small minority and exacting tribute from them. If the minority is so small that it cannot organize for self-defense, then the minority is too small to be profitably exploited. Here, again, coalitional preferences may make a difference. If the minority is hated, the overhead cost $L$ of organizing exaction from them is offset, and perhaps even reversed, by the satisfaction of the hate motive. The tragic consequences of such a hate motive are familiar in the history of the twentieth century.

We see that the stable outcome of the Gang Game, excluding coalitional preferences, is a unique grand coalition. The key point is that the state-coalition *must* be unique, as that uniqueness is a condition for its principal benefit to its members, that is, the monopoly of predictable violence. At the same time, it would be in the interest of its members for such a unique grand coalition to adopt measures that would improve efficiency, such as the production of efficient quantities of public goods.

To what extent may we say that existing states are cooperative grand coalitions for defense in their respective territories? Clearly some are not. It is not difficult to find instances of non-cooperative leviathans in history

and in current events. To what extent can we say that any existing state approximates a grand coalition in the Gang Game? The theory given here is subject to the same criticism as a social contract theory, that it is at best an as-if theory, in that citizens now living have not been the signatories of any social contract, nor have had any opportunity either to make or decline a grand defensive coalition. (New states, such as Israel in 1948, might be exceptions.) What we can say is that *to the extent that* ideal rationality shapes human actions and institutions, we would expect to see such a grand coalition as a state, and where an entrenched leviathan might exist, we might expect to see a (nearly) grand coalition to resist it; and that we would expect this grand coalition also to perform other activities that might be in the mutual interests of the citizens, such as the production of public goods and efficiency-improving regulation. On the other hand, to the extent that human action and institutions are shaped by "imperfect recall" and perfect rationality, we would expect a real state to fall short of the ideal and to struggle with tax evasion, criminality, and opportunistic corruption.

## 15.5 INEQUALITIES, INEFFICIENCIES, AND POLICIES

Symmetry is a powerful, but implausible, assumption in the Gang Game. It seems likely that asymmetries, such as differences in $K$ across coalitions or in $y$ across individuals, could generate stable grand coalitions with dissimilar payoffs, including stable payments of tribute from some groups to others. Moreover, the players in the Gang Game have been described as a group of residents of a particular territory, but it may be that not all residents are players in the Gang Game, or indeed that the players are a small minority. Those who, for some reason, have no possibility of forming a defensive coalition are not players. They may be subject to exaction by coalitions of players, another complication the game set aside for simplicity. Slaves, serfs, and prisoners of war certainly are not players in this sense. Consider, for example, a feudal society. In such a society, players in the Gang Game are landlords wealthy enough to maintain a stable of horses and an arsenal of armor and cavalry weapons. Others (at least in rural society) are not players.

Further, the Gang Game as presented here simplifies by assuming that the territory is a point. The coalitional cost $K$ will rise if the coalition is expanded by including players who are at a greater distance from one another. For predominantly rural feudal societies, this is a crucial consideration, and because of it the stable territorial units seem quite small. In

ancient and medieval urban societies, $K$ can reasonably be supposed constant for a city, but rises rapidly with intercity consolidation, so that the city predictably becomes the political unit, and broader political unities are better thought of as interstate non-cooperative arrangements.

Clearly, the operations of the state are very much influenced by imperfect recall, and by bounded rationality as well. Simply by virtue of its monopoly of predictable violence, the state can impose penalties and establish rules to determine the circumstances in which the penalties are imposed. Thus, in particular, it can tax. But, again, by assumption it is not a leviathan (a non-cooperative equilibrium construct) but a cooperative coalition, so that, in the simple model given here, the taxation would be limited to the minimum necessary to defray $K$. In practice, of course, the decisions that must be made on behalf of the state-coalition will be much more complex. The simple game assumes that $z$ is a given constant. In fact, state decisions, such as the rules for imposition of tax and prohibition of activities that (however nonviolently) impose negative externalities, may influence the value of $z$. Thus even the minimal state has a very complex decision to make as to the best set of rules for imposition of taxes, a decision that calls on expert knowledge and complex calculation. If, in addition, there are opportunities to increase $z$ by means of efficient production of public goods, regulation of production and of markets, and so on, these decisions will be all the more complex.

That being so, rationality being bounded, and recall (knowledge of the strategic commitments of agents) being imperfect, the state-coalition will have to establish an apparatus for routine decision-making, a constitution. One way to do this would be simply to entrust the decisions to a single individual. For an illiterate society, in which the main determinant of individual welfare is the level of brigandage and exaction by privileged landlords, this may be as good a way as any, and it has the advantage that it can be implemented simply by the non-cooperative Hobbesian game described in McCain (1992). Perhaps it is not too far-fetched to construe Hobbes as making that very point. In more modern societies, it seems that representative institutions have some advantages. Thus, there will be yet another set of rules defining an imbedded game: the game of pressure groups, political parties, and parliamentary coalitions. To the extent that people do not know with certainty what is in their interest, economists, social scientists, and think-tanks become players in this game.

The rule-making function of the state may also extend to setting limits on coalitions that can or must be made in imbedded games. Koczy's (2007) cooperative solution for partition function games assumes "partitional deviations," but these have been excluded from this part of the book by the assumption that there are no cooperative arrangements outside of

coalitions. A cooperative game theory that includes partitional deviations could be a useful tool of normative analysis to select the limitations that the state might impose. Suppose, for example, that an analysis of this kind were to find that industry structures comprising minimal efficient firms would be more efficient than one with more concentrated oligopolies, but nevertheless because existing monopolies and oligopolies benefit from their monopoly power, the efficient structure cannot be achieved by means of "coalitional deviations." A government antitrust policy that forces the dissolution of existing, inefficiently concentrated industries would be a way of bringing about the "partitional deviation" to an efficiently organized economy. Constitutional establishment of a federal political structure or limitation on the formation or operation of political parties could be similarly defended in terms of "partitional deviations" in the politics game. Utopian thinking (for example, Buber, 1958) may also be understood as demanding partitional deviations.

The politics game might lead to policies that are not themselves efficient, except in a second-best sense that they are not as bad as civil war. Systematic discussion of this politics game is, however, beyond the scope of this book.

## 15.6 SUMMARY

This chapter has argued that the state may be thought of as a Hobbesian grand defensive coalition, and that such a grand coalition could also be expected to adopt public policies that promote efficiency. Discussion of equity will be beyond the scope of the book. In the government game, we proceed from the assumption that "law and order" is not a public good but an excludable good with zero (or very low) marginal cost within a given territory to an interpretation of the state as a grand coalition to provide "law and order." Since a coalition is formed to advance the mutual interests of its members, we would expect that the state would also produce public goods and adopt other policies in the general interest, within the limits of "imperfect recall" and, in practice, bounded rationality.

## NOTE

1. Recall note 6, Chapter 2.

# 16. Toward political economy

Political economy is an integrated study of economics and politics that allows us to pose and tentatively to answer both positive questions such as "whose interest is state action likely to advance?" and "what are the likely consequences of such-and-such economic policy?" and normative questions such as "what policies would best advance the interest of the whole population?" Since the formation of coalitions is a foundation both of politics (states, parties) and economics (business firms) the framework developed in this part of the book would seem to provide a language for political economy. This chapter will sketch some concepts toward such a political economy.

