2 Game Theory

Introduction

"Interactive Decision Theory" would perhaps be a more descriptive name for the discipline usually called Game Theory. This discipline concerns the behaviour of decision makers (*players*) whose decisions affect each other. As in non-interactive (one-person) decision theory, the analysis is from a rational, rather than a psychological or sociological viewpoint. The term "Game Theory" stems from the formal resemblance of interactive decision problems (*games*) to parlour games such as chess, bridge, poker, monopoly, diplomacy or battleship. The term also underscores the rational, "cold," calculating nature of the analysis.

The major applications of game theory are to economics, political science (on both the national and international levels), tactical and strategic military problems, evolutionary biology, and, most recently, computer science. There are also important connections with accounting, statistics, the foundations of mathematics, social psychology, and branches of philosophy such as epistemology and ethics. Game theory is a sort of umbrella or "unified field" theory for the rational side of social science, where "social" is interpreted broadly, to include human as well as non-human players (computers, animals, plants). Unlike other approaches to disciplines like economics or political science, game theory does not use different, ad hoc constructs to deal with various specific issues, such as perfect competition, monopoly, oligopoly, international trade, taxation, voting, deterrence, and so on. Rather, it develops methodologies that apply in principle to *all* interactive situations, then sees where these methodologies lead in each specific application. Often it turns out that there are close relations between results obtained from the general game-theoretic methods and from the more ad hoc approaches. In other cases, the game-theoretic approach leads to new insights, not suggested by other approaches.

We use a historical framework for discussing some of the basic ideas of the theory, as well as a few selected applications. But the viewpoint will be modern; the older ideas will be presented from the perspective of where they have led. Needless to say, we do not even attempt a systematic historical survey.

This chapter originally appeared in *The New Palgrave: A Dictionary of Economics*, Volume 2, edited by J. Eatwell, M. Milgate, and P. Newman, pp. 460–482, Macmillan, London, 1987. Reprinted with permission.

1910–1930

During these earliest years, game theory was preoccupied with *strictly competitive* games, more commonly known as *two-person zero-sum* games. In these games, there is no point in cooperation or joint action of any kind: if one outcome is preferred to another by one player, then the preference is necessarily reversed for the other. This is the case for most two-person parlour games, such as chess or two-sided poker; but it seems inappropriate for most economic or political applications. Nevertheless, the study of the strictly competitive case has, over the years, turned out remarkably fruitful; many of the concepts and results generated in connection with this case are in fact much more widely applicable, and have become cornerstones of the more general theory. These include the following:

i. The *extensive* (or *tree*) *form* of a game, consisting of a complete formal description of how the game is played, with a specification of the sequence in which the players move, what they know at the times they must move, how chance occurrences enter the picture, and the *payoff* to each player at the end of play. Introduced by von Neumann (1928), the extensive form was later generalized by Kuhn (1953), and has been enormously influential far beyond zero-sum theory.

ii. The fundamental concept of *strategy* (or pure strategy) of a player, defined as a complete plan for that player to play the game, as a function of what he observes during the course of play, about the play of others and about chance occurrences affecting the game. Given a strategy for each player, the rules of the game determine a unique outcome of the game and hence a payoff for each player. In the case of two-person zero-sum games, the sum of the two payoffs is zero; this expresses the fact that the preferences of the players over the outcomes are precisely opposed.

iii. The strategic (or matrix) form of a game. Given strategies $s^1, \ldots, s^n$ for each of the n players, the rules of the game determine a unique outcome, and hence a payoff $H^i(s^1, \ldots, s^n)$ for each player i. The *strategic* form is simply the function that associates to each profile $s := (s^1, \ldots, s^n)$ of strategies, the *payoff profile*

$$H(s) := (H^1(s), \ldots, H^n(s)).$$

For two-person games, the strategic form often appears as a matrix: the rows and columns represent pure strategies of Players 1 and 2 respectively, whereas the entries are the corresponding payoff profiles. For zero-sum games, of course, it suffices to give the payoff to Player 1. It has been said that the simple idea of thinking of a game in its matrix form is in itself one of the greatest contributions of Game Theory. In facing an interactive situation, there is a great temptation to think only in terms of "what should I do?" When one writes down the matrix, one is led to a different viewpoint, one that explicitly takes into account that the other players are also facing a decision problem.

iv. The concept of *mixed* or *randomized* strategy, indicating that rational play is not in general describable by specifying a single pure strategy. Rather, it is often non-deterministic, with specified probabilities associated with each one of a specified set of pure strategies. When randomized strategies are used, payoff must be replaced by expected payoff. Justifying the use of expected payoff in this context is what led to expected utility theory, whose influence extends far beyond game theory (see *1930–1950*, viii).

v. The concept of "individual rationality." The *security level* of Player i is the amount max min $H^i(s)$ that he can guarantee to himself, independent of what the other players do (here the max is over i's strategies, and the min is over $(n-1)$-tuples of strategies of the players other than i). An outcome is called *individually rational* if it yields each player at least his security level. In the game tic-tac-toe, for example, the only individually rational outcome is a draw; and indeed, it does not take a reasonably bright child very long to learn that 'correct' play in tic-tac-toe always leads to a draw.

Individual rationality may be thought of in terms of pure strategies or, as is more usual, in terms of mixed strategies. In the latter case, what is being "guaranteed" is not an actual payoff, but an expectation; the word "guarantee" means that this level of payoff can be attained in the mean, regardless of what the other players do. This "mixed" security level is always at least as high as the 'pure' one. In the case of tic-tac-toe, each player can guarantee a draw even in the stronger sense of pure strategies. Games like this—i.e. having only one individually rational payoff profile in the 'pure' sense—are called *strictly determined.*

Not all games are strictly determined, not even all two-person zero-sum games. One of the simplest imaginable games is the one that game theorists call "matching pennies," and children call "choosing up" ("odds and evens"). Each player privately turns a penny either heads up or tails up. If the choices match, 1 gives 2 his penny; otherwise, 2 gives 1 his penny. In the pure sense, neither player can guarantee more than -1, and hence the game is not strictly determined. But in expectation, each player can guarantee 0, simply by turning the coin heads up or tails up with 1/2–1/2 probabilities. Thus (0,0) is the only payoff profile that is individually rational in the mixed sense. Games like this—i.e. having only one individually rational payoff profile in the "mixed" sense—are called *determined*. In a determined game, the (mixed) security level is called the *value*, strategies guaranteeing it *optimal*.

vi. *Zermelo's theorem*. The very first theorem of Game Theory (Zermelo, 1913) asserts that chess is strictly determined. Interestingly, the proof does not construct "correct" strategies explicitly; and indeed, it is not known to this day whether the "correct" outcome of chess is a win for white, a win for black, or a draw. The theorem extends easily to a wide class of parlour games, including checkers, go, and chinese checkers, as well as less well-known games such as hex and gnim (Gale, 1979, 1974); the latter two are especially interesting in that one can use Zermelo's theorem to show that Player 1 can force a win, though the proof is non-constructive, and no winning strategy is in fact known. Zermelo's theorem does not extend to card games such as bridge and poker, nor to the variant of chess known as kriegsspiel, where the players cannot observe their opponents' moves directly. The precise condition for the proof to work is that the game be a two-person zero-sum game of *perfect information*. This means that there are no simultaneous moves, and that everything is open and 'above-board': at any given time, all relevant information known to one player is known to all players.

The domain of Zermelo's theorem—two-person zero-sum games of perfect information—seems at first rather limited; but the theorem has reverberated through the decades, creating one of the main strands of game theoretic thought. To explain some of the developments, we must anticipate the notion of *strategic equilibrium* (Nash, 1951; see *1950–1960*, i). To remove the two-person zero-sum restriction, H. W. Kuhn (1953) replaced the notion of "correct," individually rational play by that of equilibrium. He then proved that *every n-person game of perfect information has an equilibrium in pure strategies*.

In proving this theorem, Kuhn used the notion of a *subgame* of a game; this turned out crucial in later developments of strategic equilibrium theory, particularly in its economic applications. A subgame relates to the whole game like a subgroup to the whole group or a linear subspace to the whole space; while part of the larger game, it is self-contained, can be played in its own right. More precisely, if at any time, all the players know everything that has happened in the game up to that time, then what happens from then on constitutes a subgame.

From Kuhn's proof it follows that every equilibrium (not necessarily pure) of a subgame can be extended to an equilibrium of the whole game. This, in turn, implies that every game has equilibria that remain equilibria when restricted to any subgame. R. Selten (1965) called such equilibria *subgame perfect*. In games of perfect information, the equilibria that the Zermelo–Kuhn proof yields are all subgame perfect.

But not all equilibria are subgame perfect, even in games of perfect information. Subgame perfection implies that when making choices, a player looks forward and assumes that the choices that will subsequently be made, by himself and by others, will be rational; i.e. in equilibrium. Threats which it would be irrational to carry through are ruled out. And it is precisely this kind of forward-looking rationality that is most suited to economic application.

Interestingly, it turns out that subgame perfection is not enough to capture the idea of forward-looking rationality. More subtle concepts are needed. We return to this subject below, when we discuss the great flowering of strategic equilibrium theory that has taken place since 1975, and that coincides with an increased preoccupation with its economic applications. The point we wished to make here is that these developments have their roots in Zermelo's theorem.

A second circle of ideas to which Zermelo's theorem led has to do with the foundations of mathematics. The starting point is the idea of a game of perfect information with an infinite sequence of stages. Infinitely long games are important models for interactive situations with an indefinite time horizon—i.e. in which the players act as if there will always be a tomorrow.

To fix ideas, let A be any subset of the unit interval (the set of real numbers between 0 and 1). Suppose two players move alternately, each choosing a digit between 1 and 9 at each stage. The resulting infinite sequence of digits is the decimal expansion of a number in the unit interval. Let G_A be the game in which 1 wins if this number is in A, and 2 wins otherwise. Using Set Theory's "Axiom of Choice," Gale and Stewart (1953) showed that Zermelo's theorem is false in this situation. One can choose A so that G_A is not strictly determined; that is, against each pure strategy of 1, Player 2 has a winning pure strategy, and against each pure strategy of 2, Player 1 has a winning pure strategy. They also showed that if A is open or closed, then G_A is strictly determined.

Both of these results led to significant developments in foundational mathematics. The axiom of choice had long been suspect in the eyes of mathematicians; the extremely anti-intuitive nature of the Gale–Stewart non-determinateness example was an additional nail in its coffin, and led to an alternative axiom, which asserts that G_A is strictly determined for every set A. This axiom, which contradicts the axiom of choice, has been used to provide an alternative axiomatization for set theory (Mycielski and Steinhaus, 1964), and this in turn has spawned a large literature (see Moschovakis, 1980, 1983). On the other hand, the positive result of Gale and Stewart was successively generalized to wider and wider families of sets A that are "constructible" in the appropriate sense (Wolfe, 1955;

Davis, 1964), culminating in the theorem of Martin (1975), according to which G_A is strictly determined whenever A is a Borel set.

Another kind of perfect information game with infinitely many stages is the *differential game*. Here time is continuous but usually of finite duration; a decision must be made at each instant, so to speak. Typical examples are games of pursuit. The theory of differential games was first developed during the 1950s by Rufus Isaacs at the Rand Corporation; his book on the subject was published in 1965, and since then the theory has proliferated greatly. A differential game need not necessarily be of perfect information, but very little is known about those that are not. Some economic examples may be found in Case (1979).

vii. *The minimax theorem*. The minimax theorem of von Neumann (1928) asserts that every two-person zero-sum game with finitely many pure strategies for each player is determined; that is, when mixed strategies are admitted, it has precisely one individually rational payoff vector. This had previously been verified by E. Borel (e.g. 1924) for several special cases, but Borel was unable to obtain a general proof. The theorem lies a good deal deeper than Zermelo's, both conceptually and technically.

For many years, minimax was considered the elegant centrepiece of game theory. Books about game theory concentrated on two-person zero-sum games in strategic form, often paying only desultory attention to the non-zero sum theory. Outside references to game theory often gave the impression that non-zero sum games do not exist, or at least play no role in the theory.

The reaction eventually set in, as it was bound to. Game theory came under heavy fire for its allegedly exclusive concern with a special case that has little interest in the applications. Game theorists responded by belittling the importance of the minimax theorem. During the fall semester of 1964, the writer of these lines gave a beginning course in Game Theory at Yale University, without once even mentioning the minimax theorem.

All this is totally unjustified. Except for the period up to 1928 and a short period in the late Forties, game theory was never exclusively or even mainly concerned with the strictly competitive case. The forefront of research was always in n-person or non-zero sum games. The false impression given of the discipline was due to the strictly competitive theory being easier to present in books, more "elegant" and complete. But for more than half a century, that is not where most of the action has been.

Nevertheless, it is a great mistake to belittle minimax. While not the centrepiece of game theory, it *is* a vital cornerstone. We have already seen how the most fundamental concepts of the general theory—extensive form, pure strategies, strategic form, randomization, utility theory—were spawned in connection with the minimax theorem. But its importance goes considerably beyond this.

The fundamental concept of non-cooperative n-person game theory—the strategic equilibrium of Nash (1951)—is an outgrowth of minimax, and the proof of its existence is modelled on a previously known proof of the minimax theorem. In cooperative n-person theory, individual rationality is used to define the set of *imputations*, on which much of the cooperative theory is based. In the theory of repeated games, individual rationality also plays a fundamental role.

In many areas of interest—stochastic games, repeated games of incomplete information, continuous games (i.e. with a continuum of pure strategies), differential games, games played by automata, games with vector payoffs—the strictly competitive case already presents a good many of the conceptual and technical difficulties that are present in general. In these areas, the two-person zero-sum theory has become an indispensable spawning and proving ground, where ideas are developed and tested in a relatively familiar, "friendly" environment. These theories could certainly not have developed as they did without minimax.

Finally, minimax has had considerable influence on several disciplines outside of game theory proper. Two of these are statistical decision theory and the design of distributed computing systems, where minimax is used for "worst case" analysis. Another is mathematical programming; the minimax theorem is equivalent to the duality theorem of linear programming, which in turn is closely related to the idea of shadow pricing in economics. This circle of ideas has fed back into game theory proper; in its guise as a theorem about linear inequalities, the minimax theorem is used to establish the condition of Bondareva (1963) and Shapley (1967) for the non-emptiness of the core of an n-person game, and the Hart–Schmeidler (1989) elementary proof for the existence of correlated equilibria.

viii. *Empirics*. The correspondence between theory and observation was discussed already by von Neumann (1928), who observed that the need to randomize arises endogenously out of the theory. Thus the phenomenon of bluffing in poker may be considered a confirmation of the theory. This kind of connection between theory and observation is typical of game theory and indeed of economic theory in general. The "observations" are often qualitative rather than quantitative; in practice, we do observe bluffing, though not necessarily in the proportions predicted by theory.

As for experimentation, strictly competitive games constitute one of the few areas in game theory, and indeed in social science, where a fairly sharp, unique "prediction" is made (though even this prediction is in general probabilistic). It thus invites experimental testing. Early experiments failed miserably to confirm the theory; even in strictly determined games, subjects consistently reached individually irrational outcomes. But experimentation in rational social science is subject to peculiar pitfalls, of which early experimenters appeared unaware, and which indeed mar many modern experiments as well. These have to do with the motivation of the subjects, and with their understanding of the situation. A determined effort to design an experimental test of minimax that would avoid these pitfalls was recently made by B. O'Neill (1987); in these experiments, the predictions of theory were confirmed to within less than one per cent.

1930–1950

The outstanding event of this period was the publication, in 1944, of the *Theory of Games and Economic Behavior* by John von Neumann and Oskar Morgenstern. Morgenstern was the first economist clearly and explicitly to recognize that economic agents must take the interactive nature of economics into account when making their decisions. He and von Neumann met at Princeton in the late Thirties, and started the collaboration that culminated in the *Theory of Games*. With the publication of this book, Game Theory came into its own as a scientific discipline.

In addition to expounding the strictly competitive theory described above, the book broke fundamental new ground in several directions. These include the notion of a cooperative game, its coalitional form, and its von Neumann–Morgenstern stable sets. Though axiomatic expected utility theory had been developed earlier by Ramsey (1931), the account of it given in this book is what made it "catch on." Perhaps most important, the book made the first extensive applications of game theory, many to economics.

To put these developments into their modern context, we discuss here certain additional ideas that actually did not emerge until later, such as the core, and the general idea of a solution concept. At the end of this section we also describe some developments of this period not directly related to the book, including games with a continuum of strategies, the computation of minimax strategies, and mathematical advances that were instrumental in later work.

i. *Cooperative games*. A game is called *cooperative* if commitments—agreements, promises, threats—are fully binding and enforceable (Harsanyi 1966, p. 616). It is called *non-cooperative* if commitments are not enforceable, even if pre-play communication between the players is possible. (For motivation, see *1950–1960*, iv.)

Formally, cooperative games may be considered a special case of non-cooperative games, in the sense that one may build the negotiation and enforcement procedures explicitly into the extensive form of the game. Historically, however, this has not been the mainstream approach. Rather, cooperative theory starts out with a formalization of games (the coalitional form) that abstracts away altogether from procedures and from the question of how each player can best manipulate them for his own benefit; it concentrates, instead, on the possibilities for agreement. The emphasis in the non-cooperative theory is on the individual, on what strategy he should use. In the cooperative theory it is on the group: What coalitions will form? How will they divide the available payoff between their members?