When we represent the "game" of social interaction either in coalition function or, as in this part of the book, in partition function form, each agent is supposed to be a member of exactly one coalition. In our actual life most of us are members of more than one coalition, and may be members of many: we may be members of a political party and a lobbying organization for a particular cause, a church, a local congregation or parish, a social club, we may be employed by, or be investors in, one or more business firms, be a member of a rural electrification cooperative, a farmer cooperative, and a mutual bank, and engaged in a very large number of agreements for the exchange of particular goods and services. In order to bring this reality within the range of a theory that employs partition functions, we must suppose that each person is simultaneously playing in two or more games.

For a modern economy, we must consider an individual agent as participating in at least three games: a government game, a production game, and an exchange game. The possibility that the same agents play simultaneously in a number of games is rarely mentioned in game theory, but plays a key part in Shapley's value theory. Shapley's theory is defined in part by the idea that the value is additive over the various games in which the agent plays.[1] Unfortunately, additivity may not be widely enough applicable in a practical context. Consider a group of rural people who (perhaps in about 1938) have neither electrical nor telephone service and are considering forming both a rural electrification cooperative and a rural telephone cooperative. These services may be complementary. With telephone service, they can organize a barn dance, and with electricity they can light the barn. In a less lighthearted mood, there may be many businesses

that need both services to be viable. To the extent that the outcomes of the games are either complementary or substitutable, additivity may fail.

## 16.1 IMBEDDING

Instead we proceed using the concept of imbedded games. In Chapter 6, Section 6.2.1, imbedding was defined for non-cooperative games in extensive form. For the model of encapsulated cooperation, each game can be resolved to a non-cooperative game, the consensus game. The government game determines parameters for both of the other games, and moreover the government (governing coalition) may participate as an individual agent in both the production and exchange games. The production game determines parameters for the exchange game, that is endowments of particular goods and services, and coalitions formed in the production game will usually participate in the exchange game as individual agents. Thus the exchange game (considered as the consensus game) is imbedded in the production game (similarly considered), and the production game is in turn imbedded in the government game.

The three-games approach is not unique but might encompass a large family of models. The distinctions between these three games may depend on the pragmatic purpose of the model. Consider, in particular, coalitions in the production game. Are the employees members of these coalitions? On the one hand, the agreement between an employer and an employee is a cooperative agreement, and some discussions of the firm treat it as a cooperative coalition among resource suppliers of all kinds, including the suppliers of labor. In some cases this is specifically in reference to firms in Japan (Aoki, 1980) or Germany (McCain, 1980). On the other hand, employment is a form of exchange (of service for money) and we conventionally speak of labor markets in capitalist economies. Thus we might subsume employer-employee relations to the exchange game, treating coalitions in the production game as coalitions of the suppliers of capital resources specifically.

The government game was considered in the previous chapter, and examples in other chapters illustrate aspects that may enter into the imbedded games of production and exchange. We need not be specific. Indeed, one of the advantages of the imbedded game procedure is that any economic game modeled as a non-cooperative equilibrium might be considered as imbedded in the government game. Many models well established in the literature could be "plugged in" without contradicting anything in the encapsulated cooperation model, and indeed this has been a major objective of the model. The example that follows in this chapter will be strictly intended as illustrative, and not in any sense final. To stress

the tentative character of the example, the example will be drawn from the history of economic thought, rather than recent economic theory, and will be a formalization in terms of encapsulated cooperation of a capitalist one-commodity "corn"[2] economy, very much in the classical framework.

## 16.2 A CORN ECONOMY

In the Ricardian framework, agents are of three types: landlords, laborers, and (capitalist) farmers. For a landlord or capitalist, to seek employment as a laborer is a dominated strategy, but laborer types have no choice, as seeking employment is the only strategy in their strategy sets.

In this model the exchange game comprises exchanges of corn for labor services (employment contracts) and of corn for land services (rental contracts). For the exchange game, landowners are endowed with specific quantities of land and each laborer with a specific quantity of potential labor services, or, in Marxist terms, labor power. Capitalists are endowed with some quantity of corn retained from past production. This stock of corn constitutes the "wages fund,"[3] which limits total wages, since wages must be paid in advance. Payment of rents, however, may be in the form of promissory notes against corn to be produced. This is a proper game. The solution will serve to determine the prices, that is, rents and wages. Classical economics suggests the following conjectures: (1) The grand coalition will support one or more imputations in the core of the exchange game. (2) The law of one price will prevail for wages for imputations in the core. (3) If landholdings are of different productivity, not all landholders will receive positive rents. The landholders excluded will be the holders of the least productive land. Other landholders will receive net payments equal to the differential productivity of the land they hold. (4) The rate of profit on stock will be equalized among the different capitalists.

Since corn will be treated as the medium of exchange, inflation and unemployment will play no part in this illustrative model, as in any classical one-commodity model. Problems of macroeconomics will be entirely beyond the scope of the example. For the example, we will proceed by backward induction, taking the exchange game first.

### 16.2.1 The Exchange Game

There are $m$ players of landlord type, $w$ players of worker type, and $c$ players of capitalist type, indexed with landlords $1, \ldots, m$; capitalists $m + 1, \ldots, m + c$; workers $m + c + 1, \ldots, m + c + a$.

In a classical corn economy, wages are limited by the wages fund, which

is the stock of corn in the hands of capitalists. For this example, each capitalist is endowed with a stock of corn amounting to $b$ units, and plans to employ $u$ workers. For the purpose of the example, for simplicity, capitalists are homogeneous. The wages fund is $cb$, and consequently the average wage is $cb/w = b/u$. Each landlord is endowed with a plot of land suited for cultivation by $u$ workers. (This defines a unit of land.) The productivity of these plots varies, and the output of plot $\acute{\imath}$, if it is cultivated, is $q_{\acute{\imath}}u$. (If a single person owns more than one plot of land then he can enter the game as a separate player for each plot.) The plots of land are indexed in such a way that $\acute{\imath} < \acute{\jmath} \Rightarrow q_{\acute{\imath}} \geq q_{\acute{\jmath}}$. It is assumed that $m > w/u$. Each worker is endowed with a certain amount of potential labor and nothing else. Then a coalition of one capitalist with $u$ workers and one landlord is (assuming the average wage is paid) the smallest that can engage in production and has no redundant resources. There are $c$ capitalists with $cu = w$.

Definition. Now consider a coalition of w′ workers, c′ capitalists, and m′ landlords comprising a set $L = \{\acute{\imath}, \acute{\jmath}, \ldots, \acute{k}\}$. Let $r = \min(m', w'/u, c')$ and let $M \subset L \ni |M| = r$ and $i \in M, j \in L \backslash M \Rightarrow i < j \Rightarrow q_{\acute{\imath}} > q_{\acute{\jmath}}$. That is, if the quantity of land is more than can be used with the labor that the coalition includes and can pay, then only the most productive plots will be used. If $r < m'$, the coalition is called a redundant-land coalition. In any case, the value of the coalition is $v = \sum_{i \in M} q_{\acute{\imath}}$.

From this definition it follows that the exchange game is proper and superadditive. This being so, we need consider only the grand coalition, since any stable imputation will be admissible for the grand coalition.

Since by assumption $m > w/u$ the grand coalition is a redundant-land coalition. In a Ricardian model the "marginal land" is the most productive land not currently in use. In this model, then, plot $(w/u) + 1$ is the marginal land and plots with $\acute{\imath} > w/u + 1$ are inframarginal. Here are some inferences about this exchange game.