There are several reasons that cooperative games came to be treated separately. One is that when one does build negotiation and enforcement procedures explicitly into the model, then the results of a non-cooperative analysis depend very strongly on the precise form of the procedures, on the order of making offers and counter-offers, and so on. This may be appropriate in voting situations in which precise rules of parliamentary order prevail, where a good strategist can indeed carry the day. But problems of negotiation are usually more amorphous; it is difficult to pin down just what the procedures are. More fundamentally, there is a feeling that procedures are not really all that relevant; that it is the possibilities for coalition forming, promising and threatening that are decisive, rather than whose turn it is to speak.

Another reason it that even when the procedures are specified, non-cooperative analyses of a cooperative game often lead to highly non-unique results, so that they are often quite inconclusive.

Finally, detail distracts attention from essentials. Some things are seen better from a distance; the Roman camps around Metzada are indiscernible when one is in them, but easily visible from the top of the mountain. The coalitional form of a game, by abstracting away from details, yields valuable perspective.

The idea of building non-cooperative models of cooperative games has come to be known as the *Nash program* since it was first proposed by John Nash (1951). In spite of the difficulties just outlined, the programme has had some recent successes (Harsanyi, 1982; Harsanyi and Selten, 1972; Rubinstein, 1982). For the time being, though, these are isolated; there is as yet nothing remotely approaching a general theory of cooperative games based on non-cooperative methodology.

ii. A *game in coalitional form*, or simply *coalitional game*, is a function v associating a real number $v(\mathrm{S})$ with each subset S of a fixed finite set I, and satisfying $v(\varnothing) = 0$ ($\varnothing$ denotes the empty set). The members of I are called *players*, the subsets S of I *coalitions*, and $v(\mathrm{S})$ is the *worth* of S.

Some notation and terminology: The number of elements in a set S is denoted $|\mathrm{S}|$. A *profile* (of strategies, numbers, etc.) is a function on I (whose values are strategies, numbers, etc.). If x is a profile of numbers and S a coalition, we write $x(\mathrm{S}) := \sum_{i \in \mathrm{S}} x^i$.

An example of a coalitional game is the *3-person voting game*; here $|\mathrm{I}| = 3$, and $v(\mathrm{S}) = 1$ or 0 according as to whether $|\mathrm{S}| \geqslant 2$ or not. A coalition S is called *winning* if $v(\mathrm{S}) = 1$, *losing* if $\mathrm{v(S)} = 0$. More generally, if w is a profile of non-negative numbers (*weights*) and q (the *quota*) is positive, define the *weighted voting game* v by $v(\mathrm{S}) = 1$ if $w(\mathrm{S}) \geqslant q$, and $v(\mathrm{S}) = 0$ otherwise. An example is a parliament with several parties. The players are the parties, rather than the individual members of parliament, w^i is the number of seats held by party i, and q is the number of votes necessary to form a government (usually a simple majority of the parliament). The weighted voting game with quota q and weights w^i is denoted $[q; w]$; e.g., the three-person voting game is [2; 1, 1, 1].

Another example of a coalitional game is a *market game*. Suppose there are l natural resources, and a single consumer product, say 'bread', that may be manufactured from these resources. Let each player i have an endowment e^i of resources (an l-vector with non-negative coordinates), and a concave production function u^i that enables him to produce the amount $u^i(x)$ of bread given the vector $x = (x_1, \ldots, x_l)$ of resources. Let $v(\mathrm{S})$ be the maximum amount of bread that the coalition S can produce; it obtains this by redistributing its resources among its members in a manner that is most efficient for production, i.e.

$$v(\mathrm{S}) = \max\left\{ \sum_{i \in \mathrm{S}} u^i(x^i) \colon \sum_{i \in \mathrm{S}} x^i = \sum_{i \in \mathrm{S}} e^i \right\}$$

where the x^i are restricted to have non-negative coordinates.

These examples illustrate different interpretations of coalitional games. In one interpretation, the payoff is in terms of some single desirable physical commodity, such as bread; $v(\mathrm{S})$ represents the maximum total amount of this commodity that the coalition S can procure for its members, and it may be distributed among the members in any desired way. This is illustrated by the above description of the market game.

Underlying this interpretation are two assumptions. First, that of *transferable utility* (TU): that the payoff is in a form that is freely transferable among the players. Second, that of *fixed threats*: that S can obtain a maximum of $v(\mathrm{S})$ no matter what the players outside of S do.

Another interpretation is that $v(\mathrm{S})$ represents some appropriate index of S's strength (if it forms). This requires neither transferable utility nor fixed threats. In voting games, for example, it is natural to define $v(\mathrm{S}) = 1$ if S is a winning coalition (e.g. can form a government or ensure passage of a bill), 0 if not. Of course, in most situations represented by voting games, utility is not transferable.

Another example is a market game in which the x^i are consumption goods rather than resources. Rather than bread, $\sum_{i \in \mathrm{S}} u^i(x^i)$ may represent a social welfare function such as is often used in growth or taxation theory. While $v(\mathrm{S})$ cannot then be divided in an arbitrary way among the members of S, it still represents a reasonable index of S's strength. This is a situation with fixed threats but without TU.

Von Neumann and Morgenstern considered strategic games with transferable payoffs, which is a situation with TU but without fixed threats. If the profile s of strategies is played, the coalition S may divide the amount $\sum_{i \in \mathrm{S}} H^i(\mathrm{S})$ among its members in any way it pleases. However, what S gets depends on what players outside S do. Von Neumann and Morgenstern defined $v(\mathrm{S})$ as the maxmin payoff of S in the two-person zero-sum game in which the players are S and I\S, and the payoff to S is $\sum_{i \in \mathrm{S}} H^i(s)$; i.e., as the expected payoff that S can assure itself (in mixed strategies), no matter what the others do. Again, this is a reasonable index of S's strength, but certainly not the only possible one.

We will use the term *TU coalitional game* when referring to coalitional games with the TU interpretation.

In summary, the coalitional form of a game associates with each coalition S a single number $v(\mathrm{S})$, which in some sense represents the total payoff that that coalition can get or may expect. In some contexts, $v(\mathrm{S})$ fully characterizes the possibilities open to S; in others, it is an index that is indicative of S's strength.

iii. *Solution concepts*. Given a game, what outcome may be expected? Most of game theory is, in one way or another, directed at this question. In the case of two-person zero-sum games, a clear answer is provided: the unique individually rational outcome. But in almost all other cases, there is no unique answer. There are different criteria, approaches, points of view, and they yield different answers.

A *solution concept* is a function (or correspondence) that associates outcomes, or sets of outcomes, with games. Usually an "outcome" may be identified with the profile of payoffs that outcome yields to the players, though sometimes we may wish to think of it as a strategy profile.

Of course a solution concept is not just any such function or correspondence, but one with a specific rationale; for example, the strategic equilibrium and its variants for strategic form games, and the core, the von Neumann–Morgenstern stable sets, the Shapley value and the nucleolus for coalitional games. Each represents a different approach or point of view.

What will "really" happen? Which solution concept is "right"? None of them; they are indicators, not predictions. Different solution concepts are like different indicators of an economy; different methods for calculating a price index; different maps (road, topo, political, geologic, etc., not to speak of scale, projection, etc.); different stock indices (Dow Jones, Standard and Poor's NYSE, etc., composite, industrials, utilities, etc.); different batting statistics (batting average, slugging average, RBI, hits, etc.); different kinds of information about rock climbs (arabic and roman difficulty ratings, route maps, verbal descriptions of the climb, etc.); accounts of the same event by different people or different media; different projections of the same three-dimensional object (as in architecture or engineering). They depict or illuminate the situation from different angles; each one stresses certain aspects at the expense of others.

Moreover, solution concepts necessarily leave out altogether some of the most vital information, namely that not entering the formal description of the game. When applied to a voting game, for example, no solution concept can take into account matters of custom, political ideology, or personal relations, since they don't enter the coalitional form. That does not make the solution useless. When planning a rock climb, you certainly want to take into account a whole lot of factors other than the physical characteristics of the rock, such as the season, the weather, your ability and condition, and with whom you are going. But you also do want to know about the ratings.

A good analogy is to distributions (probability, frequency, population, etc.). Like a game, a distribution contains a lot of information; one is overwhelmed by all the numbers. The median and the mean summarize the information in different ways; though other than by simply stating the definitions, it is not easy to say how. The definitions themselves do have a certain fairly clear intuitive content; more important, we gain a feeling for the relation between a distribution and its median and mean from experience, from working with various specific examples and classes of examples over the course of time.

The relationship of solution concepts to games is similar. Like the median and the mean, they in some sense summarize the large amount of information present in the formal description of a game. The definitions themselves have a certain fairly clear intuitive content, though they are not predictions of what will happen. Finally, the relations between a game and its core, value, stable sets, nucleolus, and so on is best revealed by seeing where these solution concepts lead in specific games and classes of games.

iv. *Domination, the core and imputations*. Continuing to identify "outcome" with "payoff profile," we call an outcome y of a game *feasible* if the all-player set I can achieve it. An outcome x *dominates* y if there exists a coalition S that can achieve at least its part of x, and each of whose members prefers x to y; in that case we also say that S can *improve upon* y. The *core* of a game is the set of all feasible outcomes that are not dominated.

In a TU coalitional game v, feasibility of x means $x(\mathrm{I}) \leqslant v(\mathrm{I})$, and x dominating y via S means that $x(\mathrm{S}) \leqslant v(\mathrm{S})$ and $x^i > y^i$ for all i in S. The core of v is the set of all feasible y with $y(\mathrm{S}) \geqslant v(\mathrm{S})$ for all S.

At first, the core sounds quite compelling; why should the players be satisfied with an outcome that some coalition can improve upon? It becomes rather less compelling when one realizes that many perfectly ordinary games have empty cores, i.e. every feasible outcome can be improved upon. Indeed, this is so even in as simple a game as the 3-person voting game.

For a coalition S to improve upon an outcome, players in S must trust each other; they must have faith that their comrades inside S will not desert them to make a coalition with other players outside S. In a TU 3-person voting game, $y := (1/3, 1/3, 1/3)$ is dominated via {1, 2} by $x := (1/2, 1/2, 0)$. But 1 and 2 would be wise to view a suggested move from y to x with caution. What guarantee does 1 have that 2 will really stick with him and not accept offers from 3 to improve upon x with, say, (0, 2/3, 1/3)? For this he must depend on 2's good faith, and similarly 2 must depend on 1's.

There are two exceptions to this argument, two cases in which domination does not require mutual trust. One is when S consists of a single player. The other is when S = I, so that there is no one outside S to lure one's partners away.

The requirement that a feasible outcome y be undominated via one-person coalitions (*individual rationality*) and via the all-person coalition (*efficiency* or *Pareto optimality*) is thus quite compelling, much more so than that it be in the core. Such outcomes are called *imputations*. For TU coalitional games, individual rationality means that $y^i \geqslant v(i)$ for all i (we do not distinguish between i and $\{i\}$), and efficiency means that $y(\mathrm{I}) = v(\mathrm{I})$. The outcomes associated with most cooperative solution concepts are imputations; the imputations constitute the stage on which most of cooperative game theory is played out.

The notion of core does not appear explicitly in von Neumann and Morgenstern, but it is implicit in some of the discussions of stable sets there. In specific economic contexts, it is implicit in the work of Edgeworth (1881) and Ransmeier (1942). As a general solution concept in its own right, it was developed by Shapley and Gillies in the early Fifties. Early references include Luce and Raiffa (1957) and Gillies (1959).

v. *Stable sets*. The discomfort with the definition of core expressed above may be stated more sharply as follows. Suppose we think of an outcome in the core as "stable." Then we should not exclude an outcome y just because it is dominated by *some* other outcome x; we should demand that x itself be stable. If x is not itself stable, then the argument for excluding y is rather weak; proponents of y can argue with justice that replacing it with x would not lead to a more stable situation, so we may as well stay where we are. If the core were the set of all outcomes not dominated by any element of the core, there would be no difficulty; but this is not so.

Von Neumann and Morgenstern were thus led to the following definition: A set K of imputations is called *stable* if it is the set of all imputations not dominated by any element of K.

This definition guarantees neither existence nor uniqueness. On the face of it, a game may have many stable sets, or it may have none. Most games do, in fact, have many stable sets; but the problem of existence was open for many years. It was solved by Lucas (1969), who constructed a ten-person TU coalitional game without any stable set. Later, Lucas and Rabie (1982) constructed a fourteen-person TU coalitional game without any stable set and with an empty core to boot.

Much of the *Theory of Games* is devoted to exploring the stable sets of various classes of TU coalitional games, such as 3- and 4-person games, voting games, market games, compositions of games, and so on. (If v and w have disjoint player sets I and J, their *composition* u is given by $u(\mathrm{S}) := v(\mathrm{S} \cap \mathrm{I}) + w(\mathrm{S} \cap \mathrm{J})$. During the 1950s many researchers carried forward with great vigour the work of investigating various classes of games and describing their stable sets. Since then work on stable sets has continued unabated, though it is no longer as much in the forefront of game-theoretic research as it was then. All in all, more than 200 articles have been published on stable sets, some 80 per cent of them since 1960. Much of the recent activity in this area has taken place in the Soviet Union.

It is impossible here even to begin to review this large and varied literature. But we do note one characteristic qualitative feature. By definition, a stable set is simply a set of imputations; there is nothing explicit in it about social structure. Yet the mathematical description of a given stable set can often best be understood in terms of an implicit social structure or form of organization of the players. Cartels, systematic discrimination, groups within groups, all kinds of subtle organizational forms spring to one's attention. These forms are endogenous, they are not imposed by definition, they emerge from the analysis. It is a mystery that just the stable set concept, and it only, is so closely allied with endogenous notions of social structure.

We adduce just one, comparatively simple example. The TU 3-person voting game has a stable set consisting of the three imputations (1/2, 1/2, 0), (1/2, 0, 1/2), (0, 1/2, 1/2). The social structure implicit in this is that all three players will *not* compromise by dividing the payoff equally. Rather, one of the three 2-person coalitions will form and divide the payoff equally, with the remaining player being left "in the cold." Because any of these three coalitions can form, competition drives them to divide the payoff equally, so that no player will prefer any one coalition to any other.

Another stable set is the interval $\{(\alpha, 1-\alpha, 0)\}$, where α ranges from 0 to 1. Here Player 3 is permanently excluded from all negotiations; he is "discriminated against." Players 1 and 2 divide the payoff in some arbitrary way, not necessarily equally; this is because a coalition with 3 is out of the question, and so competition no longer constrains 1 and 2 in bargaining with each other.

vi. *Transferable utility*. Though it no longer enjoys the centrality that it did up to about 1960, the assumption of transferable utility has played and continues to play a major role in the development of cooperative game theory. Some economists have questioned the appropriateness of the TU assumption, especially in connection with market models; it has been castigated as excessively strong and unrealistic.

This situation is somewhat analogous to that of strictly competitive games, which as we pointed out above (*1930–1950*, vii), constitute a proving ground for developing and testing ideas that apply also to more general, non-strictly competitive games. The theory of NTU (non-transferable utility) coalitional games is now highly developed (see *1960–1970*, i), but it is an order of magnitude more complex than that of TU games. The TU theory is an excellent laboratory or model for working out ideas that are later applied to the more general NTU case.

Moreover, TU games are both conceptually and technically much closer to NTU games than strictly competitive games are to non-strictly competitive games. A very large part of the important issues arising in connection with non-strictly competitive games do not have any counterpart at all in strictly competitive games, and so simply cannot be addressed in that context. But by far the largest part of the issues and questions arising in the NTU theory do have counterparts in the TU theory, they can at least be addressed and dealt with there.

Almost every major advance in the NTU theory—and many a minor advance as well—has had its way paved by a corresponding advance in the TU theory. Stable sets, core, value, and bargaining set were all defined first for TU games, then for NTU. The enormous literature on the core of a market and the equivalence between it and competitive equilibrium (c.e.) in large markets was started by Martin Shubik (1959a) in an article on TU markets. The relation between the value and c.e. in large markets was also explored first for the TU case (Shapley, 1964; Shapley and Shubik, 1969b; Aumann and Shapley, 1974; Hart, 1977a), then for NTU (Champsaur, 1975, but written and circulated circa 1970; Aumann, 1975; Mas-Colell, 1977; Hart, 1977b). The same holds for the bargaining set; first TU (Shapley and Shubik, 1984); then NTU (Mas-Colell, 1988). The connection between balanced collections of coalitions and the non-emptiness of the core (*1960–1970*, viii) was studied first for TU (Bondavera, 1963; Shapley, 1967), then for NTU (Scarf, 1967; Billera, 1970b; Shapley, 1973a); this development led to the whole subject of Scarf's algorithm for finding points in the core, which he and others later extended to algorithms for finding market equilibria and fixed points of mappings in general. Games arising from markets were first abstractly characterized in the TU case (Shapley and Shubik, 1969a), then in the NTU case (Billera and Bixby, 1973; Mas-Colell, 1975). Games with a continuum of players were conceived first in a TU application (Milnor and Shapley, 1978, but written and circulated in 1961), then NTU (Aumann, 1964). Strategic models of bargaining where time is of the essence were first treated for TU (Rubinstein, 1982), then NTU (Binmore, 1987). One could go on and on.

In each of these cases, the TU development led organically to the NTU development; it isn't just that the one came before the other. TU is to cooperative game theory what *Drosophila* is to genetics. Even if it had no direct economic interest at all, the study of TU coalitional games would be justified solely by their role as an outstandingly suggestive research tool.

vii. *Single play*. Von Neumann and Morgenstern emphasize that their analysis refers to "one-shot" games, games that are played just once, after which the players disperse, never to interact again. When this is not the case, one must view the whole situation—including expected future interactions of the same players—as a single larger game, and it, too, is to be played just once.

To some extent this doctrine appears unreasonable. If one were to take it literally, there would be only one game to analyse, namely the one whose players include all persons ever born and to be born. Every human being is linked to every other through some chain of interactions; no person or group is isolated from any other.

Savage (1954) has discussed this in the context of one-person decisions. In principle, he writes, one should "envisage every conceivable policy for the government of his whole life in its most minute details, and decide here and now on one policy. This is utterly ridiculous ..." (p. 16). He goes on to discuss the *small worlds* doctrine, "the practical necessity of confining attention to, or isolating, relatively simple situations ..." (p. 82).

To a large extent, this doctrine applies to interactive decisions too. But one must be careful, because here "large worlds" have qualitative features totally absent from "small worlds." We return to this below (*1950–1960*, ii, iii).

viii. *Expected utility*. When randomized strategies are used in a strategic game, payoff must be replaced by expected payoff (*1910–1930*, iv). Since the game is played only once, the law of large numbers does not apply, so it is not clear why a player would be interested specifically in the mathematical expectation of his payoff.

There is no problem when for each player there are just two possible outcomes, which we may call "winning" and "losing," and denominate 1 and 0 respectively. (This involves no zero-sum assumption; e.g. all players could win simultaneously.) In that case the expected payoff is simply the probability of winning. Of course each player wants to maximize this probability, so in that case use of the expectation is justified.

Suppose now that the values of *i*'s payoff function H^i are numbers between 0 and 1, representing win probabilities. Thus, for the "final" outcome there are still only two possibilities; each pure strategy profile s induces a random process that generates a win for i with probability $H^i(s)$. Then the payoff expectation when randomized strategies are used still represents *i*'s overall win probability.

Now in any game, each player has a most preferred and a least preferred outcome, which we take as a win and a loss. For each payoff h, there is some probability p such that i would as soon get h with certainty as winning with probability p and losing with probability $1 - p$. If we replace all the h's by the corresponding p's in the payoff matrix, then we are in the case of the previous paragraph, so use of the expected payoff is justified.

The probability p is a function of h, denoted $u^i(h)$, and called *i*'s von Neumann–Morgenstern *utility*. Thus, to justify the use of expectations, each player's payoff must be replaced by its utility.

The key property of the function u^i is that if h and g are random payoffs, then i prefers h to g iff $\mathrm{E}u^i(h) > \mathrm{E}u^i(g)$, where E denotes expectation. This property continues to hold when we replace u^i by a linear transform of the form $\alpha u^i + \beta$, where α and β are constants with $\alpha > 0$. All these transforms are also called utility functions for i, and any one of them may be used rather than u^i in the payoff matrix.

Recall that a strictly competitive game is defined as a two-person game in which if one outcome is preferred to another by one player, the preference is reversed for the other. Since randomized strategies are admitted, this condition applies also to "mixed outcomes" (probability mixtures of pure outcomes). From this it may be seen that a two-person game is strictly competitive if and only if, for an appropriate choice of utility functions, the utility payoffs of the players sum to zero in each square of the matrix.

The case of TU coalitional games deserves particular attention. There is no problem if we assume fixed threats and continue to denominate the payoff in bread (see ii). But without fixed threats, the total amount of bread obtainable by a coalition S is a random variable depending on what players outside S do; since this is not denominated in utility, there is no justification for replacing it by its expectation. But if we do denominate payoffs in utility terms, then they cannot be directly transferred. The only way out of this quandary is to assume that the utility of bread is linear in the amount of bread (Aumann, 1960). We stress again that no such assumption is required in the fixed threat case.

ix. *Applications*. The very name of the book, *Theory of Games and Economic Behavior*, indicates its underlying preoccupation with the applications. Von Neumann had already mentioned *Homo Economicus* in his 1928 paper, but there were no specific economic applications there.

The method of von Neumann and Morgenstern has become the archetype of later applications of game theory. One takes an economic problem, formulates it as a game, finds the game-theoretic solution, then translates the solution back into economic terms. This is to be distinguished from the more usual methodology of economics and other social sciences, where the building of a formal model and a solution concept, and the application of the solution concept to the model, are all rolled into one.

Among the applications extensively treated in the book is voting. A qualitative feature that emerges is that many different weight-quota configurations have the same coalitional form; [5; 2, 3, 4] is the same as [2; 1, 1, 1]. Though obvious to the sophisticated observer when pointed out, this is not widely recognized; most people think that the player with weight 4 is considerably stronger than the others (Vinacke and

Arkoff, 1957). The Board of Supervisors of Nassau County operates by weighted voting; in 1964 there were six members, with weights of 31, 31, 28, 21, 2, 2, and a simple majority quota of 58 (Lucas, 1983, p. 188). Nobody realized that three members were totally without influence, that [58; 31, 31, 28, 21, 2, 2] = [2; 1, 1, 1, 0, 0, 0].

In a voting game, a winning coalition with no proper winning subsets is called *minimal winning* (mw). The game [q; w] is *homogeneous* if $w(S) = q$ for all minimal winning S; thus [3; 2, 1, 1, 1] is homogeneous, but [5; 2, 2, 2, 1, 1, 1] is not. A *decisive* voting game is one in which a coalition wins if and only if its complement loses; both the above games are decisive, but [3; 1, 1, 1, 1] is not. TU decisive homogeneous voting games have a stable set in which some mw coalition forms and divides the payoff in proportion to the weights of its members, leaving nothing for those outside. This is reminiscent of some parliamentary democracies, where parties in a coalition government get cabinet seats roughly in proportion to the seats they hold in parliament. But this fails to take into account that the actual number of seats held by a party may well be quite disproportional to its weight in a homogeneous representation of the game (when there is such a representation).

The book also considers issues of monopoly (or monopsony) and oligopoly. We have already pointed out that stable set theory concerns the endogenous emergence of social structure. In a market with one buyer (monopsonist) and two sellers (duopolists) where supply exceeds demand, the theory predicts that the duopolists will form a cartel to bargain with the monopsonist. The core, on the other hand, predicts cut-throat competition; the duopolists end up by selling their goods for nothing, with the entire consumer surplus going to the buyer.

This is a good place to point out a fundamental difference between the game-theoretic and other approaches to social science. The more conventional approaches take institutions as given, and ask where they lead. The game theoretic approach asks how the institutions came about, what led to them? Thus general equilibrium theory takes the idea of market prices for granted; it concerns itself with their existence and properties, calculating them, and so on. Game Theory asks, *why* are there market prices? How did they come about? Under what conditions will all traders trade at given prices?

Conventional economic theory has several approaches to oligopoly, including competition and cartelization. Starting with any particular one of these, it calculates what is implied in specific applications. Game Theory proceeds differently. It starts with the physical description of the situation only, making no institutional or doctrinal assumptions, then applies a solution concept and sees where it leads.

In a sense, of course, the doctrine is built into the solution concept; as we have seen, the core implies competition, the stable set cartelization. It is not that game theory makes no assumptions, but that the assumptions are of a more general, fundamental nature. The difference is like that between deriving the motion of the planets from Kepler's laws or from Newton's laws. Like Kepler's laws, which apply to the planets only, oligopoly theory applies to oligopolistic markets only. Newton's laws apply to the planets and also to apples falling from trees; stable sets apply to markets and also to voting.

To be sure, conventional economics is also concerned with the genesis of institutions, but on an informal, verbal, ad hoc level. In Game Theory, institutions like prices or cartels are outcomes of the formal analysis.

x. Games with a *continuum of pure strategies* were first considered by Ville (1938), who proved the minimax theorem for them, using an appropriate continuity condition. To guarantee the minimax (security) level, one may need to use a continuum of pure strategies, each with probability zero. An example due to Kuhn (1952) shows that in general one cannot guarantee anything even close to minimax using strategies with finite support. Ville's theorem was extended in the fifties to strategic equilibrium in non-strictly competitive games.

xi. *Computing* security levels, and strategies that will guarantee them, is highly non-trivial. The problem is equivalent to that of linear programming, and thus succumbed to the simplex method of George Dantzig (1951a, 1951b).

xii. The major advance in relevant mathematical methods during this period was *Kakutani's fixed point theorem* (1941). An abstract expression of the existence of equilibrium, it is the vital active ingredient of countless proofs in economics and game theory. Also instrumental in later work were Lyapounov's theorem on the range of a vector measure (1940) and von Neumann's selection theorem (1949).

1950–1960

The 1950s were a period of excitement in game theory. The discipline had broken out of its cocoon, and was testing its wings. Giants walked the earth. At Princeton, John Nash laid the groundwork for the general non-cooperative theory, and for cooperative bargaining theory; Lloyd Shapley defined the value for coalitional games, initiated the theory of stochastic games, co-invented the core with D. B. Gillies, and, together with John Milnor, developed the first game models with continua of players; Harold Kuhn worked on behaviour strategies and perfect recall; Al Tucker discovered the prisoner's dilemma; the Office of Naval Research was unstinting in its support. Three Game Theory conferences were held at Princeton, with the active participation of von Neumann and Morgenstern themselves. Princeton University Press published the four classic volumes of *Contributions to the Theory of Games*. The Rand Corporation, for many years to be a major centre of game theoretic research, had just opened its doors in Santa Monica. R. Luce and H. Raiffa (1957) published their enormously influential *Games and Decisions*. Near the end of the decade came the first studies of repeated games.

The major applications at the beginning of the decade were to tactical military problems: defense from missiles, Colonel Blotto, fighter-fighter duels, etc. Later the emphasis shifted to deterrence and cold war strategy, with contributions by political scientists like Kahn, Kissinger, and Schelling. In 1954, Shapley and Shubik published their seminal paper on the value of a voting game as an index of power. And in 1959 came Shubik's spectacular rediscovery of the core of a market in the writings of F. Y. Edgeworth (1881). From that time on, economics has remained by far the largest area of application of game theory.

i. An *equilibrium* (Nash, 1951) of a strategic game is a (pure or mixed) strategy profile in which each player's strategy maximizes his payoff given that the others are using their strategies.

Strategic equilibrium is without doubt the single game theoretic solution concept that is most frequently applied in economics. Economic applications include oligopoly, entry and exit, market equilibrium, search, location, bargaining, product quality, auctions, insurance, principal-agent, higher education, discrimination, public goods, what have you. On the political front, applications include voting, arms control, and inspection, as well as most international political models (deterrence, etc.). Biological applications of game theory all deal with forms of strategic equilibrium; they suggest a simple interpretation of equilibrium quite different from the usual overt rationalism (see *1970–1986*, i). We cannot even begin to survey all this literature here.

ii. *Stochastic and other dynamic games*. Games played in stages, with some kind of stationary time structure, are called *dynamic*. They include stochastic games, repeated games with or without complete information, games of survival (Milnor and Shapley, 1957; Luce and Raiffa, 1957; Shubik, 1959b) or ruin (Rosenthal and Rubinstein, 1984), recursive games (Everett, 1957), games with varying opponents (Rosenthal, 1979), and similar models.

This kind of model addresses the concerns we expressed above (*1930–1950*, vii) about the single play assumption. The present can only be understood in the context of the past and the future: "Know whence you came and where you are going" (Ethics of the Fathers III:1). Physically, current actions affect not only current payoff but also opportunities and payoffs in the future. Psychologically, too, we learn: past experience affects our current expectations of what others will do, and therefore our own actions. We also teach: our current actions affect others' future expectations, and therefore their future actions.

Two dynamic models—stochastic and repeated games—have been especially "successful." *Stochastic* games address the physical point, that current actions affect future opportunities. A strategic game is played at each stage; the profile of strategies determines both the payoff at that stage and the game to be played at the next stage (or a probability distribution over such games). In the strictly competitive case, with future payoff discounted at a fixed rate, Shapley (1953a) showed that stochastic games are determined; also, that they have optimal strategies that are stationary, in the sense that they depend only on the game being played (not on the history or even on the date). Bewley and Kohlberg (1976) showed that as the discount rate tends to 0 the value tends to a limit; this limit is the same as the limit, as $k \to \infty$, of the values of the k-stage games, in each of which the payoff is the mean payoff for the k stages. Mertens and Neyman (1981) showed that the value exists also in the undiscounted infinite stage game, when payoff is defined by the Cesaro limit (limit, as $k \to \infty$, of the average payoff in the first k stages). For an understanding of some of the intuitive issues in this work, see Blackwell and Ferguson (1968), which was extremely influential in the modern development of stochastic games.

The methods of Shapley, and of Bewley and Kohlberg, can be used to show that non-strictly competitive stochastic games with fixed discounts have equilibria in stationary strategies, and that when the discount tends to 0, these equilibria converge to a limit (Mertens, 1982). But unlike in the strictly competitive case, the payoff to this limit need not correspond to an equilibrium of the undiscounted game (Sorin, 1986b). It is not known whether undiscounted non-strictly competitive stochastic games need at all have strategic equilibria.

iii. *Repeated* games model the psychological, informational side of ongoing relationships. Phenomena like cooperation, altruism, trust, punishment, and revenge are predicted by the theory. These may be called "subjective informational" phenomena, since what is at issue is information about the behaviour of the players. Repeated games of incomplete information (*1960–1970*, ii) also predict "objective informational" phenomena such as secrecy, and signalling of substantive information. Both kinds of informational issue are quite different from the "physical" issues addressed by stochastic games.

Given a strategic game G, consider the game G^∞ each play of which consists of an infinite sequence of repetitions of G. At each stage, all players know the actions taken by all players at all previous stages. The payoff in G^∞ is some kind of average of the stage payoffs; we will not worry about exact definitions here.

Here we state only one basic result, known as the *Folk Theorem*. Call an outcome (payoff profile) x *feasible* in G if it is achievable by the all-player set when using a correlated randomizing device; i.e. is in the convex hull of the "pure" outcomes. Call it *strongly individually rational* if no player i can be prevented from achieving x^i by the other players, when they are randomizing independently; i.e. if $x^i \geqslant \min \max H^i(s)$, where the max is over i's strategies, and the min is over $(n-1)$-tuples of mixed strategies of the others. The Folk Theorem then says that the equilibrium outcomes in the repetition G^∞ *coincide with the feasible and strongly individually rational outcomes in the one-shot game G.*

The authorship of the Folk Theorem, which surfaced in the late Fifties, is obscure. Intuitively, the feasible and strongly individually rational outcomes are the outcomes that could arise in cooperative play. Thus the Folk Theorem points to a strong relationship between repeated and cooperative games. Repetition is a kind of enforcement mechanism; agreements are enforced by "punishing" deviators in subsequent stages.

iv. The *Prisoner's Dilemma* is a two-person non-zero sum strategic game with payoff matrix as depicted in Figure 1. Attributed to A. W. Tucker, it has deservedly attracted enormous attention; it is said that in the social

	f	g
f	3,3	0,4
g	4,0	1,1

Figure 1

psychology literature alone, over a thousand papers have been devoted to it.

One may think of the game as follows: Each player decides whether he will receive \$1000 or the other will receive \$3000. The decisions are simultaneous and independent, though the players may consult with each other before deciding.

The point is that ordinary rationality leads each player to choose the \$1000 for himself, since he is thereby better off *no matter what the other player does*. But the two players thereby get only \$1000 each, whereas they could have gotten \$3000 each if both had been "friendly" rather than "greedy."

The universal fascination with this game is due to its representing, in very stark and transparent form, the bitter fact that when individuals act for their own benefit, the result may well be disaster for all. This principle has dozens of applications, great and small, in everyday life. *People who fail to cooperate for their own mutual benefit are not necessarily foolish or irrational*; they may be acting perfectly rationally. The sooner we accept this, the sooner we can take steps to design the terms of social intercourse so as to encourage cooperation.

One such step, of very wide applicability, is to make available a mechanism for the enforcement of voluntary agreements. "Pray for the welfare of government, without whose authority, man would swallow man alive" (Ethics of the Fathers III:2). The availability of the mechanism is itself sufficient; once it is there, the players are naturally motivated to use it. If they can make an *enforceable* agreement yielding (3, 3), they would indeed be foolish to end up with (1, 1). It is this that motivates the definition of a cooperative game (*1930–1950*, i).

The above discussion implies that (g, g) is the unique strategic equilibrium of the prisoner's dilemma. It may also be shown that in any finite repetition of the game, all strategic equilibria lead to a constant stream of "greedy" choices by each player; but this is a subtler matter than the simple domination argument used for the one-shot case. In the infinite repetition, the Folk Theorem (iii) shows that (3, 3) is an equilibrium outcome; and indeed, there are equilibria that lead to a constant stream of "friendly" choices by each player. The same holds if we discount future payoff in the repeated game, as long as the discount rate is not too large (Sorin, 1986a).

R. Axelrod (1984) has carried out an experimental study of the repeated prisoner's dilemma. Experts were asked to write computer programmes for playing the game, which were matched against each other in a "tournament." At each stage, the game ended with a fixed (small) probability; this is like discounting. The most successful program in the tournament turned out to be a "cooperative" one: Matched against itself, it yields a constant stream of "friendly" choices; matched against others, it "punishes" greedy choices. The results of this experiment thus fit in well with received theoretical doctrine.

The design of this experiment is noteworthy because it avoids the pitfalls so often found in game experiments: lack of sufficient motivation and understanding. The experts chosen by Axelrod understood the game as well as anybody. Motivation was provided by the investment of their time, which was much more considerable than that of the average subject, and by the glory of a possible win over distinguished colleagues. Using computer programmes for strategies presaged important later developments (*1970–1986*, iv).