(1) Marginal and inframarginal land receives no rent. Consider landlord $\acute{\imath}$ with $\acute{\imath} > w/u$. Suppose the payoff to landlord $\acute{\imath}$, $y_{\acute{\imath}} > 0$ and consider the deviation to $C = N \backslash \{\acute{\imath}\}$. Then $v(C) = v(N)$, so the excess for this deviation is exactly $y_{\acute{\imath}}$; and it follows that the grand coalition with a positive rent for a parcel of marginal or inframarginal land is not $\xi$-stable.

(2) For any set of one capitalist $\acute{\imath}$ and $u$ workers $\{\acute{\jmath}, \ldots, \acute{k}\} = L$, $y^* = y_{\acute{\imath}} + \sum_{j \in L} y_{\acute{\jmath}} = q_{(w/u+1)}$. That is, a capitalist and a standard team of workers together receive exactly the productivity of the marginal land. Let $S = \{\acute{\imath}, \acute{\jmath} \ldots \acute{k}, w + 1\}$.

   (a) Suppose instead that $y^* > q_{(w/u)+1}$. Consider the deviation to $C = N \backslash S$. The excess for this deviation is $y^* - q_{(w/u)+1} > 0$, so a candidate solution with $y^* > q_{(w/u)+1}$ is $\xi$-unstable.

(b) Suppose instead that $y^* < q_{(w/u)+1}$. Consider the deviation to S. The excess for this deviation is $q_{(w/u)+1} - y^* > 0$, so a candidate solution with $y^* < q_{(w/u)+1}$ is $\xi$-unstable.

(3) The law of one price applies to wages. By the above result, if any two workers $j$, $k$ have different payoffs $y_j \neq y_k$, the substitution of one for the other in the elements of S would result in different values for the sum of the payoffs to S, contrary to proposition 2.

(4) The rate of profit is equalized. For any two capitalists $i$, $j$ we must have $y_i = y_j$ by the same reasoning. Moreover, the common rate of profit is $\pi = (q_{w+1} - b)/b$.

(5) Rent is differential productivity. Consider a landlord with $i \leq w$. Then the payment to this landlord is $y_i = q_i - q_{w+1}$. Let S be $\{i, j, \ldots k, l\}$ where $j, \ldots, k$ are any $u$ workers and $l$ any capitalist. Suppose $y_i < q_i - q_{w+1}$ and consider a deviation to N\S. The value of the deviation is $v(N) - q_i$. The payoff to N\S in the candidate solution is $V(N) - (y_i - q_{w+1})$ (by 2 above) so that the excess is $y_i - (q_i - q_{w+1}) > 0$. Suppose $y_i < q_i - q_{w+1}$ and consider the deviation to S. The excess for this deviation is $q_i - y_i - q_{w+1} > 0$. Thus a payment to a landlord that differs from the differential productivity will not be $\xi$-stable.

(6) The grand coalition is $\xi$-stable with payments of differential rent to landlords, $ca/w$ to workers, and $q_{w+1} - a$ to capitalists. Consider any deviation S with $w'$ workers, $c'$ capitalists, and $m'$ landlords comprising a set $L = \{i, j, \ldots k\}$. The payoff to this group in the grand coalition is $y = (cb/w)w' + (q_{w+1} - b)c' + \sum_{i \in L}(q_i - q_{w+1})$. The value of the deviation is $v(S) = \sum_{i \in M} q_i$ with $M$ as in the definition; and with $r$ as in the definition this can be rewritten as $(cb/w)r + (q_{w+1} - b)r + \sum_{i \in L'}(q_i - q_{w+1})$. Term by term, each of these terms is less than or equal to the corresponding term in $y$. Thus the excess is nonpositive and the deviation does not disrupt the candidate solution.

Thus, the exchange game in this case will generate predictable prices, rents, and profits. Agents in the production game will anticipate this, and make their decisions accordingly. We now proceed to consider the production game.

### 16.2.2 The Production Game

Players in the production game are agents with enough wealth to supply their own (negligible) consumption of corn and to command the labor of a team of workers. We may assume that relations between employees and workers in production are non-cooperative and that employers face

a cost of monitoring effort, and that this cost includes both an overhead component and a variable component that increases with the size of the work force. We suppose that a work force of just $u$ workers is the team size that optimally balances the overhead cost against the variable cost. Thus, each production coalition will plan for a work force of just that size.

In a corn economy, the formation of a coalition for production can be identified with the formation of capital through frugality, or obtaining capital by alienating land. Classical economics assumes that landlords never save nor employ labor in cultivation. Nevertheless they command purchasing power in the form of corn, from their rental income. In a classical model, they spend their income on the wages of servants. These servants are consequently not available as employees in the exchange game as just described, and constitute the "unproductive labor" in the classical schema.

Accordingly, we suppose that agents in the production game are of two types, with different intertemporal preferences: one type with a lower discount rate for future consumption and one with a higher discount rate. An agent of the first type, endowed with land, and an agent of the second kind, endowed with capital, might form a coalition to transfer the capital to the agent of the first type in exchange for land at a price in corn between the agents' discounted present values of future rents. We also suppose that any attempt to consolidate production on two or more plots of land results in loss of efficiency due to "imperfect recall," so that production coalitions with more than one capitalist will be unstable. Classical economics did not envision externalities, so they will play no part in this model. This is a counterfactual simplifying assumption. As noted, a property-owner with more than enough land or labor to employ one work team in cultivation will enter the exchange game as a plurality of agents, one for each plot to be cultivated, and this is the stable partition for the production game.

### 16.2.3 The Government Game

Now consider the government game. We will take as given that the government is formed as a grand coalition of all players in the government game. Marx (who after all originated the concept of capitalism as an economic system) regarded capitalist government as a "dictatorship of the bourgeoisie," that is, a system in which laborers would be excluded from participation in politics by a property test for the franchise or some other unprivileged status. (This does not conflict in any way with the views of Smith, Malthus, and Ricardo). Thus, for Marx and Engels

(1848), "The executive of the modern state is but a committee for managing the common affairs of the whole bourgeoisie." Accordingly, laborers are not players in the government game. The governing grand coalition of capitalists and landlords is not an agent in the production or exchange games. Again, further formal development will be beyond the scope of this sketch.

An interesting question that arises is why absolute monarchies proved stable in many nineteenth-century capitalist governments. We might conjecture that absolutism could be stable as a way of resolving potential rivalry between landlords and non-landlord capitalists. Acemoglu et al. (2008) offer a model of this kind that lends itself to interpretation in terms of ω-stability. The landlords (in a coalition with the capitalists) might have the power to bring about a shift from absolute to constitutional monarchy, from which they would benefit in a naive sense, but they foresee that the capitalists would then have the power to abolish all vestiges of traditional privilege, leaving the landlords worse off on net; therefore, the landlords would not disrupt the absolute monarchy and instead support it against the pressures of the capitalist class. Of course, such rivalries also played a role in the historic "corn laws" controversy in Britain that influenced the growth of classical political economy.