Much that is fallacious has been written on the one-shot prisoner's dilemma. It has been said that for the reasoning to work, pre-play communication between the players must be forbidden. This is incorrect. The players can communicate until they are blue in the face, and agree solemnly on (f, f); when faced with the actual decision, rational players will still choose g. It has been said that the argument depends on the notion of strategic equilibrium, which is open to discussion. This too is incorrect; the argument depends only on strong domination, i.e. on the simple proposition that people always prefer to get another $1000. "Resolutions" of the "paradox" have been put forward, suggesting that rational players will play f after all; that my choosing f has some kind of "mirror" effect that makes you choose it also. Worse than just nonsense, this is actually vicious, since it suggests that the prisoner's dilemma does not represent a real social problem that must be dealt with.

Finally, it has been said that the experimental evidence—Axelrod's and that of others—contradicts theory. This too is incorrect, since most of the experimental evidence relates to repeated games, where the friendly outcome is perfectly consonant with theory; and what evidence there is in one-shot games does point to a preponderance of "greedy" choices. It is true that in long finite repetitions, where the only equilibria are greedy, most experiments nevertheless point to the friendly outcome; but fixed finite repetitions are somewhat artificial, and besides, this finding, too, can be explained by theory (Neyman, 1985a; see *1970–1986*, iv).

v. We turn now to cooperative issues. A model of fundamental importance is the *bargaining problem* of Nash (1950). Formally, it is defined as a convex set C in the Euclidean plane, containing the origin in its interior. Intuitively, two players bargain; they may reach any agreement whose payoff profile is in C; if they disagree, they get nothing. Nash listed four *axioms*—conditions that a reasonable compromise solution might be expected to satisfy—such as symmetry and efficiency. He then showed that there is one and only one solution satisfying them, namely the point x in the non-negative part of C that maximizes the product x^1x^2. An appealing economic interpretation of this solution was given by Harsanyi (1956).

By varying the axioms, other authors have obtained different solutions to the bargaining problem, notably Kalai–Smorodinski (1975) and Maschler–Perles (1981). Like Nash's solution, each of these is characterized by a formula with an intuitively appealing interpretation.

Following work of A. Rubinstein (1982), K. Binmore (1987) constructed an explicit bargaining model which, when analyzed as a non-cooperative strategic game, leads to Nash's solution of the bargaining problem. This is an instance of a successful application of the "Nash program" (see *1930–1950*, vi). Similar constructions have been made for other solutions of the bargaining problem.

An interesting qualitative feature of the Nash solution is that it is very sensitive to risk aversion. A risk loving or risk neutral bargainer will get a better deal than a risk averse one; this is so even when there are no overt elements of risk in the situation, nothing random. The very willingness to take risks confers an advantage, though in the end no risks are actually taken.

Suppose, for example, that two people may divide \$600 in any way they wish; if they fail to agree, neither gets anything. Let their utility functions be $u^1(\$x) = x$ and $u^2(\$x) = \sqrt{x}$, so that 1 is risk neutral, 2 risk averse. Denominating the payoffs in utilities rather than dollars, we find that the Nash solution corresponds to a dollar split of \$400–\$200 in favour of the risk neutral bargainer.

This corresponds well with our intuitions. A fearful, risk averse person will not bargain well. Though there are no overt elements of risk, no random elements in the problem description, the bargaining itself constitutes a risk. A risk averse person is willing to pay, in terms of a less favourable settlement, to avoid the risk of the other side's being adamant, walking away, and so on.

vi. The *value* (Shapley, 1953b) is a solution concept that associates with each coalitional game v a unique outcome ϕv. Fully characterized by a set of axioms, it may be thought of as a reasonable compromise or arbitrated outcome, given the power of the players. Best, perhaps, is to think of it simply as an index of power, or what comes to the same thing, of social productivity.

It may be shown that Player i's value is given by

$$\phi^i v = (1/n!) \sum v^i(\mathrm{S}^i_R),$$

where $\sum$ ranges over all $n!$ orders on the set I of all players, S_R^i is the set of players up to and including i in the order R, and $v^i(S)$ is the *contribution* $v(S) - v(S\backslash i)$ of i to the coalition S; note that this implies linearity of ϕv in v. In words, $\phi^i v$ is i's mean contribution when the players are ordered at random; this suggests the social productivity interpretation, an interpretation that is reinforced by the following remarkable theorem (Young, 1985): Let ψ be a mapping from games v to efficient outcomes ψv, that is symmetric among the players in the appropriate sense. Suppose $\psi^i v$ depends only on the 2^{n-1} contributions $v^i(S)$, and monotonically so. Then ψ must be the value ϕ. In brief, if it depends on the contributions only, it's got to be the value, even though we don't assume linearity to start with.

An intuitive feel for the value may be gained from examples. The value of the 3-person voting game is (1/3, 1/3, 1/3), as is suggested by symmetry. This is not in the core, because {1, 2} can improve upon it. But so can {1, 3} and {2, 3}; starting from (1/3, 1/3, 1/3), the players might be well advised to leave things as they are (see *1930–1950*, iv). Differently viewed, the symmetric stable set predicts one of the three outcomes (1/2, 1/2, 0), (1/2, 0, 1/2), (0, 1/2, 1/2). Before the beginning of bargaining, each player may figure that his chances of getting into a ruling coalition are 2/3, and conditional on this, his payoff is 1/2. Thus his "expected outcome" is the value, though in itself, this outcome has no stability.

In the homogeneous weighted voting game [3; 2, 1, 1, 1], the value is (1/2, 1/6, 1/6, 1/6); the large player gets a disproportionate share, which accords with intuition: "l'union fait la force."

Turning to games of economic interest, we model the market with two sellers and one buyer discussed above (*1930–1950*, ix) by the TU weighted voting game [3; 2, 1, 1]. The core consists of the unique point (1, 0, 0), which means that the sellers must give their merchandise, for nothing, to the buyer. While this has clear economic meaning—cutthroat competition—it does not seem very reasonable as a compromise or an index of power. After all, the sellers do contribute something; without them, the buyer could get nothing. If one could be sure that the sellers will form a cartel to bargain with the buyer, a reasonable compromise would be (1/2, 1/4, 1/4). In fact, the value is (2/3, 1/6, 1/6), representing something between the cartel solution and the competitive one; a cartel is possible, but is not a certainty.

Consider next a market in two perfectly divisible and completely complementary goods, which we may call right and left gloves. There are four players; initially 1 and 2 hold one and two left gloves respectively, 3 and 4 hold one right glove each. In coalitional form, $v(1234) = v(234) = 2$, $v(ij) = v(12j) = v(134) = 1$, $v(S) = 0$ otherwise, where $i = 1, 2$, and

$j = 3, 4$. The core consists of (0, 0, 1, 1) only; that is, the owners of the left gloves must simply give away their merchandise, for nothing. This in itself seems strange enough. It becomes even stranger when one realizes that Player 2 could make the situation entirely symmetric (as between 1, 2 and 3, 4) simply by burning one glove, an action that he can take alone, without consulting anybody.

The value can never suffer from this kind of pathological breakdown in monotonicity. Here $\phi v = (1/4, 7/12, 7/12, 7/12)$, which nicely reflects the features of the situation. There *is* an oversupply of left gloves, and 3 and 4 do benefit from it. Also 2 benefits from it; he always has the option of nullifying it, but he can also use it (when he has an opportunity to strike a deal with both 3 and 4). The brunt of the oversupply is thus born by 1 who, unlike 2, cannot take measures to correct it.

Finally, consider a market with 2,000,001 players, 1,000,000 holding one right glove each, and 1,000,001 holding one left glove each. Again, the core stipulates that the holders of the left gloves must all give away their merchandise, for nothing. True, there *is* a slight oversupply of left gloves; but one would hardly have imagined so drastic an effect from one single glove out of millions. The value, too, takes the oversupply into account, but not in such an extreme form; altogether, the left-glove holders get about 499,557 pairs, the right about 500,443 (Shapley and Shubik, 1969b). This is much more reasonable, though the effect is still surprisingly large: The short side gains an advantage that amounts to almost a thousand pairs.

The value has many different characterizations, all of them intuitively meaningful and interesting. We have already mentioned Shapley's original axioms, the value formula, and Young's characterization. To them must be added Harsanyi's (1959) dividend characterization, Owen's (1972) fuzzy coalition formula, Myerson's (1977) graph approach, Dubey's (1980) diagonal formula, the potential of Hart and Mas-Colell (1989), the reduced game axiomatization by the same authors, and Roth's (1977) formalization of Shapley's (1953b) idea that the value represents the utility to the players of playing a game. Moreover, because of its mathematical tractability, the value lends itself to a far greater range of applications than any other cooperative solution concept. And in terms of general theorems and characterizations for wide classes of games and economies, the value has a greater range than *any* other solution concept, bar none.

Previously (*1930–1950*, iii), we compared solution concepts of games to indicators of distributions, like mean and median. In fact the value is in many ways analogous to the mean, whereas the median corresponds to something like the core, or to core-like concepts such as the nucleolus

(*1960–1970*, iv). Like the core, the median has an intuitively transparent and compelling definition (the point that cuts the distribution exactly in half), but lacks an algebraically neat formula; and like the value, the mean has a neat formula whose intuitive significance is not entirely transparent (though through much experience from childhood on, many people have acquired an intuitive feel for it). Like the value, the mean is linear in its data; the core, nucleolus, and median are not. Both the mean and the value are very sensitive to their data: change one datum by a little, and the mean (or value) will respond in the appropriate direction; neither the median nor the core is sensitive in this way: one can change the data in wide ranges without affecting the median (or core) at all. On the other hand, the median can suddenly jump because of a moderate change in just one datum; thus the median of 1,000,001 zeros and 1,000,000 ones is 0, but jumps to 1 if we change just one datum from 0 to 1. We have already seen that the core may behave similarly, but the mean and the value cannot. Both the mean and the value are mathematically very tractable, resulting in a wide range of applications, both theoretical and practical; the median and core are less tractable, resulting in a narrower (though still considerable) range of applications.

The first extensive applications of the value were to various voting games (Shapley and Shubik, 1954). The key observation in this seminal paper was that the value of a player equals his probability of *pivoting*—turning a coalition from losing to winning—when the players are ordered at random. From this there has grown a very large literature on voting games. Other important classes of applications are to market games (*1960–1970*, v) and political-economic games (e.g. Aumann and Kurz, 1977; Neyman, 1985b).

vii. *Axiomatics*. The Shapley value and Nash's solution to the bargaining problem are examples of the axiomatic approach. Rather than defining a solution concept directly, one writes down a set of conditions for it to satisfy, then sees where they lead. In many contexts, even a relatively small set of fairly reasonable conditions turn out to be self-contradictory; there is no concept satisfying all of them. The most famous instance of this is Arrow's (1951) impossibility theorem for social welfare functions, which is one of the earliest applications of axiomatics in the social sciences.

It is not easy to pin down precisely what is meant by "the axiomatic method." Sometimes the term is used for any formal deductive system, with undefined terms, assumptions, and conclusions. As understood today, all of game theory and mathematical economics fits that definition. More narrowly construed, an axiom system is a small set of individually transparent conditions, set in a fairly general and abstract framework, which when taken together have far-reaching implications. Examples are Euclid's axioms for geometry, the Zermelo–Fraenkel axioms for set theory, the conditions on multiplication that define a group, the conditions on open sets that define a topological space, and the conditions on preferences that define utility and/or subjective probability.

Game theoretic solution concepts often have both direct and axiomatic characterizations. The direct definition applies to each game separately, whereas most axioms deal with relationships between games. Thus the formula for the Shapley value ϕv enables one to calculate it without referring to any game other than v. But the axioms for ϕ concern relationships between games; they say that if the values of certain games are so and so, then the values of certain other, related games must be such and such. For example, the additivity axiom is $\phi(v + w) = \phi v + \phi w$. This is analogous to direct vs. axiomatic approaches to integration. Direct approaches such as limit of sum work on a single function; axiomatic approaches characterize the integral as a linear operator on a *space* of functions. (Harking back to the discussion at (vi), we note that the axioms for the value are quite similar to those for the integral, which in turn is closely related to the mean of a distribution.)

Shapley's value and the solutions to the bargaining problem due to Nash (1950), Kalai–Smorodinski (1975) and Maschler–Perles (1981) were originally conceived axiomatically, with the direct characterization coming afterwards. In other cases the process was reversed; for example, the nucleolus, NTU Shapley value, and NTU Harsanyi value were all axiomatized only years after their original direct definition (see *1960–1970*). Recently the core, too, has been axiomatized (Peleg, 1985, 1986).

Since axiomatizations concern relations between different games, one may ask why the players of a given game should be concerned with other games, which they are not playing. This has several answers. Viewed as an indicator, a solution of a game doesn't tell us much unless it stands in some kind of coherent relationship to the solutions of other games. The ratings for a rock climb tell you something if you have climbed other rocks whose ratings you know; topographic maps enable you to take in a situation at a glance if you have used them before, in different areas. If we view a solution as an arbitrated or imposed outcome, it is natural to expect some measure of consistency from an arbitrator or judge. Indeed, much of the law is based on precedent, which means relating the solution of the given "game" to those of others with known solutions. Even when viewing a solution concept as a norm of actual behaviour, the very word "norm" implies that we are thinking of a function on classes of games rather than of a single game; outcomes are largely based on mutual expectations, which are determined by previous experience with other games, by "norms."

Axiomatizations serve a number of useful purposes. First, like any other alternative characterization, they shed additional light on a concept, enable us to "understand" it better. Second, they underscore and clarify important similarities between concepts, as well as differences between them. One example of this is the remarkable "reduced game property" or "consistency principle," which is associated in various different forms with just about every solution concept, and plays a key role in many of the axiomatizations (see *1970–1986*, vi). Another example consists of the axiomatizations of the Shapley and Harsanyi NTU values. Here the axioms are exact analogues, except that in the Shapley case they refer to payoff profiles, and in the Harsanyi case to 2^n-tuples of payoff profiles, one for each of the 2^n coalitions (Hart, 1985a). This underscores the basic difference in outlook between those two concepts: The Shapley value assumes that the all-player coalition eventually forms, the intermediate coalitions being important only for bargaining chips and threats, whereas the Harsanyi value takes into account a real possibility of the intermediate coalitions actually forming.

Last, an important function of axiomatics relates to "counterintuitive examples," in which a solution concept yields outcomes that seem bizarre; e.g. the cores of some of the games discussed above in (vi). Most axioms appearing in axiomatizations do seem reasonable on the face of it, and many of them are in fact quite compelling. The fact that a relatively small selection of such axioms is often categoric (determines a unique solution concept), and that different such selections yield different answers, implies that all together, these reasonable-sounding axioms are contradictory. This, in turn, implies that any one solution concept will necessarily violate at least some of the axioms that are associated with other solution concepts; thus if the axioms are meant to represent intuition, counter-intuitive examples are inevitable.

In brief, axiomatics underscores the fact that a "perfect" solution concept is an unattainable goal, a *fata morgana*; there is something "wrong" —some quirk—with every one. Any given kind of counterintuitive example can be eliminated by an appropriate choice of solution concept, but only at the cost of another quirk turning up. Different solution concepts can therefore be thought of as results of choosing not only which properties one likes, but also which examples one wishes to avoid.

1960–1970

The Sixties were a decade of growth. Extensions such as games of incomplete information and NTU coalitional games made the theory much more widely applicable. The fundamental underlying concept of common knowledge was formulated and clarified. Core theory was extensively developed and applied to market economies; the bargaining set and related concepts such as the nucleolus were defined and investigated; games with many players were studied in depth. The discipline expanded geographically, outgrowing the confines of Princeton and Rand; important centres of research were established in Israel, Germany, Belgium and the Soviet Union. Perhaps most important was the forging of a strong, lasting relationship with mathematical economics and economic theory.

i. *NTU coalitional games and NTU value*. Properly interpreted, the coalitional form (*1930–1950*, ii) applies both to TU and to NTU games; nevertheless, for many NTU applications one would like to describe the opportunities available to each coalition more faithfully than can be done with a single number. Accordingly, define a game in *NTU coalitional form* as a function that associates with each coalition S a set V(S) of S-tuples of real numbers (functions from S to $\mathbb{R}$). Intuitively, V(S) represents the set of payoff S-tuples that S can achieve. For example, in an exchange economy, V(S) is the set of utility S-tuples that S can achieve when its members trade among themselves only, without recourse to agents outside of S. Another example of an NTU coalitional game is Nash's bargaining problem (*1950–1960*, iii), where one can take $V(\{1,2\}) = C$, $V(1) = \{0\}$, $V(2) = \{0\}$.

The definitions of stable set and core extend straightforwardly to NTU coalitional games, and these solution concepts were among the first to be investigated in that context (Aumann and Peleg, 1960; Peleg, 1963a; Aumann, 1961). The first definitions of NTU value were proposed by Harsanyi (1959, 1963), but they proved difficult to apply. Building on Harsanyi's work, Shapley (1969) defined a value for NTU games that has proved widely applicable and intuitively appealing.

For each profile λ of non-negative numbers and each outcome x, define the *weighted outcome* λx by $(\lambda x)^i = \lambda^i x^i$. Let $v_\lambda(S)$ be the maximum total weight that the coalition S can achieve,

$$v_\lambda(S) = \max\left\{ \sum_{i \in S} \lambda^i x^i, x \in V(S) \right\}.$$

Call an outcome x an *NTU value* of V if $x \in V(N)$ and there exists a weight profile λ with $\lambda x = \phi v_\lambda$; in words, if x is feasible and corresponds to the value of one of the coalitional games v_λ.

Intuitively, $v_\lambda(S)$ is a numerical measure of S's total worth and hence $\phi^i v_\lambda$ measures i's social productivity. The weights λ^i are chosen so that the resulting value is feasible; an infeasible result would indicate that some people are overrated (or underrated), much like an imbalance between supply and demand indicates that some goods are overpriced (or underpriced).

The NTU value of a game need not be unique. This may at first sound strange, since unlike stability concepts such as the core, one might expect an "index of social productivity" to be unique. But perhaps it is not so strange when one reflects that even a person's net worth depends on the prevailing (equilibrium) prices, which are not uniquely determined by the exogenous description of the economy.

The Shapley NTU value has been used in a very wide variety of economic and political-economic applications. To cite just one example, the Nash bargaining problem has a unique NTU value, which coincides with Nash's solution. For a partial bibliography of applications, see the references of Aumann (1985).

We have discussed the historical importance of TU as pointing the way for NTU results (*1930–1950*, vi). There is one piquant case in the reverse direction. Just as positive results are easier to obtain for TU, negative results are easier for NTU. Non-existence of stable sets was first discovered in NTU games (Stearns, 1964), and this eventually led to Lucas's famous example (1969) of non-existence for TU.

ii. *Incomplete information*. In 1957, Luce and Raiffa wrote that a fundamental assumption of game theory is that "each player ... is fully aware of the rules of the game and the utility functions of each of the players ... this is a serious idealization which only rarely is met in actual situations" (p. 49). To deal with this problem, John Harsanyi (1967) constructed the theory of games of incomplete information (sometimes called differential or asymmetric information). This major conceptual breakthrough laid the theoretical groundwork for the great blooming of information economics that got under way soon thereafter, and that has become one of the major themes of modern economics and game theory.

For simplicity, we confine attention to strategic form games in which each player has a fixed, known set of strategies, and the only uncertainty is about the utility functions of the other players; these assumptions are removable. Bayesian rationality in the tradition of Savage (1954) dictates that all uncertainty can be made explicit; in particular, each player has a personal probability distribution on the possible utility (payoff) functions of the other player. But these distributions are not sufficient to describe the situation. It is not enough to specify what each player thinks about the other's payoffs; one must also know what he thinks they think about his (and each others') payoffs, what he thinks they think he thinks about their payoffs, and so on. This complicated infinite regress would appear to make useful analysis very difficult.

To cut this Gordian knot, Harsanyi postulated that each player may be one of several *types*, where a type determines both a player's own utility function and his personal probability distribution on the types of the other players. Each player is postulated to know his own type only. This enables him to calculate what he thinks the other players' types—and therefore their utilities—are. Moreover, his personal distribution on their types also enables him to calculate what he thinks they think about his type, and therefore about his utility. The reasoning extends indefinitely, and yields the infinite regress discussed above *as an outcome*.

Intuitively, one may think of a player's type as a possible state of mind, which would determine his utility as well as his distribution over others' states of mind. One need not assume that the number of states of mind (types) is finite; the theory works as well for, say, a continuum of types. But even with just two players and two types for each player, one gets a non-trivial infinite string of beliefs about utilities, beliefs about beliefs, and so on.

A model of this kind—with players, strategies, types, utilities, and personal probability distributions—is called an *I-game* (incomplete information game). A *strategic equilibrium* in an I-game consists of a strategy for each *type* of each player, which maximizes that type's expected payoff given the strategies of the other players' types.

Harsanyi's formulation of I-games is primarily a device for thinking about incomplete information in an orderly fashion, bringing that wild, bucking infinite regress under conceptual control. An (incomplete) analogy is to the strategic form of a game, a conceptual simplification without which it is unlikely that game theory would have gotten very far. Practically speaking, the strategic form of a particular game such as chess is totally unmanageable, one can't even begin to write it down. The advantage of the strategic form is that it is a comparatively simple formulation, mathematically much simpler than the extensive form; it enables one to formulate and calculate examples, which suggest principles that can be formulated and proved as general theorems. All this would be much more difficult—probably unachievable—with the extensive form; one would be unable to see the forest for the trees. A similar relationship holds between Harsanyi's I-game formulation and direct formulations in terms of beliefs about beliefs. (Compare the discussion of perspective made in connection with the coalitional form (*1930–1950*, i). That situation is somewhat different, though, since in going to the coalitional form, substantive information is lost. Harsanyi's formulation of I-games loses no information; it is a more abstract and simple—and hence transparent and workable—formulation of the same data as would be contained in an explicit description of the infinite regress.)

Harsanyi called an I-game *consistent* if all the personal probability distributions of all the types are derivable as posteriors from a single prior distribution p on all n-tuples of types. Most applications of the theory have assumed consistency. A consistent I-game is closely related to the ordinary strategic game (*C-game*) obtained from it by allowing 'nature' to choose an n-tuple of types at random according to the distribution p, then informing each player of his type, and then playing the I-game as before. In particular, the strategic equilibria of a consistent I-game are essentially the same as the strategic equilibria of the related C-game. In the cooperative theory, however, an I-game is rather different from the related C-game, since binding agreements can only be made after the players know their types. Bargaining and other cooperative models have been treated in the incomplete information context by Harsanyi and Selten (1972), Wilson (1978), Myerson (1979, 1984), and others.

In a repeated game of incomplete information, the same game is played again and again, but the players do not have full information about it; for example, they may not know the others' utility functions. The actions of the players may implicitly reveal private information, e.g. about preferences; this may or may not be advantageous for them. We have seen (*1950–1960*, iii) that repetition may be viewed as a paradigm for cooperation. Strategic equilibria of repeated games of incomplete information may be interpreted as a subtle bargaining process, in which the players gradually reach wider and wider agreement, developing trust for each other while slowly revealing more and more information (Hart 1985b).

iii. *Common knowledge*. Luce and Raiffa, in the statement quoted at the beginning of (ii), missed a subtle but important point. It is not enough that each player be fully aware of the rules of the game and the utility functions of the players. Each player must also be aware of this fact, i.e. of the awareness of all the players; moreover, each player must be aware that each player is aware that each player is aware, and so on ad infinitum. In brief, the awareness of the description of the game by all players must be a part of the description itself.

There is evidence that game theorists had been vaguely cognizant of the need for some such requirement ever since the late Fifties or early Sixties; but the first to give a clear, sharp formulation was the philosopher D. K. Lewis (1969). Lewis defined an event as *common knowledge* among a set of agents if all know it, all know that all know it, and so on ad infinitum.

The common knowledge assumption underlies all of game theory and much of economic theory. Whatever be the model under discussion, whether complete or incomplete information, consistent or inconsistent, repeated or one-shot, cooperative or non-cooperative, the model itself must be assumed common knowledge; otherwise the model is insufficiently specified, and the analysis incoherent.

iv. *Bargaining set, kernel, nucleolus*. The core excludes the unique symmetric outcome (1/3, 1/3, 1/3) of the three-person voting game, because any two-person coalition can improve upon it. Stable sets (*1930–1950*, v) may be seen as a way of expressing our intuitive discomfort with this exclusion. Another way is the bargaining set (Davis and Maschler, 1967). If, say, 1 suggests (1/2, 1/2, 0) to replace (1/3, 1/3, 1/3), then 3 can suggest to 2 that he is as good a partner as 1; indeed, 3 can even offer 2/3 to 2, still leaving himself with the 1/3 he was originally assigned. Formally, if we call (1/2, 1/2, 0) an *objection* to (1/3, 1/3, 1/3), then (0, 2/3, 1/3) is a *counter-objection*, since it yields to 3 at least as much as he was originally assigned, and yields to 3's partners in the counter-objection at least as much as they were assigned either originally or in the objection. In brief, the counter-objecting player tells the objecting one, "I can maintain my level of payoff and that of my partners, while matching your offers to players we both need." An imputation is in the core iff there is no objection to it. It is in the *bargaining set* iff there is no *justified* objection to it, i.e. one that has no counter-objection.

Like the stable sets, the bargaining set includes the core (dominating and objecting are essentially the same). Unlike the core and the set of stable sets, the bargaining set is for TU games never empty (Peleg, 1967). For NTU it may be empty (Peleg, 1963b); but Asscher (1976) has defined a non-empty variant; see also Billera (1970a).

Crucial parameters in calculating whether an imputation x is in the bargaining set of v are the *excesses* $v(S) - x(S)$ of coalitions S w.r.t. x, which measure the ability of members of S to use x in an objection (or counter-objection). Not, as is often wrongly assumed, because the initiator of the objection can assign the excess to himself while keeping his partners at their original level, but for precisely the opposite reason: because he can parcel out the excess to his partners, which makes counter-objecting more difficult.

The excess is so ubiquitous in bargaining set calculations that it eventually took on intuitive significance on its own. This led to the formulation of two additional solution concepts: the *kernel* (Davis and Maschler, 1965), which is always included in the bargaining set but is often much smaller, and the *nucleolus* (Schmeidler, 1969), which always consists of a single point in the kernel.

To define the nucleolus, choose first all those imputations x whose maximum excess (among the 2^n excesses $v(S) - x(S)$) is minimum (among all imputations). Among the resulting imputations, choose next those whose second largest excess is minimum, and so on. Schmeidler's theorem asserts that by the time we have gone through this procedure 2^n times, there is just one imputation left.

We have seen that the excess is a measure of a coalition's 'manoeuvring ability'; in these terms the greatest measure of stability, as expressed by the nucleolus, is reached when all coalitions have manoeuvring ability as nearly alike as possible. An alternative interpretation of the excess is as a measure of S's total dissatisfaction with x, the volume of the cry that S might raise against x. In these terms, the nucleolus suggests that the final accommodation is determined by the loudest cry against it. Note that the *total* cry is determining, not the average cry; a large number of moderately unhappy citizens can be as potent a force for change as a moderate number of very unhappy ones. Variants of the nucleolus that use the average excess miss this point.

When the core is non-empty, the nucleolus is always in it. The nucleolus has been given several alternative characterizations, direct (Kohlberg, 1971, 1972) as well as axiomatic (Sobolev, 1975). The kernel was axiomatically characterized by Peleg (1986), and many interesting relationships have been found between the bargaining set, core, kernel, and nucleolus (e.g. Maschler, Peleg and Shapley, 1979). There is a large body of applications, of which we here cite just one: In a decisive weighted voting game, the nucleolus constitutes a set of weights (Peleg, 1968). Thus the nucleolus may be thought of as a natural generalization of "voting weights" to arbitrary games. (We have already seen that value and weights are quite different: see *1950–1960*, vi.)

v. *The Equivalence Principle*. Perhaps the most remarkable single phenomenon in game and economic theory is the relationship between the price equilibria of a competitive market economy, and all but one of the major solution concepts for the corresponding game (the one exception is the stable set, about which more below). By a "market economy" we here mean a pure exchange economy, or a production economy with constant returns.

We call an economy "competitive" if it has many agents, each individual one of whom has too small an endowment to have a significant effect. This has been modelled by three approaches. In the *asymptotic approach*, one lets the number of agents tend to infinity, and shows that in an appropriate sense, the solution concept in question—core, value, bargaining set, or strategic equilibrium—tends to the set of competitive allocations (those corresponding to price equilibria). In the *continuum approach*, the agents constitute a (non-atomic) continuum, and one shows that the solution concept in question actually equals the set of competitive allocations. In the *non-standard* approach, the agents constitute a non-standard model of the integers in the sense of Robinson (1970), and again one gets equality. Both the continuum and the non-standard approaches require extensions of the theory to games with infinitely many players; see vi.

Intuitively, the equivalence principle says that the institution of market prices arises naturally from the basic forces at work in a market, (almost) no matter what we assume about the way in which these forces work. Compare (*1930–1950*, ix).

For simplicity in this section, unless otherwise indicated, the terms "core," "value," etc., refer to the limiting case. Thus "core" means the limit of the cores of the finite economies, or the core of the continuum economy, or of the non-standard economy.

For the core, the asymptotic approach was pioneered by Edgeworth (1881), Shubik (1959a) and Debreu and Scarf (1963). Anderson (1986) is an excellent survey of the large literature that ensued. Early writers on the continuum approach included Aumann (1964) and Vind (1965); the non-standard approach was developed by Brown and Robinson (1975). Except for Shubik's, all these contributions were NTU. See the entry on CORE. After a twenty-year courtship, this was the honeymoon of game theory and mathematical economics, and it is difficult to convey the palpable excitement of those early years of intimacy between the two disciplines.

Some early references for the value equivalence principle, covering both the asymptotic and continuum approaches, were listed above (*1930–1950*, vi). For the non-standard approach, see Brown and Loeb (1976). Whereas the core of a competitive economy equals the set of *all* competitive allocations, this holds for the value only when preferences are smooth (Shapley, 1964; Aumann and Shapley, 1974; Aumann 1975; Mas-Colell, 1977). Without smoothness, every value allocation is competitive, but not every competitive allocation need be a value allocation. When preferences are kinky (non-differentiable utilities), the core is often quite large, and then the value is usually a very small subset of the core; it gives much more information. In the TU case, for example, the value is always a single point, even when the core is very large. Moreover, it occupies a central position in the core (Hart, 1980; Tauman, 1981; Mertens, 1987); in particular, when the core has a centre of symmetry, the value is that centre of symmetry (Hart, 1977a).

For example, suppose that in a glove market (*1950–1960*, vi), the number (or measure) of left-glove holders equals that of right-glove holders. Then at a price equilibrium, the price ratio between left and right gloves may be anything between 0 and ∞ (inclusive!). Thus the left-glove holders may end up giving away their merchandise for nothing to the right-glove holders, or the other way around, or anything in between. The same, of course, holds for the core. But the value prescribes precisely equal prices for right and left gloves.

It should be noted that in a finite market, the core contains the competitive allocations, but usually also much more. As the number of agents increases, the core "shrinks," in the limit leaving only the competitive allocations. This is not so for the value; in finite markets, the value allocations may be disjoint from the core, and a fortiori from the competitive allocations (*1950–1960*, vi).

We have seen (*1930–1950*, iv) that the core represents a very strong and indeed not quite reasonable notion of stability. It might therefore seem perhaps not so terribly surprising that it shrinks to the competitive allocations. What happens, one may ask, when one considers one of the more reasonable stability concepts that are based on domination, such as the bargaining set or the stable sets?

For the bargaining set of TU markets, an asymptotic equivalence theorem was established by Shapley and Shubik in the mid-Seventies, though it was not published until 1984. Extending this result to NTU, to the continuum, or to both seemed difficult. The problems were conceptual as well as mathematical; it was difficult to give a coherent formulation. In 1986, Shapley presented the TU proof at a conference on the equivalence principle that took place at Stony Brook. A. Mas-Colell, who was in the audience, recognized the relevance of several results that he had obtained in other connections; within a day or two he was able to formulate and prove the equivalence principle for the bargaining set in NTU continuum economies (Mas-Colell, 1988). In particular, this implies the core equivalence principle; but it is a much stronger and more satisfying result.

For the strategic equilibrium the situation had long been less satisfactory, though there were results (Shubik, 1973; Dubey and Shapley, 1994). The difficulty was in constructing a satisfactory strategic (or extensive) model of exchange. Very recently Douglas Gale (1986) provided such a model and used it to prove a remarkable equivalence theorem for strategic equilibria in the continuum mode.

The one notable exception to the equivalence principle is the case of stable sets, which predict the formation of cartels even in fully competitive economies (Hart, 1974). For example, suppose half the agents in a continuum initially hold 2 units of bread each, half initially hold 2 units of cheese, and the utility functions are concave, differentiable, and symmetric (e.g., $u(x, y) = \sqrt{x} + \sqrt{y}$). There is then a unique price equilibrium, with equal prices for bread and cheese. Thus each agent ends up with one piece of bread and one piece of cheese; this is also the unique point in the core and in the bargaining set, and the unique NTU value. But stable set theory predicts that the cheese holders will form a cartel, the bread holders will form a cartel, and these two cartels will bargain with each other as if they were individuals. The upshot will depend on the bargaining, and may yield an outcome that is much better for one side than for the other. Thus at each point of the unique stable set with the full symmetry of the game, each agent on each side gets as much as each other agent on that side; but these two amounts depend on the bargaining, and may be quite different from each other.

In a sense, the failure of stable set theory to fall into line makes the other results even more impressive. It shows that there isn't some implicit tautology lurking in the background, that the equivalence principle makes a substantive assertion.

In the *Theory of Games*, von Neumann and Morgenstern (1944) wrote that

> when the number of participants becomes really great, some hope emerges that the influence of every particular participant will become negligible ... These are, of course, the classical conditions of "free competition" ... The current assertions concerning free competition appear to be very valuable surmises and inspiring anticipations of results. But they are not results, and it is scientifically unsound to treat them as such.

One may take the theorems constituting the equivalence principle as embodying precisely this kind of "result." Yet it is interesting that Morgenstern himself, who died in 1977, never became convinced of the validity of the equivalence principle; he thought of it as mathematically correct but economically wrongheaded. It was his firm opinion that economic agents organize themselves into coalitions, that perfect competition is a fiction, and that stable sets explain it all. The greatness of the man is attested to by the fact that though scientifically opposed to the equivalence principle, he gave generous support, both financial and moral, to workers in this area.

vi. *Many players*. The preface to *Contributions to the Theory of Games* I (Kuhn and Tucker, 1950) contains an agenda for future research that is remarkable in that so many of its items—computation of minimax, existence of stable sets, n-person value, NTU games, dynamic games—did in fact become central in subsequent work. Item 11 in this agenda reads, "establish significant asymptotic properties of n-person games, for large n." We have seen (v) how this was realized in the equivalence principle for large economies. But actually, political game models with many players are at least as old as economic ones, and may be older. During the early Sixties, L. S. Shapley, working alone and with various collaborators, wrote a series of seven memoranda at the Rand Corporation under the generic title "Values of Large Games," several of which explored models of large elections, using the asymptotic and the continuum approaches. Among these were models which had both "atoms"—players who are significant as individuals—and an 'ocean' of individually insignificant players. One example of this is a corporation with many small stockholders and a few large stockholders; see also Milnor and Shapley (1978). "Mixed" models of this kind—i.e., with an ocean as well as atoms—have been explored in economic as well as political contexts using various solution notions, and a large literature has developed. The core of mixed markets has been studied by Drèze, Gabszewicz and Gepts (1969), Gabszewicz and Mertens (1971), Shitovitz (1973) and many others. For the nucleolus of "mixed" voting games, see Galil (1974). Among the studies of values of mixed games are Hart (1973), Fogelman and Quinzii (1980), and Neyman (1987).