In passing we might sketch two alternatives to capitalism in the same classical framework. For a Marxist workers' state, drawing mostly on the (early and brief) *Communist Manifesto*, only worker types are players in the government game. The governing grand coalition of workers becomes an agent in the production game and may be the only agent permitted. Nevertheless, employer-employee relations remain an exchange game. Imperfect recall in the determination of effort plays no part in this conception, so coalitions in the production game are supposed to be superadditive. In a "cooperative commonwealth", (Altenberg, 1990) coalitions in the production game are coalitions of workers, that is, worker cooperatives, not coalitions of owners. These coalitions may participate in exchange games to secure finance as well as selling their products, or finance may be arranged through government, while in some conceptions of this kind government may again be construed as a workers' state, a grand coalition in a government game in which only workers are players.

## 16.3 CONCLUSION AND SUMMARY

This chapter has sketched some concepts for a political economy in terms of encapsulated cooperation. Since agents will typically participate

in different coalitions for different purposes, the political economy is modeled as three nested consensus games, where the consensus game is a non-cooperative interpretation of the game of coalition formation. The production game is nested within the government game and thus reflects the policies and actions of the government among the rules of the production game. The production game determines the partition of property owners into producing coalitions, and the endowments with which those coalitions and individuals play the exchange game, which is in turn imbedded in the production game. Externalities are important in the government and production games, and although they are assumed to be absent from the exchange game, "imperfect recall" is important in each of the three nested games.

This is illustrated with a sketch of a classical "corn" economy with land of differential productivity. In a more modern (and neoclassical) economic model of a capitalist economy, we would relax the one-commodity assumption and explicitly allow for many distinct commodities. We might model a single exchange game, or perhaps distinct exchange games for the different goods and services exchanged for money. In such a game, banking might be modeled as exchange, and issues of macroeconomics might be addressed. The production game could reflect a given technology, with "substitution of factors of production," but also incorporate organizational costs as a representation of imperfect recall along Marshallean lines. Increasing returns due to indivisibilities or to division of labor might also be incorporated. For the government game, a more modern treatment would reflect the extension of the voting franchise to (almost?) all adults.

Clearly, there is much more to be done, and the work could go in a number of directions. Nevertheless, this must conclude the book. The objective has been to survey and criticize game theory as a tool for public policy analysis, and to propose an alternative approach at roughly the same level of generality. Specific applications could be the work of many more books by other authors, and I hope they will be.

## NOTES

1. This assumption allows one to decompose a game in coalition function form into a set of elementary games, a decomposition that underlies some of the mathematical properties of Shapley's value assignment. Additivity and decomposition are not needed for the core analysis nor for the calculation of the nucleolus.
2. Recall that "corn economy" is a phrase in the British language, not the American, so that "corn" means "grain," such as wheat or oats, depending on the country, rather than maize specifically. For this chapter "corn" is an abstract wage-good.
3. This wages fund assumption was perhaps one of the least persuasive of classical ideas, and the first to be abandoned, but contains an important true insight: if production takes

time, it will be necessary for the people who do the work to eat while production takes place, and if the money wage is raised while there is no increase in the quantity of wage-goods available, inflation is the only result. This fact has reasserted itself in the context of industrialization in the twentieth century and as recently as the food crisis of the spring of 2008. The unique function of capitalists in the classical corn economy is to supply the wage-goods from their stock of accumulated wage-goods.

# References


Abreu, Dilip, Prajit K. Dutta and Lones Smith (1994), "The Folk Theorem for Repeated Games: A New Condition," *Econometrica*, **64**, (4) (July), 939–48.

Acemoglu, Daron, Georgy Egorov and Konstantin Sonin (2008), "Dynamics and Stability of Constitutions, Coalitions, and Clubs," (Paper presented at the Third World Congress of the Game Theory Society, Evanston, Ill, August).

*The Age* (2002), "The Sage of Omaha's Trans-Atlantic Game," http://www.theage.com.au/articles/2002/09/23/1032734111833.html, 24 September (accessed 26 November 2007).

Alchian, Armen and H. Demsetz (1972), "Production, Information Costs, and Economic Organization," *American Economic Review*, **62** (5) (December), 777–95.

Allen, Paul (1998), "Outline History of Game Theory," http://www.econ.canterbury.ac.nz/hist.htm, (accessed 11 April 2009).

Altenberg, Lee (1990), "Beyond Capitalism: Leland Stanford's Forgotten Vision," *Sandstone and Tile*, **14** (1) (Winter), 8–20.

Andreoni, James and Emily Blanchard (2006), "Testing Subgame Perfection Apart from Fairness in Ultimatum Games," *Experimental Economics*, **9** (4), 307–21.

Aoki, M. (1980), "A Model of the Firm as a Stockholder-Employee Cooperative Game," *American Economic Review*, **70** (4) (September), 600–610.

Arrow, Kenneth J. (1951), *Social Choice and Individual Values*, New York: Wiley.

Aumann, Robert J. (1973), "Disadvantageous Monopolies," *Journal of Economic Theory*, **6** (1) (February), 1–11.

Aumann, Robert J. (1974), "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, **1**, 67–96.

Aumann, R. J. (1976), "Agreeing to Disagree," *Annals of Statistics*, **4**, 1236–9.

Aumann, R. J. (1987), "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, **55**, 1–18.

Aumann, R. J. (1997), "Rationality and Bounded Rationality," *Games and Economic Behavior*, **21**, 2–14.

Aumann, Robert J. (2003), "Presidential Address," *Games and Economic Behavior*, **45**, 2–14.
Aumann, R. J. (2004), "Address," Second World Congress of the Game Theory Society, Marseilles.
Aumann, R. J. (2005), "War and Peace (Nobel Prize Lecture)," The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2005, http://nobelprize.org/nobel_prizes/economics/laureates/2005/aumann-lecture.html (accessed 9 June 2007).
Aumann, R. J. and J. H. Dreze, (1974), "Cooperative Games with Coalition Structure," *International Journal of Game Theory*, **3**, 217–37.
Aumann, R. J. and J. Dreze (2005), "When All Is Said and Done, How Should You Play and What Should You Expect?" discussion paper, Departement des Sciences Economiques de l'Université catholique de Louvain.
Aumann, Robert and Michael Maschler (1964), "The Bargaining Set for Cooperative Games," in M. Dresher, L. S. Shapley, and A. W. Tucker, (eds), *Advances in Game Theory*, Annals of Mathematics Studies Number 52, Princeton: Princeton University Press, pp. 443–76.
Aumann, Robert J. and Michael Maschler (1972), "Some Thoughts on the Minimax Principle," *Management Science*, **18** (5) (January), 54–63.
Aumann, Robert J. and Michael Maschler (1985), "Game Theoretic Analysis of a Bancruptcy Problem from the Talmud," *Journal of Economic Theory*, **36** (2) (August), 195–213.
Aumann, Robert and S. Sorin (1989), "Cooperation and Bounded Recall," *Games and Economic Behavior*, **1** (1) (March), 5–39.
Axelrod, Robert (1981), "The Emergence of Cooperation among Egoists," *American Political Science Review*, **75** (2) (June), 306–18.
Axelrod, Robert (1984), *The Evolution of Cooperation*, New York: Basic Books.
Azariadis, Costas (1981), "Self-Fulfilling Prophesies," *Journal of Economic Theory*, **25**, 380-96.
Baharad, Eyal and Zvika Neeman (2002), "The Asymptotic Strategy Proofness of Scoring and Condorcet Consistent Rules," *Review of Economic Design*, **7**, 311–40.
Baumol, William and Wallace Oates (1975), *The Theory of Environmental Policy*, Englewood Cliffs, NJ: Prentice-Hall.
Baumol, William J., John C. Panzar and Robert D. Willig (1982), *Contestable Markets and the Theory of Industry Structure*, New York: Harcourt Brace Jovanovich.
Berg, Joyce, John Dickhaut and Kevin McCabe (1995), "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, **10** (1), 122–42.

Bernheim, B. Douglas (1984), "Rationalizable Strategic Behavior," *Econometrica*, **52** (4) (July), 1007–28.
Bernheim, B. Douglas, Bezalel Peleg and Michael D. Whinston (1987), "Coalition-Proof Nash Equilibria I. Concepts," *Journal of Economic Theory*, **42**, 1–12.
Brandenburger, Adam M. and Barry J. Nalebuff (1997), *Co-opetition*, New York: Doubleday Business.
Bray, Marianne (2003), "Arm-Wrestle Settles Network Battle," CNN.com, http://edition.cnn.com/2003/WORLD/asiapcf/auspac/03/10/offbeat.nz.wrestle/, 11 March (accessed 15 November 2007).
Bresnahan, Timothy (1981), "Duopoly Models with Consistent Conjectures," *American Economic Review*, **71** (5) (December), 934–45.
Broder, John (2007), "Governors Join in Creating Regional Pacts on Climate Change," New York Times, http://www.nytimes.com/2007/11/15/washington/15climate.html?em&ex=1195275600&en=45651e5591715bd9&ei=5087%0A, 15 November (accessed 15 November 2007).
Brothwell, Don (1987), *The Bog Man and the Archeology of People*, Cambridge, MA: Harvard University Press.
Brown, A.C. (1975), *A Bodyguard of Lies*, New York: Harper and Row.
Buber, Martin (1958), *Paths in Utopia*, Boston, MA: Beacon.
Camerer, Colin (1987), "Do Biases in Probability Judgement Matter in Markets: Experimental Evidence," *American Economic Review*, **77** (5) (December), 981–97.
Carraro, Carlo (2003), *The Endogenous Formation of Economic Coalitions*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Carter, J. R. and M. D. Irons (1991), "Are Economists Different, and If So, Why?," *Journal of Economic Perspectives*, **5**, 171–7.
Cass, David and Karl Shell (1983), "Do Sunspots Matter?," *Journal of Political Economy*, **91** (2), 193-227.
Chatterjee, Satyajit, Russell Cooper and B Ravikumar (1993), "Strategic Complementarity in Business Formation: Aggregate Fluctuations and Sunspot Equilibria," *Review of Economic Studies*, **60** (4) (October), 795–811.
Cheung, S. N. S. (1968), "Private Property Rights and Sharecropping," *Journal of Political Economy*, **76** (6), 1107–22.
Chwe, Michael Suk-Young (1994), "Farsighted Coalitional Stability," *Journal of Economic Theory*, **63** (2) (August), 229–325.
Clark, John Bates (1899), *The Distribution of Wealth*, New York: Macmillan.
Colby, Bonnie G. (2000), "Cap-and-Trade Policy Challenges: A Tale of Three Markets," *Land Economics*, **76** (4) (November), 638–58.

Coleman, Andrew M. (2003), "Cooperation, Psychological Game Theory, and Limitations of Rationality in Social Interaction," *Behavioral and Brain Sciences*, **26** (2) (April), 139–53.

Conlon, John R. (1993), "Can the Government Talk Cheap? Communication, Announcements, and Cheap Talk," *Southern Economic Journal*, **60** (2) (October), 418–29.

Cooper, Russell W., Douglas V. DeJong, Robert Forsythe and Thomas W. Ross (1990), "Selection Criteria in Coordination Games: Some Experimental Results," *American Economic Review*, **80** (1) (March), 218–33.

Cramton, Peter (1997), "Spectrum Auctions: An Early Assessment," *Journal of Economics and Management Strategy*, **6** (3) (Fall), 431–96.

Crandall, John (2007), "Who Invented the Traffic Light? William Potts of Detroit," Suite101.com, http://transportationhistory.suite101.com/article.cfm/who_invented_the_traffic_light (accessed 15 November 2007).

Dasgupta, Partha and Eric Maskin (2008), "On the Robustness of Majority Rule," *Journal of the European Economic Association*, **6** (5), 949–73.

Davis, M. and M. Maschler (1965), "The Kernel of a Cooperative Game," *Naval Research Logistics Quarterly*, **12**, 223–59.

Dawes, Robyn M. (1980), "Social Dilemmas," *Annual Review of Psychology*, **31**, 169–93.

Debreu, G. and Herbert E. Scarf (1963), "A Limit Theorem on the Core of an Economy," *International Economic Review*, **4** (3) (September), 235–46.

Dresher, M., A. W. Tucker and P. Wolfe (1957), *Contributions to the Theory of Games, Volume III*, Annals of Mathematics Studies, Number 39, Princeton: Princeton University Press.

Dresher, M., L. Shapley and A. W. Tucker (1964), *Advances in Game Theory*, Annals of Mathematics Studies, Number 52, Princeton: Princeton University Press.

Dutta, Prajit (1999), *Strategies and Games: Theory and Practice*, Cambridge, MA: MIT Press.

Elster, Jon (1977), "Ulysses and the Sirens: A Theory of Imperfect Rationality," *Social Science Information*, **16** (October), 469–526.

Etzioni, Amitai (1988), *The Moral Dimension: Toward A New Economics*, New York: The Free Press.

Famous-inventors.com (2006), "Biography of Garrett Morgan," http://www.famous-inventors.com/biography-of-garrett-morgan.html (accessed 15 November 2007).

Fehr, E. and Urs Fischbacher (2004), "Third-Party Punishment and Social Norms," *Evolution and Human Behavior*, **25**, 63–87.

Feldman, Allan (1979), "Manipulating Voting Procedures," *Economic Inquiry*, **17** (July), 452-74.
Foley, Duncan K. (1967), "Resource Allocation in the Public Sector," *Yale Economic Essays*, **7** (Spring), 73–6.
Forgo, Ferenc, Jeno Szep and Ferenc Szidarovszky (1999), *Introduction to the Theory of Games: Concepts, Methods, Applications*, Dordrecht: Kluwer.
Foster, Dean P. and Rakesh V. Vohra (1997), "Calibrated Learning and Correlated Equilibrium," *Games and Economic Behavior*, **21** (12) (October–November), 40–55.
Friedman, Daniel (1998), "Evolutionary Economics Goes Mainstream: A Review of the Theory of Learning in Games," *Evolutionary Economics*, **8** (4), 423–32.
Fudenberg, Drew and David K. Levine (1981), "Perfect Equilibria of Finite and Infinite Horizon Games," UCLA Department of Economics Working Paper, no. 216.
Fudenberg, Drew and David K. Levine (1999), "Conditional Universal Consistency," *Games and Economic Behavior*, **29**, 104–30.
Fudenberg, D. and E. Maskin (1986), "The Folk Theorem in Repeated Games with Discounting and with Incomplete Information," *Econometrica*, **54**, 533–54.
Galbraith, John Kenneth (1973), *Economics and the Public Purpose*, Boston, Mass: Houghton Mifflin.
Gardner, Roy (2003), *Game Theory for Business and Economics*, 2nd edition, Hoboken: Wiley.
Gibbard, Alan (1973), "Manipulation of Voting Schemes: A General Result," *Econometrica*, **41** (4), 587–601.
Gillies, D.B. (1953), "Some Theorems on n-Person Games", doctoral dissertation, Princeton University, Princeton, NJ.
Gintis, Herbert (2007), "A Framework for the Unification of the Behavioral Sciences," *Behavioral and Brain Sciences*, **30**, 1–16.
Glenn, David (2007), "3 Americans Win Nobel Prize in Economics," *Chronicle of Higher Education*, 26 October, 16.
Greenberg, J. (1994), "Coalition Structures," in R. J. Aumann and S. Hart (eds), *Handbook of Game Theory*, Amsterdam: Elsevier, pp. 1305–37.
Greenberg, Joseph (1990), *The Theory of Social Situations: An Alternative Game-Theoretic Approach*, Cambridge: Cambridge University Press.
Groves, T. and J. Ledyard, (1977), "Optimal Allocation of Public Goods: A Solution to the 'Free Rider' Problem," *Econometrica*, **45** (4) (May), 783–809.
Guth, W., R. Schmittberger and B. Schwartze (1982), "An Experimental

Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization*, **3**, 376–88.
Haag, Matthew and Roger Lagunoff (2005), "On the Size and Structure of Group Cooperation," working paper, Department of Economics, Georgetown University, http://www9.georgetown.edu/faculty/lagunofr/size2.pdf, 28 June (accessed 11 September 2008).
Hahn, Robert (1989), "Economic Prescriptions for Environmental Problems: How the Patient Followed the Doctor's Orders," *Journal of Economic Perspectives*, **3** (2) (Spring), 95–114.
Hamilton, Richard, Frank Linnehan and Roger McCain (2008), "Emergency Department Overcrowding as a Nash Equilibrium: Hypothesis and Test by Questionnaire," working paper, LeBow College of Business, Drexel University, http://faculty.lebow.drexel.edu/mccainr/top/Working_Papers.html, 11 September (accessed 11 September 2008).
Hanson, Niel (2001), *The Custom of the Sea*, New York: John Wiley and Sons.
Hargreaves Heap, Shaun P. and Yanis Varoufakis (1995), *Game Theory: A Critical Introduction*, London: Routledge.
Harris, Robert and Jeremy Paxman (2002), *A Higher Form of Killing: The Secret History of Chemical and Biological Warfare*, New York: Random House.
Harsanyi, John C. (1963), "A Simplified Bargaining Model for the n-Person Cooperative Game", *International Economic Review*, **4** (2), 194–220.
Harsanyi, John C. (1967–8), "Games with Incomplete Information Played by 'Bayesian' Players," *Management Science*, **14**, Part I, pp.159–83; Part II, pp. 320–34; Part III, pp. 486–502.
Harsanyi, John (1975), "Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls' Theory," *American Political Science Review*, **69**, 594–606.
Harsanyi, John and Reinhard Selten (1972), "A Generalized Nash Solution for Two-Person Bargaining Games with Incomplete Information," *Management Science*, **18** (5) (January), 80–106.
Harsanyi, John and Reinhard Selten (1988), *A General Theory of Equilibrium Selection in Games*, Cambridge, MA: MIT Press.
Hart, Sergiu and Andreu Mas-Colell (2000), "A Simple Adaptive Procedure Leading to Correlated Equilibrium," *Econometrica*, **68** (5) (September), 1127–50.
Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, Richard McElreath, Michael Alvard, Abigail Barr, Jean Ensminger, Natalie Smith Henrich, Kim Hill, Francisco Gil-White, Michael Gurven, Frank W. Marlowe, John Q. Patton and

David Tracer (2005), "'Economic man' in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies," *Behavioral and Brain Sciences*, **28** (6) (December), 795–815.
Hobbes, Thomas (1968 [1660]), *Leviathan*, Harmondsworth: Penguin.
Hodgson, Geoffrey M. (2002), "Darwinism in Economics: From Analogy to Ontology," *Journal of Evolutionary Economics*, **12** (3) (July), 259–81.
Hurwicz, Leonid (1973), "The Design of Mechanisms for Resource Allocation," *American Economic Review*, **63** (2) (May), 1–30.
Irwin, Frank (1971), *Intentional Behavior and Motivation*, Philadelphia: Lippincott.
Kaldor, Nicholas (1934), "The Equilibrium of the Firm," *Economic Journal*, **44** (173) (March), 60–76.
Keystone Automobile Club (1927), Eastern Tours (Keystone Automobile Club).
Klemperer, Paul (2002), "How (Not) to Run Auctions: The European Telecom Auctions," *European Economic Review*, **46**, (45) (April), 829–45.
Koczy, Laszlo (2007), "A Recursive Core for Partition Function Form Games," *Theory and Decision*, **63**, 41–51.
Kreps, D.M., Paul Milgrom, John Roberts and R. Wilson (1982), "Rational Cooperation in the Finitely Repeated Prisoners Dilemma," *Journal of Economic Theory*, **27** (2), 245–52.
Kuhn, H.W. (1997), *Classics in Game Theory*, Princeton: Princeton University Press.
Kuhn, H. W. and A. W. Tucker (1950), *Contributions to the Theory of Games, Volume I*, Annals of Mathematics Studies, Number 24, Princeton: Princeton University Press.
Kuhn, H. W. and A. W. Tucker (1953), *Contributions to the Theory of Games, Volume II*, Annals of Mathematics Studies, Number 28, Princeton: Princeton University Press.
Lamar-Sterling, Sara (2006), "Drawing Straws," Park Avenue Methodist Church, New York, http://www.parkavemethodist.org/sermon.php?s=1, 28 May (accessed 15 November 2007).
Lave, L. B. (1965), "Factors Affecting Cooperation in the Prisoner's Dilemma," *Behavioral Science*, **10**, 26–38.
Lehman, William (2007), "US Airways: A Heritage Story," US Airways, http://www.usairways.com/awa/content/aboutus/pressroom/history/allegheny.aspx (accessed 26 November 2006).
Lohr, Steve (2007), "Three Share Nobel in Economics for Work on Social Mechanisms," *New York Times*, http://www.nytimes.com/2007/10/16/business/16nobel.html?_r=1&ref=business&oref=slogin, 16 October (accessed 16 October 2007).