Large games in which *all* the players are individually insignificant—*non-atomic* games—have also been studied extensively. Among the early contributions to value theory in this connection are Kannai (1966), Riker and Shapley (1968), and Aumann and Shapley (1974). The subject has proliferated greatly, with well over a hundred contributions since 1974, including theoretical contributions as well as economic and political applications.

There are also games with infinitely many players in which *all* the players are atoms, namely games with a denumerable infinity of players. Again, values and voting games loom large in this literature. See, e.g., Shapley (1962), Artstein (1972) and Berbee (1981).

vii. *Cores of finite games and markets*. Though the core was defined as an independent solution concept by Gillies and Shapley already in the early Fifties, it was not until the Sixties that a significant body of theory was developed around it. The major developments centre around conditions for the core to be non-empty; gradually it came to be realized that such conditions hold most naturally and fully when the game has an "economic" rather than a "political" flavour, when it may be thought of as arising from a market economy.

The landmark contributions in this area were the following: the Gale–Shapley 1962 paper on the core of a marriage market; the work of

Bondareva (1963) and Shapley (1967) on the balancedness condition for the non-emptiness of the core of a TU game; Scarf's 1967 work on balancedness in NTU games; the work of Shapley and Shubik (1969a) characterizing TU market games in terms of non-emptiness of the core; and subsequent work, mainly associated with the names of Billera and Bixby (1973), that extended the Shapley–Shubik condition to NTU games. Each of these contributions was truly seminal, in that it inspired a large body of subsequent work.

Gale and Shapley (1962) asked whether it is possible to match m women with m men so that there is no pair consisting of an unmatched woman and man who prefer each other to the partners with whom they were matched. The corresponding question for homosexuals has a negative answer: the preferences of four homosexuals may be such that no matter how they are paired off, there is always an unmatched pair of people who prefer each other to the person with whom they were matched. This is so, for example, if the preferences of a, b, and c are cyclic, whereas d is lowest in all the others' scales. But for the heterosexual problem, Gale and Shapley showed that the answer is positive.

This may be stated by saying that the appropriately defined NTU coalitional game has a non-empty core. Gale and Shapley proved not only the non-emptiness but also provided a simple algorithm for finding a point in it.

This work has spawned a large literature on the cores of discrete market games. One fairly general recent result is Kaneko and Wooders (1982), but there are many others. A fascinating application to the assignment of interns to hospitals has been documented by Roth (1984). It turns out that American hospitals, after fifty years of turmoil, finally developed in 1950 a method of assignment that is precisely a point in the core.

We come now to general conditions for the core to be non-empty. Call a TU game v *superadditive at* a coalition U if $v(\mathrm{U}) \geqslant \sum_j v(\mathrm{S}_j)$ for any partition of U into disjoint coalition S_j. This may be strengthened by allowing partitions of U into disjoint "part-time" coalitions $\theta\mathrm{S}$, interpreted as coalitions S operating during a proportion θ of the time $(0 \leqslant \theta \leqslant 1)$. Such a partition is therefore a family $\{\theta_j \mathrm{S}_j\}$, where the total amount of time that each player in U is employed is exactly 1; i.e., where $\sum_j \theta_j \chi_{\mathrm{S}_j} = \chi_{\mathrm{U}}$, where χ_{S} is the indicator function of S. If we think of $v(\mathrm{S})$ as the revenue that S can generate when operating full-time, then the part-time coalition $\theta\mathrm{S}$ generates $\theta v(\mathrm{S})$. Superadditivity at U for part-time coalitions thus means that

$$\sum_j \theta_j \chi_{\mathrm{S}_j} = \chi_{\mathrm{U}} \text{ implies } v(\mathrm{U}) \geqslant \sum_j \theta_j v(\mathrm{S}_j).$$

A TU game v obeying this condition for U = I is called *balanced*; for all U, *totally balanced*.

Intuitively, it is obvious that a game with a non-empty core must be superadditive at I; and once we have the notion of part-time coalitions, it is only slightly less obvious that it must be balanced. The converse was established (independently) by Bondareva (1963) and Shapley (1967). Thus a *TU game has a non-empty core if and only if it is balanced*.

The connection between the core and balancedness (generalized superadditivity) led to several lines of research. Scarf (1967) extended the notion of balancedness to NTU games, then showed that every balanced NTU game has a non-empty core. Unlike the Bondareva–Shapley proof, which is based on linear programming methods, Scarf's proof was more closely related to fixed-point ideas. Eventually, Scarf realized that his methods could be used actually to prove Brouwer's fixed-point theorem, and moreover, to develop effective algorithms for approximating fixed points. This, in turn, led to the development of algorithms for approximating competitive equilibria of economies (Scarf, 1973), and to a whole area of numerical analysis dealing with the approximation of fixed points.

An extension of the Bondareva–Shapley result to the NTU case that is different from Scarf's was obtained by Billera (1970a).

Another line of research that grew out of balancedness deals with characterizing markets in purely game-theoretic terms. When can a given coalitional game v be expressed as a market game (*1930–1950*, ii)? The Bondareva–Shapley theorem implies that market games have non-empty cores, and this also follows from the fact that outcomes corresponding to competitive equilibria are always in the core. Since a subgame of a market game is itself a market game, it follows that *for v to be a market game, it is necessary that it and all its subgames have non-empty cores*, i.e., that the game be totally balanced. (A *subgame* of a coalitional game v is defined by restricting its domain to subcoalitions of a given coalition *U*.) Shapley and Shubik (1969a) showed that *this necessary condition is also sufficient*. Balancedness itself is not sufficient, since there exist games with non-empty cores having subgames with empty cores (e.g., $|I| = 4$, $v(S) := 0, 0, 1, 1, 2$ when $|S| = 0, 1, 2, 3, 4$, respectively).

For the NTU case, characterizations of market games have been obtained by Billera and Bixby (1973), Mas-Colell (1975), and others.

Though the subject of this section is finite markets, it is nevertheless worthwhile to relate the results to non-atomic games (where the players constitute a non-atomic continuum, an "ocean"). The total balancedness condition then takes on a particularly simple form. Suppose, for simplicity, that v is a function of finitely many measures, i.e., $v(S) = f(\mu(S))$, where $\mu = (\mu_1, \ldots, \mu_n)$, and the μ_j are non-atomic measures. Then v is a market game iff f is concave and 1-homogeneous ($f(\theta x) = \theta f(x)$) when $\theta \geqslant 0$). This is equivalent to saying that v is superadditive (at all coalitions), and f is 1-homogeneous (Aumann and Shapley, 1974).

Perhaps the most remarkable expression of the connection between superadditivity and the core has been obtained by Wooders (1983). Consider coalitional games with a fixed finite number k of 'types' of players, the coalitional form being given by $v(\mathrm{S}) = f(\mu(\mathrm{S}))$, where $\mu(\mathrm{S})$ is the profile of type sizes in S, i.e. it is a vector whose i'th coordinate represents the number of type i players in S. (To specify the game, $\mu(\mathrm{I})$ must also be specified.) Assume that f is superadditive, i.e. $f(x + y) \geqslant f(x) + f(y)$ for all x and y with non-negative integer coordinates; this assures the superadditivity of v. Moreover, assume that f obeys a 'Lipschitz' condition, namely that $|f(x) - f(y)|/\|x - y\|$ is uniformly bounded for all $x \neq y$, where $\|x\| := \max_j |x_j|$. Then for each $\varepsilon > 0$, when the number of players is sufficiently large, the ε-core is non-empty. (The ε-core is defined as the set of all outcomes x such that $x(\mathrm{S}) \geqslant v(\mathrm{S}) - \varepsilon|\mathrm{S}|$ for all S.) Roughly, the result says that the core is "almost" non-empty for sufficiently large games that are superadditive and obey the Lipschitz condition. Intuitively, the superadditivity together with the Lipschitz condition yield "approximate" 1-homogeneity, and in the presence of 1-homogeneity, superadditivity is equivalent to concavity. Thus f is approximately a 1-homogeneous concave function, so that we are back in a situation similar to that treated in the previous paragraph. What makes this result so remarkable is that other than the Lipschitz condition, the only substantive assumption is superadditivity.

Wooders (1983) also obtained a similar theorem for NTU; Wooders and Zame (1984) obtained a formulation that does away with the finite type assumption.

1970–1986

We do not yet have sufficient distance to see the developments of this period in proper perspective. Political and political-economic models were studied in depth. Non-cooperative game theory was applied to a large variety of particular economic models, and this led to the study of important variants on the refinements of the equilibrium concept. Great strides forward were made in almost all the areas that had been initiated in previous decades, such as repeated games (both of complete and of incomplete information), stochastic games, value, core, nucleolus, bargaining theory, games with many players, and so on (many of these developments have been mentioned above). Game Theory was applied to biology, computer science, moral philosophy, cost allocation. New light was shed on old concepts such as randomized strategies.

Sociologically, the discipline proliferated greatly. Some 16 or 17 people participated in the first international workshop on game theory held in Jerusalem in 1965; the fourth one, held in Cornell in 1978, attracted close to 100, and the discipline is now too large to make such workshops useful. An international workshop in the relatively restricted area of repeated games, held in Jerusalem in 1985, attracted over fifty participants. The *International Journal of Game Theory* was founded in 1972; *Mathematics of Operations Research*, founded in 1975, was organized into three major "areas," one of them Game Theory. Economic theory journals, such as the *Journal of Mathematical Economics*, the *Journal of Economic Theory*, *Econometrica*, and others devoted increasing proportions of their space to game theory. Important centres of research, in addition to the existing ones, sprang up in France, Holland, Japan, England, and India, and at many universities in the United States.

Gradually, game theory also became less personal, less the exclusive concern of a small "in" group whose members all know each other. For years, it had been a tradition in game theory to publish only a fraction of what one had found, and then only after great delays, and not always what is most important. Many results were passed on by word of mouth, or remained hidden in ill-circulated research memoranda. The "Folk Theorem" to which we alluded above (*1950–1960*, iii) is an example. This tradition had both beneficial and deleterious effects. On the one hand, people did not rush into print with trivia, and the slow cooking of results improved their flavour. As a result, phenomena were sometimes rediscovered several times, which is perhaps not entirely bad, since you understand something best when you discover it yourself. On the other hand, it was difficult for outsiders to break in; non-publication caused less interest to be generated than would otherwise have been, and significantly impeded progress.

Be that as it may, those day are over. There are now hundreds of practitioners, they do not all know each other, and sometimes have never even heard of one another. It is no longer possible to communicate in the old way, and as a result, people are publishing more quickly. As in other disciplines, it is becoming difficult to keep abreast of the important developments. Game theory has matured.

i. *Applications to biology*. A development of outstanding importance, whose implications are not yet fully appreciated, is the application of game thory to evolutionary *biology*. The high priest of this subject is John

Maynard Smith (1982), a biologist whose concept of *evolutionarily stable strategy*, a variant of strategic equilibrium, caught the imagination both of biologists and of game theorists. On the game theoretic side, the theme was taken up by Reinhard Selten (1980, 1983) and his school; a conference on "Evolutionary theory in biology and economics," organized by Selten in Bielefeld in 1985, was enormously successful in bringing field biologists together with theorists of games to discuss these issues. A typical paper was tit for tat in the great tit (Regelmann and Curio, 1986); using actual field observations, complete with photographs, it describes how the celebrated "tit for tat" strategy in the repeated prisoners' dilemma (Axelrod, 1984) accurately describes the behaviour of males and females of a rather common species of bird called the great tit, when protecting their young from predators.

It turns out that ordinary, utility maximizing rationality is much more easily observed in animals and even plants than it is in human beings. There are even situations where rats do significantly better than human beings. Consider, for example, the famous probability matching experiment, where the subject must predict the values of a sequence of i.i.d. random variables taking the values L and R with probabilities 3/4 and 1/4 respectively; each correct prediction is rewarded. It is of course optimal always to predict L; but human subjects tend to match the probabilities, i.e. to predict L about 3/4 of the time. On the other hand, while rats are not perfect (i.e. do not predict L *all* the time), they do predict L significantly more often than human beings.

Several explanations have been suggested. One is that in human experimentation, the subjects try subconsciously to "guess right," i.e., to guess what the experimenter "wants" them to do, rather than maximizing utility. Another is simply that the rats are more highly motivated. They are brought down to 80 per cent of their normal body weight, are literally starving; it is much more important for them to behave optimally than it is for human subjects.

Returning to theory, though the notion of strategic equilibrium seems on the face of it simple and natural enough, a careful examination of the definition leads to some doubts and questions as to why and under what conditions the players in a game might be expected to play a strategic equilibrium. Evolutionary theory suggests a simple rationale for strategic equilibrium, in which there is no conscious or overt decision making at all. For definiteness, we confine attention to two-person games, though the same ideas apply to the general case. We think of each of the two players as a whole species rather than an individual; reproduction is assumed asexual. The set of pure strategies of each player is interpreted as the locus of some gene (examples of a locus are eye colour, degree of aggressiveness, etc.); individual pure strategies are interpreted as alleles (blue or green or brown eyes, aggressive or timid behaviour, etc.). A given individual of each species possesses just one allele at the given locus; he interacts with precisely one individual in the other species, who also has just one allele at the locus of interest. The result of the interaction is a definite increment or decrement in the fitness of each of the two individuals, i.e. the number (or expected number) of his offspring; thus the payoff in the game is denominated in terms of fitness.

In these terms, a mixed strategy is a distribution of alleles throughout the population of the species (e.g., 40% aggressive, 60% timid). If each individual of each species is just as likely to meet any one individual of the other species as any other one, then the probability distribution of alleles that each individual faces is precisely given by the original mixed strategy. It then follows that a given pair of mixed strategies is a strategic equilibrium if and only if it represents a population equilibrium, i.e. a pair of distributions of characteristics (alleles) that does not tend to change.

Unfortunately, sexual reproduction screws up this story, and indeed the entire Maynard Smith approach has been criticized for this reason. But to be useful, the story does not have to be taken entirely literally. For example, it applies to evolution that is cultural rather than biological. In this approach, a "game" is interpreted as a *kind* of confrontational situation (like shopping for a car) rather than a specific instance of such a situation; a "player" is a role ("buyer" or "salesman"), not an individual human being; a pure strategy is a possible kind of behaviour in this role ("hard sell" or "soft sell"). Up to now this is indeed not very different from traditional game theoretic usage. What is different in the evolutionary interpretation is that pure or mixed strategic equilibria do not represent conscious rational choices of the players, but rather a population equilibrium which evolves as the result of how successful certain behaviour is in certain roles.

ii. *Randomization as ignorance*. In the traditional view of strategy randomization, the players use a randomizing device, such as a coin flip, to decide on their actions. This view has always had difficulties. Practically speaking, the idea that serious people would base important decisions on the flip of a coin is difficult to swallow. Conceptually, too, there are problems. The reason a player must randomize in equilibrium is only to keep others from deviating. For himself, randomizing is unnecessary; he will do as well by choosing any pure strategy that appears with positive probability in his equilibrium mixed strategy.

Of course, there is no problem if we adopt the evolutionary model described above in (i); mixed strategies appear as population distributions, and there is no explicit randomization at all. But what is one to make of randomization within the more usual paradigm of conscious, rational choice?

According to Savage (1954), randomness is not physical, but represents the ignorance of the decision maker. You associate a probability with every event about which you are ignorant, whether this event is a coin flip or a strategic choice by another player. The important thing in strategy randomization is that the *other* players be ignorant of what you are doing, and that they ascribe the appropriate probabilities to each of your pure strategies. It is not necessary for you actually to flip a coin.

The first to break away from the idea of explicit randomization was J. Harsanyi (1973). He showed that if the payoffs to each player *i* in a game are subjected to small independent random perturbations, known to *i* but not to the other players, then the resulting game of incomplete information has *pure* strategy equilibria that correspond to the mixed strategy equilibria of the original game. In plain words, nobody really randomizes. The appearance of randomization is due to the payoffs not being exactly known to all; each player, who does know his own payoff exactly, has a unique optimal action against his estimate of what the others will do.

This reasoning may be taken one step further. Even without perturbed payoffs, the players simply do not know which strategies will be chosen by the other players. At an equilibrium of "matching pennies," each player knows very well what he himself will do, but ascribes 1/2–1/2 probabilities to the other's actions; he also knows that the other ascribes those probabilities to his own actions, though it is admittedly not quite obvious that this is necessarily the case. In the case of a general n-person game, the situation is essentially similar; the mixed strategies of *i* can always be understood as describing the uncertainty of players other than *i* about what *i* will do (Aumann, 1987).

iii. *Refinements of strategic equilibrium*. In analysing specific economic models using the strategic equilibrium—an activity carried forward with great vigour since about 1975—it was found that Nash's definition does not provide adequately for rational choices given one's information at each stage of an extensive game. Very roughly, the reason is that Nash's definition ignores contingencies "off the equilibrium path." To remedy this, various "refinements" of strategic equilibrium have been defined, starting with Selten's (1975) "trembling hand" equilibrium. Please refer to our discussion of Zermelo's theorem (*1930–1950*, vi).