Lucas, W. F. (1968), "A Game with no Solution," *Bulletin of the American Mathematical Society*, **74**, 237–39.
Luce, R. Duncan and Howard Raiffa (1957), *Games and Decisions*, New York: Wiley and Sons.
Mailath, G. J. (1998), "Do People Play Nash Equilibrium? Lessons from Evolutionary Game Theory," *Journal of Economic Literature*, **36** (3), 1347–74.
Market Design, Inc. (2007), "MDI Projects,"http://www.market-design.com/projects-telecommunications.html, (accessed 1 November 2007).
Marquis, Don (1950), *The Lives and Times of Archy and Mehitabel*, Garden City, New York: Doubleday.
Marwell, Gerald and Ruth E. Ames (1981), "Economists Free Ride: Does Anybody Else?," *Journal of Public Economics*, **15**, 295–310.
Marx, Karl (1845), "Theses on Feuerbach," Marxists.org, http://www.marxists.org/archive/marx/works/1845/theses/index.htm, (accessed 20 October 2007).
Marx, Karl and F. Engels (1848), "Manifesto of the Communist Party," http://www.marxists.org/archive/marx/works/1848/communist-manifesto/, (accessed 7 September 2004.
Maskin, E. (1999), "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies* (paper presented at the summer workshop of the Econometric Society in Paris, June), **66**, 23–38.
Maskin, Eric (2004), "Bargaining, Coalitions and Externalities," plenary lecture, Second World Congress of the Game Theory Society, Marseille.
Maskin, Eric and Jean Tirole (1987), "Correlated Equilibria and Sunspots," *Journal of Economic Theory*, **43**, 364–73.
McCain, Roger A. (1972), "Distributional Equality and Aggregate Utility: Further Comment," *American Economic Review*, **62** (3) (June), 497–500.
McCain, Roger A. (1978), "Endogenous Bias in Technical Progress and Environmental Policy," *American Economic Review*, **68** (4) (September), 538-46.
McCain, Roger A. (1980), "A Theory of Codetermination," *Zeitschift für Nationalökonomie*, **40** (12), 65–90.
McCain, Roger A. (1985), "Economic Planning for Market Economies: The Optimality of Planning in an Economy with Uncertainty and Asymmetrical Information," *Economic Modelling*, **2** (4) (October), 317–23.
McCain, Roger A. (1991), "A Theory of Economic Planning for Market Economies: The Optimality of Planning," in S. Baghwan Dahiya (ed.),

*Theoretical Foundations of Development Planning*, Vol. 3, New Delhi: Concept Publishing.

McCain, Roger A. (1992), "Heuristic Coordination Games: Rational Action Equilibrium and Objective Social Constraints in a Linguistic Conception of Rationality," *Social Science Information*, **31** (4), 711–34.

McCain, Roger A. (2000), "Differences, Games, and Pluralism," *Behavioral and Brain Science*, **23** (5), 688.

McCain, Roger A. (2004), *Game Theory: A Nontechnical Introduction to the Analysis of Strategy*, Mason, Ohio: Thomson South-Western.

McCain, Roger (2007a), "Cooperation and Effort, Reciprocity and Mutual Supervision in Worker Cooperatives," in Sonja Novkovic (ed.), *Advances in the Economic Analysis of Participatory and Labor-Managed Firms*, Greenwich, CT: JAI Press.

McCain, Roger (2007b), "Welfare Economics," in William Darity (ed.), *International Encyclopedia of the Social Sciences*, 2nd edition, Farmington Hills, MI: Gale.

McKelvey, R. D. and T. R. Palfrey (1992), "An Experimental Study of the Centipede Game," *Econometrica*, **60** (4), 803–36.

McKinsey, J. C. C. (1952), *Introduction to the Theory of Games*, New York: McGraw-Hill.

McMillan, John, Michael Rothschild and Robert Wilson (1997), "Introduction (to Special Issue on Market Design and Spectrum Auctions)," *Journal of Economics and Management Strategy*, **6** (3) (Fall), 425–30.

Montet, Christian and Daniel Serra (2003), *Game Theory and Economics*, Basingstoke: Palgrave Macmillan.

Morehouse, L. G. (1967), "One-Play, Two-Play, Five-Play and Ten-Play Runs of Prisoner's Dilemma," *Journal of Conflict Resolution*, **11**, 354–62.

Morgenstern, Oskar and G. Schwödiauer (1976), "Competition and Collusion in Bilateral Markets," *Zeitschrift für Nationalökonomie*, **36** (4), 217–45.

Moulin, Herve (1982), *Game Theory for the Social Sciences*, New York: New York University Press.

Moulin, H. and B. Peleg, (1982), "Cores of Effectivity Functions and Implementation Theory," *Journal of Mathematical Economics*, **10** (1) (June), 115–45.

Myerson, Roger B. (1976), "Values of Games in Partition Function Form," *International Journal of Game Theory*, **6** (1) (September), 23–31.

Myerson, Roger B. (1979), "Incentive Compatibility and the Bargaining Problem," *Econometrica*, **47**, 61–73.

Myerson, R. (1986), "Multistage Games with Communication," *Econometrica*, **54**, 323–58.
Nash, John (1950a), "Equilibrium Points in n-Person Games," *Proceedings of the National Academy of Science*, **36**, 48–9.
Nash, John (1950b), "The Bargaining Problem," *Econometrica*, **18**, 155–62.
Nash, John (1951), "Non-Cooperative Games," *Annals of Mathematics*, **2** (September), 286–95.
Nash, John (1953), "Two-Person Cooperative Games," *Econometrica*, **21** (January), 128–40.
New School History of Economic Thought Website (2007), "William Stanley Jevons, 1835–1882," http://cepa.newschool.edu/het/profiles/jevons.htm (accessed 29 November 2007).
Newell, A. and H. A. Simon, (1972), *Human Problem Solving*, Englewood Cliffs, NJ: Prentice-Hall.
Nutter, Warren (1964), "Duopoly, Oligopoly and Emerging Competition," *Southern Economic Journal*, **30** (April), 342–52.
Oosterbeek, Hessel, Randolph Sloof and Gijs van de Kuilen (2004), "Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-analysis," *Experimental Economics*, **7** (2) (June), 171–88.
Osborne, Martin (2004), *An Introduction to Game Theory*, Oxford: Oxford University Press.
Pareto, Vilfredo (1971 [1906]), *Manual of Political Economy*, Fairfield, NJ: A. M. Kelley.
Pearce, D.G. (1984), "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, **52** (4) (July), 1029–50.
Peck, James and Karl Shell (1991), "Market Uncertainty: Correlated and Sunspot Equilibria in Imperfectly Competitive Economies," *Review of Economic Studies*, **58** (5) (October), 1011–29.
Peleg, B. and P. Sudhölter (2003), *Introduction to the Theory of Cooperative Games*, Dordrecht: Kluwer.
Pigou, A. C. (1920), *Economics of Welfare*, London: Macmillan.
Plott, Charles R. (1997), "Laboratory Experimental Testbeds: Application to the PCS Auction", *Journal of Economics and Management Strategy*, **6** (3), 605–38.
Poundstone, William (1992), *Prisoner's Dilemma*, New York: Doubleday.
Rapoport, Anatole and Albert M. Chammah (1965), *Prisoner's Dilemma*, Ann Arbor: University of Michigan Press.
Rawls, John (1971), *A Theory of Justice*, Cambridge, MA: Belknap Press.
Ray, Debraj and Rajiv Vohra (1999), "A Theory of Endogenous Coalition Structures," *Games and Economic Behavior*, **26**, 286–336.
Rosenstein-Rodan, Paul (1943), "Problems of Industrialization of Eastern

and South-Eastern Europe," *Economic Journal*, **53**, (June–September), 202–11.
Rosenthal, R. (1981), "Games of Perfect Information, Predatory Pricing, and the Chain Store Paradox," *Journal of Economic Theory*, **25**, 92–100.
Roth, Alvin and Ido Erev (1995), "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior*, **8**, 164–212.
Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara and Shmuel Zamir (1991), "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," *American Economic Review*, **91** (5) (December), 1068–95.
Rousseau, Jean-Jacques (1754), "Discourse on Inequality (G. D. H. Cole Translation), Part II," the Constitution Society, http://www.constitution.org/jjr/ineq_04.htm (accessed 9 September 2006).
Royal Swedish Academy of Sciences (2007), "Scientific Background on Mechanism Design Theory," http://nobelprize.org/nobel_prizes/economics/laureates/2007/ecoadv07.pdf (accessed 1 November 2007).
Rubinstein, Ariel (1979), "Equilibrium in Supergames with the Overtaking Criterion," *Journal of Economic Theory*, **31**, 227–50.
Samuelson, Paul (1954), "The Pure Theory of Public Expenditures," *Review of Economics and Statistics*, **36**, (4) (November), 387–9.
Satterthwaite, M. (1975), "Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *Journal of Economic Theory*, **10** (2) (April), 189–203.
Scarf, Herbert E. (1967), "The Core of an N Person Game," *Econometrica*, **35** (1) (January), 50–69.
Schelling, Thomas (1960), *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.
Schelling, T.C. (1978), *Micromotives and Macrobehavior*, New York: Norton.
Schelling, T.C. (1980), "The Intimate Contest for Self-Command," *Public Interest*, **60** (Summer), 94–118.
Schmeidler, David (1969), "The Nucleolus of a Characteristic Function Game," *SIAM Journal on Applied Mathematics*, **17** (6) (November), 1163–70.
Scitovsky, Tibor (1954), "Two Concepts of External Economies," *Journal of Political Economy*, **62** (2) (April), 143–51.
Selten, Reinhard (1964), "Valuation of N-Person Games," in M. Dresher, L. S. Shapley and A. W. Tucker (eds), *Advances in Game Theory*, Annals of Mathematics Studies, Number 52, Princeton: Princeton University Press, pp. 577–626.

Selten, Reinhard (1965), "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragetragheit," *Zeitschrift für die gesamte Staatswissenschaft*, **121**, 301–24.
Selten, Reinhard (1975), "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, **4** , 25–55.
Selten, Reinhard and G. Gigerenzer (2001), *Bounded Rationality: The Adaptive Toolbox*, Cambridge, MA: MIT Press. Retrieved 28 November 2007 from Drexel University, Hagerty Library, Net Library, http://www.netlibrary.com.
Sen, Amartya (1970), "The Impossibility of a Paretian Liberal," *Journal of Political Economy*, **78** (1) (January), 152–7.
Sen, Amartya (1985), *Commodities and Capabilities*, Amsterdam: North-Holland.
Serrano, Roberto (2003), "Fifty Years of the Nash Program, 1953–2003," working paper, Brown University.
Shapley, L.S. (1953), "Stochastic Games," *Proceedings, National Academy of Sciences*, **39**, 1095–100.
Shapley, Lloyd and Martin Shubik (1952), "Solutions of N-Person Games with Ordinal Utilities (abstract)," *Econometrica*, **21**, 348–9.
Shapley, Lloyd and Martin Shubik (1954), "A Method for Evaluating the Distribution of Power in a Committee System," *American Political Science Review*, **48** (3) (September), 787–92.
Shapley, Lloyd and Martin Shubik (1969), "On the Core of an Economic System with Externalities," *American Economic Review*, **59** (4) (September), 678–84.
Shubik, M. (1962), "Incentives, Decentralized Control, the Assignment of Joint Costs and Internal Pricing," *Management Science*, **8** (3) (April), 325–43.
Simon, H.A. (1995), "Artificial Intelligence: An Empirical Science," *Artificial Intelligence*, **77**, 95–127.
Smith, Adam (1994 [1776]), *The Wealth of Nations*, New York: The Modern Library.
Smith, John Maynard (1972), "Game Theory and the Evolution of Fighting," in John Maynard Smith (ed.), *On Evolution*, Edinburgh: Edinburgh University Press, pp. 8–28.
Stahl, Dale O. and Paul W. Wilson (1995), "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, **10**, 218–54.
Stanley, T.D and Ume Tran (1998), "Economics Students Need Not Be Greedy: Fairness and the Ultimatum Game," *Journal of Socio-Economics*, **27** (6), 657–63.

Stein, Jeremy (1989), "Cheap Talk and the Fed: A Theory of Imprecise Policy Announcements," *American Economic Review*, **79**, 32–42.
Stiglitz, Joseph (1974), "Incentives and Risk Sharing in Sharecropping," *Review of Economic Studies*, **41** (2) (April), 219–55.
Stuart, H.W, Jr (2001), "Cooperative Games and Business Strategy," in K. Chatterjee and William Samuelson (eds), *Game Theory and Business Applications*, Norwell, MA: Kluwer, pp. 189-211.
Telser, Lester (1978), *Economic Theory and the Core*, Chicago: University of Chicago Press.
Telser, Lester (1997), *Joint Ventures of Labor and Capital*, Ann Arbor: University of Michigan Press.
Thrall, R.M. and W. F. Lucas (1963), "N-Person Games in Partition Function Form," *Naval Research in Logistics Quarterly*, **10**, 281–98.
Tsebelis, George (1990), *Nested Games: Rational Choice in Comparative Politics*, Berkeley, CA: University of California Press.
Tucker, A.W. and R. D. Luce (eds) (1959), *Contributions to the Theory of Games, Volume IV*, Annals of Mathematics Studies Number 40, Princeton: Princeton University Press.
*USA Today* (2006), "Shared Sacrifice? Not For These Airline Executives," http://www.usatoday.com/news/opinion/editorials/2006-02-01-airlines-edit_x.htm, 1 February (accessed 26 November 2007).
Van Huyck, J., R. Battalio and R. Beil (1990), "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure," *American Economic Review*, **80**, 234–48.
Vickrey, William (1960), "Utility, Strategy, and Social Decision Rules," *Quarterly Journal of Economics*, **74**, 507–35.
Vickrey, W. (1961), "Counterspeculation, Auctions, and Competitive Sealed Tenders," *Journal of Finance*, **16** (March), 8–37.
von Neumann, John (1959), "On the Theory of Games of Strategy," trans. Sonya Bargmann, in A. W. Tucker and R. D. Luce (eds), *Contributions to the Theory of Games, Volume IV*, Annals of Mathematics Studies Number 40, Princeton: Princeton University Press, pp. 13–42.
von Neumann, John and and Oskar Morgenstern (2004), *Theory of Games and Economic Behavior*, 60th anniversary edition, Princeton: Princeton University Press.
Waud, Roger (1970), "Public Interpretation of Federal Reserve Discount Rate Changes: Evidence on the 'Announcement Effect'," *Econometrica*, **38**, 231–50.
Weintraub, E. Roy (1992), "Introduction," in E. Roy Weintraub (ed.), *Toward a History of Game Theory*, Durham: Duke University Press, pp. 3–12.
Williamson, O. (1964), *The Economics of Discretionary Behavior:*

*Managerial Objectives in the Theory of the Firm*, Englewood Cliffs, NJ: Prentice-Hall.
Zaretsky, Adam M. (1998), "Going Once, Going Twice, Sold: Auctions and the Success of Economic Theory," The Federal Reserve Bank of St. Louis, http://stlouisfed.org/publications/re/1998/a/re1998a4.html, (accessed 1 November 2007).
Zermelo, E. (1913), "Uber eine Anwendung der Mengenlehre auf die theorie des schachspiels," *Proceedings, Fifth International Congress of Mathematicians*, **2**, 501–4.
Zeuthen, Frederick (1930), *Problems of Monopoly and Economic Warfare*, London: Routledge and Kegan Paul.
Zhao, J. (1992), "The Hybrid Solutions of an n-person Game," *Games and Economic Behavior*, **6** (3), 145–60.

# Index