The interesting aspect of these refinements is that they use *irrationality* to arrive at a strong form of rationality. In one way or another, all of them work by assuming that irrationality cannot be ruled out, that the players ascribe irrationality to each other with a small probability. True rationality requires "noise"; it cannot grow in sterile ground, it cannot feed on itself only.

iv. *Bounded rationality*. For a long time it has been felt that both game and economic theory assume too much rationality. For example, the hundred-times repeated prisoner's dilemma has some $2^{2^{100}}$ pure strategies; all the books in the world are not large enough to write this number even once in decimal notation. There is no practical way in which all these strategies can be considered truly available to the players. On the face of it, this would seem to render statements about the equilibrium points of such games (*1950–1960*, iv) less compelling, since it is quite possible that if the sets of strategies were suitably restricted, the equilibria would change drastically.

For many years, little on the formal level was done about these problems. Recently the theory of automata has been used for formulations of bounded rationality in repeated games. Neyman (1985a) assumes that only strategies that are programmable on an automaton of exogenously fixed size can be considered "available" to the players. He then shows that even when the size is very large, one obtains results that are qualitatively different from those when all strategies are permitted. Thus in the n-times repeated prisoner's dilemma, only the greedy–greedy outcome can occur in equilibrium; but if one restricts the players to using automata with as many as $e^{o(n)}$ states, then for sufficiently large n, one can approximate in equilibrium any feasible individually rational outcome, and in particular the friendly–friendly outcome. For example, this is the case if the number of states is bounded by any fixed polynomial in n. Recently, Neyman (*MOR* 1998) generalized this result from the prisoner's dilemma to arbitrary games; specifically, he shows that a result similar to the Folk Theorem holds in any long finitely repeated game, when the automaton size is limited as above to subexponential.

Another approach has been used by Rubinstein (1986), with dramatically different results. In this work, the automaton itself is endogenous; all states of the automaton must actually be used on the equilibrium path. Applied to the prisoner's dilemma, this assumption leads to the conclusion that in equilibrium, one cannot get anywhere near the friendly–friendly outcome. Intuitively, the requirement that all states be used in equilibrium rules out strategies that punish deviations from equilibrium, and these are essential to the implicit enforcement mechanism that underlies the folk theorem. See the discussion at (*1950–1960*, iii) above.

v. *Distributed computing*. In the previous subsection (iv) we discussed applications of computer science to game theory. There are also applications in the opposite direction; with the advent of distributed computing, game theory has become of interest in computer science. Different units of a distributed computing system are viewed as different players, who must communicate and coordinate. Breakdowns and failures of one unit are often modelled as malevolent, so as to get an idea as to how bad the worst case can be. From the point of view of computer tampering and crime, the model of the malevolent player is not merely a fiction; similar remarks hold for cryptography, where the system must be made proof against purposeful attempts to "break in." Finally, multi-user systems come close to being games in the ordinary sense of the word.

vi. *Consistency* is a remarkable property which, in one form or another, is common to just about all game-theoretic solution concepts. Let us be given a game, which for definiteness we denote v, though it may be NTU or even non-cooperative. Let x be an outcome that "solves" the game in some sense, like the value or nucleolus or a point in the core. Suppose now that some coalition S wishes to view the situation as if the players outside S get their components of x so to speak exogenously, without participating in the play. That means that the players in S are playing the "reduced game" v_x^S, whose all-player set is S. It is not always easy to say just how v_x^S should be defined, but let's leave that aside for the moment. Suppose we apply to v_x^S the same solution concept that when applied to v yields x. Then the consistency property is that $x|S$ (x restricted to S) is the resulting solution. For example, if x is the nucleolus of v, then for each v, the restriction $x|S$ is the nucleolus of v_x^S.

Consistency implies that it is not too important how the player set is chosen. One can confine attention to a 'small world,' and the outcome for the denizens of this world will be the same as if we had looked at them in a "big world."

In a game theoretic context, consistency was first noticed by J. Harsanyi (1959) for the Nash solution to the n-person bargaining game. This is simply an NTU game V in which the only significant coalitions are the single players and the all-player coalition, and the single players are normalized to get 0. The Nash solution, axiomatized by Harsanyi (1959), is the outcome x that maximizes the product $x^1x^2\ldots x^n$. To explain the consistency condition, let us look at the case $n = 3$, in which case V({1, 2, 3}) is a subset of 3-space. If we let $S = \{1, 2\}$, and if x_0 is the Nash solution, then 3 should get x_0^3. That means that 1 and 2 are confined to bargaining within that slice of V({1, 2, 3}) that is determined by the plane $x^3 = x_0^3$. According to the Nash solution for the two-person case, they should maximize x^1x^2 over this slice; it is not difficult to see that this maximum is attained at (x_0^1, x_0^2), which is exactly what consistency requires.

Davis and Maschler (1965) proved that the kernel satisfies a consistency condition; so do the bargaining set, core, stable set, and nucleolus, using the same definition of the reduced game v_x^S as for the kernel (Aumann and Drèze, 1974). Using a somewhat different definition of v_x^S, consistency can be established for the value (Hart and Mas-Colell, 1989). Note that strategic equilibria, too, are consistent; if the players outside S play their equilibrium strategies, an equilibrium of the resulting game on S is given by having the players in S play the same strategies that they were playing in the equilibrium of the large game.

Consistency often plays a key role in axiomatizations. Strategic equilibrium is axiomatized by consistency, together with the requirement that in one-person maximization problems, the maximum be chosen. A remarkable axiomatization of the Nash solution to the bargaining problem (including the 2-person case discussed at *1950–1960*, v), in which the key role is played by consistency, has been provided by T. Lensberg (1988). Axiomatizations in which consistency plays the key role have been provided for the nucleolus (Sobolev, 1975), core (Peleg, 1985, 1986), kernel (Peleg, 1986), and value (Hart and Mas-Colell, 1989). Consistency-like conditions have also been used in contexts that are not strictly game-theoretic, e.g. by Balinski and Young (1982), W. Thomson, J. Roemer, H. Moulin, H. P. Young and others.

In law, the consistency criterion goes back at least to the 2000-year old Babylonian Talmud (Aumann and Maschler, 1985). Though it is indeed a very natural condition, its huge scope is still somewhat startling.

vii. The fascination of *cost allocation* is that it retains the formal structure of cooperative game theory in a totally different interpretation. The question is how to allocate joint costs among users. For example, the cost of a water supply or sewage disposal system serving several municipalities (e.g. Bogardi and Szidarovsky, 1976); or the cost of telephone calls in an organization such as a university or corporation (Littlechild and Owen, 1973, 1976). In the airport case, for example, each "player" is one landing of one airplane, and $v(S)$ is the cost of building and running an airport large enough to accommodate the set S of landings. Note that $v(S)$ depends not only on the number of landings in S but also on its composition; one would not charge the same for a landing of a 747 as for a Piper, for example because the 747 requires a longer runway. The allocation of cost would depend on the solution concept; for example, if we are using the Shapley value ϕ, then the fee for each landing i would be $\phi^i v$.

The axiomatic method is particularly attractive here, since in this application the axioms often have rather transparent meaning. Most frequently used has been the Shapley value, whose axiomatic characterization is particularly transparent (Billera and Heath, 1982).

The literature on the game theoretic approach to cost allocation is quite large, probably several hundred items, many of them in the accounting literature (e.g., Roth and Verrecchia, 1979).

Concluding Remarks

i. *Ethics*. While game theory does have intellectual ties to ethics, it is important to realize that in itself, it has no moral content, makes no moral recommendations, is ethically neutral. Strategic equilibrium does not tell us to maximize utility, it explores what happens when we do. The Shapley value does not recommend dividing payoff according to power, it simply measures the power. Game Theory is a tool for telling us where incentives will lead. History and experience teach us that if we want to achieve certain goals, including moral and ethical ones, we had better see to the incentive effects of what we are doing; and if we do not want people to usurp power for themselves, we had better build institutions that spread power as thinly and evenly as possible. Blaming game theory—or, for that matter, economic theory—for selfishness is like blaming bacteriology for disease. Game theory studies selfishness, it does not recommend it.

ii. *Mathematical methods*. We have had very little to say about mathematical methods in the foregoing, because we wished to stress the conceptual side. Worth noting, though, is that mathematically, game theoretic results developed in one context often have important implications in completely different contexts. We have already mentioned the implications of two-person zero-sum theory for the theory of the core and for correlated equilibria (*1910–1930*, vii). The first proofs of the existence of competitive equilibrium (Arrow and Debreu, 1954) used the existence of strategic equilibrium in a generalized game (Debreu, 1952). Blackwell's 1956 theory of two-person zero-sum games with vector payoffs is of fundamental importance for n-person repeated games of complete information (Aumann, 1961) and for repeated games of incomplete information (e.g. Mertens, 1982; Hart, 1985b). The Lemke–Howson algorithm (1962) for finding equilibria of 2-person non-zero sum non-cooperative games was seminal in the development of the algorithms of Scarf (1967, 1973) for finding points in the core and finding economic equilibria.

iii. *Terminology*. Game Theory has sometimes been plagued by haphazard, inappropriate terminology. Some workers, notably L. S. Shapley (1973b), have tried to introduce more appropriate terminology, and we have here followed their lead. What follows is a brief glossary to aid the reader in making the proper associations.

Used here	*Older term*
Strategic form	Normal form
Strategic equilibrium	Nash equilibrium
Coalitional form	Characteristic function
Transferable utility	Side payment
Decisive voting game	Strong voting game
Improve upon	Block
Worth	Characteristic function value
Profile	*n*-tuple
1-homogeneous	Homogeneous of degree 1

Bibliography

Anderson, R. M. 1986. Notions of core convergence. In Hildenbrand and Mas-Colell (1986), 25–46.

Arrow, K. J. 1951. *Social Choice and Individual Values*. New York: John Wiley.

Arrow, K. J. and Debreu, G. 1954. Existence of an equilibrium for a competitive economy. *Econometrica* 22, 265–90.

Artstein, Z. 1972. Values of games with denumerably many players. *International Journal of Game Theory* 3, 129–40.

Asscher, N. 1976. An ordinal bargaining set for games without side payments. *Mathematics of Operations Research* 1, 381–9.

Aumann, R. J. 1960. Linearity of unrestrictedly transferable utilities. *Naval Research Logistics Quarterly* 7, 281–4 [Chapter 16].

Aumann, R. J. 1961. The core of a cooperative game without side payments. *Transactions of the American Mathematical Society* 98, 539–52 [Chapter 39].

Aumann, R. J. 1964. Markets with a continuum of traders. *Econometrica* 32, 39–50 [Chapter 46].

Aumann, R. J. 1975. Values of markets with a continuum of traders. *Econometrica* 43, 611–46 [Chapter 51].

Aumann, R. J. 1985. On the non-transferable utility value: a comment on the Roth–Shafer examples. *Econometrica* 53, 667–77 [Chapter 61c].

Aumann, R. J. 1987. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica* 55, 1–18 [Chapter 33].

Aumann, R. J. and Drèze, J. H. 1974. Cooperative games with coalition structures. *International Journal of Game Theory* 3, 217–38 [Chapter 44].

Aumann, R. J. and Kurz, M. 1977. Power and taxes. *Econometrica* 45, 1137–61 [Chapter 52].

Aumann, R. J. and Maschler, M. 1985. Game theoretic analysis of a bankruptcy problem from the Talmud. *Journal of Economic Theory* 36, 195–213 [Chapter 45].

Aumann, R. J. and Peleg, B. 1960. Von Neumann–Morgenstern solutions to cooperative games without side payments. *Bulletin of the American Mathematical Society* 66, 173–9 [Chapter 38].

Aumann, R. J. and Shapley, L. S. 1974. *Values of Non-Atomic Games.* Princeton: Princeton University Press.

Axelrod, R. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Balinski, M. L. and Young, H. P. 1982. *Fair Representation*. New Haven: Yale University Press.

Berbee, H. 1981. On covering single points by randomly ordered intervals. *Annals of Probability* 9, 520–28.

Bewley, T. and Kohlberg, E. 1976. The asymptotic theory of stochastic games. *Mathematics of Operations Research* 1, 197–208.

Billera, L. J. 1970a. Existence of general bargaining sets for cooperative games without side payments. *Bulletin of the American Mathematical Society* 76, 375–9.

Billera, L. J. 1970b. Some theorems on the core of an n-person game without side payments. *SIAM Journal of Applied Mathematics* 18, 567–79.

Billera, L. J. and Bixby, R. 1973. A characterization of polyhedral market games. *International Journal of Game Theory* 2, 253–61.

Billera, L. J. and Heath, D. C. 1982. Allocation of shared costs: a set of axioms yielding a unique procedure. *Mathematics of Operations Research* 7, 32–9.

Billera, L. J., Heath, D. C. and Raanan, J. 1978. Internal telephone billing rates—a novel application of non-atomic game theory. *Operations Research* 26, 956–65.

Binmore, K. 1987. Perfect equilibria in bargaining models. In Binmore, K. and Dasgupta, P. (eds), *The Economics of Bargaining*, Blackwell, Oxford, 77–105.

Blackwell, D. 1956. An analogue of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics* 6, 1–8.

Blackwell, D. and Ferguson, T. S. 1968. The big match. *Annals of Mathematical Statistics* 39, 159–63.

Bogardi, I. and Szidarovsky, F. 1976. Application of game theory in water management. *Applied Mathematical Modelling* 1, 11–20.

Bondareva, O. N. 1963. Some applications of linear programming methods to the theory of cooperative games (in Russian). *Problemy kibernetiki* 10, 119–39.

Borel, E. 1924. Sur les jeux où interviennent l'hasard et l'habilité des joueurs. In *Eléments de la theorie des probabilités*, ed. J. Hermann, Paris; Librairie Scientifique, 204–24.

Braithwaite, R. B. (ed.) 1950. F. P. Ramsey, *The Foundations of Mathematics and Other Logical Essays*. New York: Humanities Press.

Brams, S. J., Lucas, W. F. and Straffin, P. D., Jr. (eds) 1983. *Political and Related Models*. New York: Springer.

Brown, D. J. and Loeb, P. 1976. The values of non-standard exchange economies. *Israel Journal of Mathematics* 25, 71–86.

Brown, D. J. and Robinson, A. 1975. Non standard exchange economies. *Econometrica* 43, 41–55.

Case, J. H. 1979. *Economics and the Competitive Process*. New York: New York University Press.

Champsaur, P. 1975. Cooperation vs. competition. *Journal of Economic Theory* 11, 394–417.

Dantzig, G. B. 1951a. A proof of the equivalence of the programming problem and the game problem. In Koopmans (1951), 330–38.

Dantzig, G. B. 1951b. Maximization of a linear function of variables subject to linear inequalities. In Koopmans (1951), 339–47.

Davis, M. 1967. Existence of stable payoff configurations for cooperative games. In Shubik (1967), 39–62.

Davis, M. and Maschler, M. 1964. Infinite games with perfect information. In Dresher, Shapley and Tucker (1964), 85–101.

Davis, M. and Maschler, M. 1965. The kernel of a cooperative game. *Naval Research Logistics Quarterly* 12, 223–59.

Debreu, G. 1952. A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences of the United States* 38, 886–93.

Debreu, G. and Scarf, H. 1963. A limit theorem on the core of an economy. *International Economic Review* 4, 236–46.

Dresher, M. A. and Shapley, L. S. and Tucker, A. W. (eds) 1964. *Advances in Game Theory*. Annals of Mathematics Studies 52, Princeton: Princeton University Press.

Dresher, M. A., Tucker, A. W. and Wolfe, P. (eds) 1957. *Contributions to the Theory of Games III*. Annals of Mathematics Studies 39, Princeton: Princeton University Press.

Drèze, J. H., Gabszewicz, J. and Gepts, S. 1969. On cores and competitive equilibria. In Guilbaud (1969), 91–114.

Dubey, P. 1980. Asymptotic semivalues and a short proof of Kannai's theorem. *Mathematics of Operations Research* 5, 267–70.

Dubey, P. and Shapley, L. S. 1994. Non-cooperative exchange with a continuum of traders: two models, *Journal of Mathematical Economics* 23, 253–293.

Edgeworth, F. Y. 1881. *Mathematical Psychics*. London: Kegan Paul.

Everett, H. 1957. Recursive games. In Dresher, Tucker and Wolfe (1957), 47–78.

Fogelman, F. and Quinzii, M. 1980. Asymptotic values of mixed games. *Mathematics of Operations Research* 5, 86–93.

Gabszewicz, J. J. and Mertens, J. F. 1971. An equivalence theorem for the core of an economy whose atoms are not "too" big. *Econometrica* 39, 713–21.

Gale, D. 1974. A curious nim-type game. *American Mathematical Monthly* 81, 876–79.

Gale, D. 1979. The game of hex and the Brouwer fixed-point theorem. *American Mathematical Monthly* 86, 818–27.

Gale, D. 1986. Bargaining and competition, Part I: Characterization; Part II: Existence. *Econometrica* 54, 785–806; 807–18.

Gale, D. and Shapley, L. S. 1962. College admissions and the stability of marriage. *American Mathematical Monthly* 69, 9–15.

Gale, D. and Stewart, F. M. 1953. Infinite games with perfect information. In Kuhn and Tucker (1953), 245–66.

Galil, Z. 1974. The nucleolus in games with major and minor players. *International Journal of Game Theory* 3, 129–40.

Gillies, D. B. 1959. Solutions to general non-zero-sum games. In Luce and Tucker (1959), 47–85.

Guilbaud, G. T. (ed.) 1969. *La décision: aggrégation et dynamique des ordres de préférence*. Paris: Editions du CNRS.

Harsanyi, J. C. 1956. Approaches to the bargaining problem before and after the theory of games: a critical discussion of Zeuthen's, Hicks' and Nash's theories. *Econometrica* 24, 144–57.

Harsanyi, J. C. 1959. A bargaining model for the cooperative n-person game. In Tucker and Luce (1959), 325–56.

Harsanyi, J. C. 1963. A simplified bargaining model for the n-person cooperative game. *International Economic Review* 4, 194–220.

Harsanyi, J. C. 1966. A general theory of rational behavior in game situations. *Econometrica* 34, 613–34.

Harsanyi, J. C. 1967–8. Games with incomplete information played by "Bayesian" players, parts I, II and III. *Management Science* 14, 159–82, 320–34, 486–502.

Harsanyi, J. C. 1973. Games with randomly disturbed payoffs: a new rationale for mixed strategy equilibrium points. *International Journal of Game Theory* 2, 1–23.

Harsanyi, J. C. 1982. Solutions for some bargaining games under the Harsanyi–Selten solution theory I: Theoretical preliminaries; II: Analysis of specific games. *Mathematical Social Sciences* 3, 179–91; 259–79.

Harsanyi, J. C. and Selten, R. 1987. *A General Theory of Equilibrium Selection in Games*. Cambridge, Mass.: MIT Press.

Harsanyi, J. C. and Selten, R. 1972. A generalized Nash solution for two-person bargaining games with incomplete information. *Management Science* 18, 80–106.

Hart, S. 1973. Values of mixed games. *International Journal of Game Theory* 2, 69–86.

Hart, S. 1974. Formation of cartels in large markets. *Journal of Economic Theory* 7, 453–66.

Hart, S. 1977a. Asymptotic values of games with a continuum of players. *Journal of Mathematical Economics* 4, 57–80.

Hart, S. 1977b. Values of non-differentiable markets with a continuum of traders. *Journal of Mathematical Economics* 4, 103–16.

Hart, S. 1980. Measure-based values of market games. *Mathematics of Operations Research* 5, 197–228.

Hart, S. 1985a. An axiomatization of Harsanyi's nontransferable utility solution. *Econometrica* 53, 1295–314.

Hart, S. 1985b. Non zero-sum two-person repeated games with incomplete information. *Mathematics of Operations Research* 10, 117–53.

Hart, S. and Mas-Colell, A. 1989. The potential: a new approach to the value in multiperson allocation problems. *Econometrica* 57, 589–614.

Hart, S. and Schmeidler, D. 1989. Correlated equilibria: an elementary existence proof. *Mathematics of Operations Research* 14, 18–25.

Hildenbrand, W. (ed.) 1982. *Advances in Economic Theory*. Cambridge: Cambridge University Press.

Hildenbrand, W. and Mas-Colell, A. (eds.) 1986. *Contributions to Mathematical Economics in Honor of G. Debreu.* Amsterdam: North-Holland.

Hu, T. C. and Robinson, S. M. (eds) 1973. *Mathematical Programming*. New York: Academic Press.

Isaacs, R. 1965. *Differential Games: A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization*. New York: John Wiley.

Kakutani, S. 1941. A generalization of Brouwer's fixed point theorem. *Duke Mathematical Journal* 8, 457–9.

Kalai, E. and Smorodinsky, M. 1975. Other solutions to Nash's bargaining problem. *Econometrica* 43, 513–18.

Kaneko, M. and Wooders, M. 1982. Cores of partitioning games. *Mathematical Social Sciences* 3, 313–27.

Kannai, Y. 1966. Values of games with a continuum of players. *Israel Journal of Mathematics* 4, 54–8.

Kohlberg, E. 1971. On the nucleolus of a characteristic function game. *SIAM Journal of Applied Mathematics* 20, 62–6.

Kohlberg, E. 1972. The nucleolus as a solution to a minimization problem. *SIAM Journal of Applied Mathematics* 23, 34–49.

Koopmans, T. C. (ed.) 1951. *Activity Analysis of Production and Allocation*. New York: Wiley.

Kuhn, H. W. 1952. *Lectures on the Theory of Games*. Issued as a report of the Logistics Research Project, Office of Naval Research, Princeton University.

Kuhn, H. W. 1953. Extensive games and the problem of information. In Kuhn and Tucker (1953), 193–216.

Kuhn, H. W. and Tucker, A. W. (eds) 1950. *Contributions to the Theory of Games I*. Annals of Mathematics Studies 24, Princeton: Princeton University Press.

Kuhn, H. W. and Tucker, A. W. (eds) 1953. *Contributions to the Theory of Games II*. Annals of Mathematics Studies 28, Princeton: Princeton University Press.

Lemke, L. E. and Howson, J. T. 1962. Equilibrium points of bimatrix games. *SIAM Journal of Applied Mathematics* 12, 413–23.

Lensberg, T. 1988. Stability and the Nash Solution. *Journal of Economic Theory* 45, 330–341.

Lewis, D. K. 1969. *Convention*. Cambridge, Mass.: Harvard University Press.

Littlechild, S. C. and Owen, G. 1973. A simple expression for the Shapley value in a special case. *Management Science* 20, 370–72.

Littlechild, S. C. and Owen, G. 1976. A further note on the nucleolus of the "airport game." *International Journal of Game Theory* 5, 91–5.

Lucas, W. F. 1969. The proof that a game may not have a solution. *Transactions of the American Mathematical Society* 137, 219–29.

Lucas, W. F. 1983. Measuring power in weighted voting systems. In Brams, Lucas and Straffin (1983), ch. 9.

Lucas, W. F. and Rabie, M. 1982. Games with no solutions and empty core. *Mathematics of Operations Research* 7, 491–500.

Luce, R. D. and Raiffa, H. 1957. *Games and Decisions, Introduction and Critical Survey*. New York: John Wiley.

Luce, R. D. and Tucker, A. W. (eds) 1959. *Contributions to the Theory of Games IV*. Annals of Mathematics Studies 40, Princeton: Princeton University Press.

Lyapounov, A. A. 1940. On completely additive vector-functions (in Russian, abstract in French). *Akademiia Nauk USSR Izvestiia Seriia Mathematicheskaia* 4, 465–78.

Martin, D. A. 1975. Borel determinacy. *Annals of Mathematics* 102, 363–71.

Maschler, M. (ed.) 1962. *Recent Advances in Game Theory*. Proceedings of a Conference, privately printed for members of the conference, Princeton: Princeton University Conferences.

Maschler, M., Peleg, B. and Shapley, L. S. 1979. Geometric properties of the kernel, nucleolus, and related solution concepts. *Mathematics of Operations Research* 4, 303–38.

Maschler, M. and Perles, M. 1981. The superadditive solution for the Nash bargaining game. *International Journal of Game Theory* 10, 163–93.

Mas-Colell, A. 1975. A further result on the representation of games by markets. *Journal of Economic Theory* 10, 117–22.

Mas-Colell, A. 1977. Competitive and value allocations of large exchange economies. *Journal of Economic Theory* 14, 419–38.

Mas-Colell, A. 1988. An equivalence theorem for a bargaining set. *Journal of Mathematical Economics*.

Maynard Smith, J. 1982. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Mertens, J. F. 1982. Repeated games: an overview of the zero-sum case. In Hildenbrand (1982), 175–82.

Mertens, J. F. 1987. The Shapley value in the non-differentiable case. *International Journal of Game Theory* 17, 1–65.

Mertens, J. F. and Neyman, A. 1981. Stochastic games. *International Journal of Game Theory* 10, 53–66.

Milnor, J. W. and Shapley, L. S. 1957. On games of survival. In Dresher, Tucker and Wolfe (1957), 15–45.

Milnor, J. W. and Shapley, L. S. 1978. Values of large games II: Oceanic games. *Mathematics of Operations Research* 3, 290–307.

Moschovakis, Y. N. 1980. *Descriptive Set Theory*. New York: North-Holland.

Moschovakis, Y. N. (ed.) 1983. *Cabal Seminar 79–81: Proceedings, Caltech-UCLA Logic Seminar 1979–81*. Lecture Notes in Mathematics 1019, New York: Springer-Verlag.

Mycielski, J. and Steinhaus, H. 1964. On the axiom of determinateness. *Fundamenta Mathematicae* 53, 205–24.

Myerson, R. B. 1977. Graphs and cooperation in games. *Mathematics of Operations Research* 2, 225–9.

Myerson, R. B. 1979. Incentive compatibility and the bargaining problem. *Econometrica* 47, 61–74.

Myerson, R. B. 1984. Cooperative games with incomplete information. *International Journal of Game Theory* 13, 69–96.

Nash, J. F. 1950. The bargaining problem. *Econometrica* 18, 155–62.

Nash, J. F. 1951. Non-cooperative games. *Annals of Mathematics* 54, 289–95.

Neyman, A. 1985a. Bounded complexity justifies cooperation in the finitely repeated prisoner's dilemma. *Economics Letters* 19, 227–30.

Neyman, A. 1985b. Semivalues of political economic games. *Mathematics of Operations Research* 10, 390–402.

Neyman, A. 1987. Weighted majority games have an asymptotic value. *Mathematics of Operations Research* 13, 556–580.

O'Neill, B. 1987. Non-metric test of the minimax theory of two-person zero-sum games. *Proceedings of the National Academy of Sciences of the United States* 84, 2106–9.

Owen, G. 1972. Multilinear extensions of games. *Management Science* 18, 64–79.

Peleg, B. 1963a. Solutions to cooperative games without side payments. *Transactions of the American Mathematical Society* 106, 280–92.

Peleg, B. 1963b. Bargaining sets of cooperative games without side payments. *Israel Journal of Mathematics* 1, 197–200.

Peleg, B. 1967. Existence theorem for the bargaining set $M_1^{(i)}$. In Shubik (1967), 53–6.

Peleg, B. 1968. On weights of constant-sum majority games. *SIAM Journal of Applied Mathematics* 16, 527–32.

Peleg, B. 1985. An axiomatization of the core of cooperative games without side payments. *Journal of Mathematical Economics* 14, 203–14.

Peleg, B. 1986. On the reduced games property and its converse. *International Journal of Game Theory* 15, 187–200.

Pennock, J. R. and Chapman, J. W. (eds) 1968. *Representation*. New York: Atherton.

Ramsey, F. P. 1931. Truth and probability. In Braithwaite (1950), 156–98.

Ransmeier, J. S. 1942. *The Tennessee Valley Authority: A Case Study in the Economics of Multiple Purpose Stream Planning*, Nashville: Vanderbilt University Press.

Regelmann, K. and Curio, E. 1986. How do great tit (Parus Major) pair mates cooperate in broad defence? *Behavior* 97, 10–36.

Riker, W. H., and Shapley, L. S. 1968. Weighted voting: a mathematical analysis for instrumental judgements. In Pennock and Chapman (1968), 199–216.

Robinson, A. 1970, 1974. *Non-Standard Analysis*. Amsterdam: North-Holland.

Rosenthal, R. W. 1979. Sequences of games with varying opponents. *Econometrica* 47, 1353–66.

Rosenthal, R. W. and Rubinstein, A. 1984. Repeated two player games with ruin. *International Journal of Game Theory* 13, 155–77.

Roth, A. E. 1977. The Shapley value as a von Neumann–Morgestern utility. *Econometrica* 45, 657–64.

Roth, A. E. 1984. The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of Political Economy* 92, 991–1016.

Roth, A. E. and Verrecchia, R. E. 1979. The Shapley value as applied to cost allocation: a reinterpretation. *Journal of Accounting Research* 17, 295–303.

Rubinstein, A. 1982. Perfect equilibrium in a bargaining model. *Econometrica* 50, 97–109.

Rubinstein, A. 1986. Finite automata play the repeated prisoner's dilemma. *Journal of Economic Theory* 39, 83–96.

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

Scarf, H. E. 1967. The core of an n-person game. *Econometrica* 35, 50–69.

Scarf, H. E. 1973. *The Computation of Economic Equilibria*. New Haven: Yale University Press.

Schelling, T. C. 1960. *The Strategy of Conflict*. Cambridge, Mass.: Harvard University Press.

Schmeidler, D. 1969. The nucleolus of a characteristic function game. *SIAM Journal of Applied Mathematics* 17, 1163–70.

Selten, R. C. 1965. Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragetragheit. *Zeitschrift für die gesamte Staatswissenschaft* 121, 301–24; 667–89.

Selten, R. C. 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* 4, 25–55.

Selten, R. C. 1980. A note on evolutionary stable strategies in asymmetric animal conflicts. *Journal of Theoretical Biology* 84, 101–101.

Selten, R. C. 1983. Evolutionary stability in extensive two-person games. *Mathematical Social Sciences* 5, 269–363.

Shapley, L. S. 1953a. Stochastic games. *Proceedings of the National Academy of Sciences of the United States* 39, 1095–100.

Shapley, L. S. 1953b. A value for n-person games. In Kuhn and Tucker (1953), 305–17.

Shapley, L. S. 1962. Values of games with infinitely many players. In Maschler (1962), 113–18.

Shapley, L. S. 1964. Values of large games, VII: a general exchange economy with money. *RAND Publication* RM-4248, Santa Monica, California.

Shapley, L. S. 1967. On balanced sets and cores. *Naval Research Logistics Quarterly* 14, 453–60.

Shapley, L. S. 1969. Utility comparison and the theory of games. In Guilbaud (1969), 251–63.

Shapley, L. S. 1973a. On balanced games without side payments. In Hu and Robinson (1973), 261–90.

Shapley, L. S. 1973b. Let's block 'block'. *Econometrica* 41, 1201–2.

Shapley, L. S. and Shubik, M. 1954. A method for evaluating the distribution of power in a committee system. *American Political Science Review* 48, 787–92.

Shapley, L. S. and Shubik, M. 1969a. On market games. *Journal of Economic Theory* 1, 9–25.

Shapley, L. S. and Shubik, M. 1969b. Pure competition, coalitional power and fair division. *International Economic Review* 10, 337–62.

Shapley, L. S. and Shubik, M. 1984. Convergence of the bargaining set for differentiable market games. Appendix B in Shubik (1984), 683–92.

Shitovitz, B. 1973. Oligopoly in markets with a continuum of traders. *Econometrica* 41, 467–501.

Shubik, M. 1959a. Edgeworth market games. In Luce and Tucker (1959), 267–78.

Shubik, M. 1959b. *Strategy and Market Structure*. New York: John Wiley.

Shubik, M. (ed.) 1967. *Essays in Mathematical Economics in Honor of Oskar Morgenstern*. Princeton: Princeton University Press.

Shubik, M. 1973. Commodity money, oligopoly, credit and bankruptcy in a general equilibrium model. *Western Economic Journal* 11, 24–36.

Shubik, M. 1982. *Game Theory in the Social Sciences: Concepts and Solutions*. Cambridge, Mass.: MIT Press.

Shubik, M. 1984. *A Game Theoretic Approach to Political Economy*. Cambridge, Mass.: MIT Press.

Sobolev, A.I. 1975. Characterization of the principle of optimality for cooperative games through functional equations (in Russian). In Vorobiev (1975), 94–151.

Sorin, S. 1986a. On repeated games of complete information. *Mathematics of Operations Research* 11, 147–60.

Sorin, S. 1986b. An asymptotic property of non-zero sum stochastic games. *International Journal of Game Theory* 15(2), 101–7.

Stearns, R. E. 1964. On the axioms for a cooperative game without side payments. *Proceedings of the American Mathematical Society* 15, 82–6.

Tauman, Y. 1981. Value on a class of non-differentiable market games. *International Journal of Game Theory* 10, 155–62.

Ville, J. A. 1938. Sur le théorie générale des jeux où intervient l'habilité des joueurs. In *Traité du calcul des probabilités et de ses applications*, Vol. 4, ed. E. Borel, Paris: Gauthier-Villars, 105–13.

Vinacke, W. E. and Arkoff, A. 1957. An experimental study of coalitions in the triad. *American Sociological Review* 22, 406–15.

Vind, K. 1965. A theorem on the core of an economy. *Review of Economic Studies* 32, 47–8.

von Neumann, J. 1928. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen* 100, 295–320.

von Neumann, J. 1949. On rings of operators. Reduction theory. *Annals of Mathematics* 50, 401–85.

von Neumann, J. and Morgenstern, O. 1944. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Vorobiev, N. N. (ed.) 1975. *Mathematical Methods in Social Science* (in Russian). Vipusk 6, Vilnius.

Wilson, R. 1978. Information, efficiency, and the core of an economy. *Econometrica* 46, 807–16.

Wolfe, P. 1955. The strict determinateness of certain infinite games. *Pacific Journal of Mathematics* 5, 841–7.

Wooders, M. H. 1983. The epsilon core of a large replica game. *Journal of Mathematical Economics* 11, 277–300.

Wooders, M. H. and Zame, W. R. 1984. Approximate cores of large games. *Econometrica* 52, 1327–50.

Young, H. P. 1985. Monotonic solutions of cooperative games. *International Journal of Game Theory* 14, 65–72.

Zermelo, E. 1913. Über eine Anwendung der Mengenlehre auf die theorie des Schachspiels. *Proceedings of the Fifth International Congress of Mathematicians* 2, 501–4.