3

Recent Advances in Game Theory and Political Science*

Frank C. Zagare
Boston University

I. INTRODUCTION

Game theory is formally a branch of mathematics developed to deal with conflict-of-interest situations in social science. Although the origins of the theory can be traced to early articles in the 1920s by mathematicians Emile Borel (1921, 1924) and John von Neumann (1928), the field was only definitely established when von Neumann and economist Oskar Morgenstern published *Theory of Games and Economic Behavior* in 1944. Since then, the literature of the field has expanded enormously, with theoretical research or applications of the theory blossoming in areas as diverse as operations research, mathematics, military science, biology, law, sports, biblical studies, and moral philosophy. The explosion of the paradigm has in fact been so large that recently Colman (1982, p. vii) was led to remark, somewhat wistfully, that "the time has (alas) long passed when a single person could reason-

*I would like to thank Steven J. Brams, Jacek Kugler and Nicholas R. Miller for their helpful comments and suggestions on an earlier version of this essay. Some parts of this chapter are drawn from Zagare (1984a).

ably hope to be an expert on all branches of game theory or on all of its applications."

The impact of the game-theoretic paradigm on the social sciences has been particularly striking. Long ago, sociologists, political scientists, and international relations specialists saw the relevance of game theory for studying the processes underlying coalition formation and behavior, important areas of concern in each of these disciplines. The enormous experimental literature on Prisoners' Dilemma and related games attests to the significant impact of game-theoretic models in social psychology. And a recent review by Schotter and Schwödiauer (1980) in the *Journal of Economic Literature* demonstrates how deeply the theory of games has penetrated economics since the publication of von Neumann and Morgenstern's monumental work. Indeed it is virtually impossible to be well-versed in any social science today without taking cognizance of the contributions of game theorists. It may even be true, as Howard (1971, p. 202) has persuasively argued, "that game theory is becoming a unifying force in the social sciences, encompassing economics, psychology, politics, and history within a single mathematical theory capable of being applied to the understanding of all interactions between conscious beings."

With appropriate modifications, the observations of both Colman and Howard also apply to the game theory and politics literature. The influence of game theory on the study of politics is both wide and diverse, with applications and extensions of the theory becoming increasingly prominent in each of the major subfields of the discipline. Rather than attempting to review this variegated literature in toto, this essay will highlight only those theoretical advances made since the mid-1970s that have immediate and general implications for the study of politics. This means that purely formal contributions, specific applications of the theory that are relevant to only certain subfields,[1] and the vast experimental gaming literature will not be reviewed.[2] Nevertheless, to place the subsequent discussion in context, a short summary of the assumptions, primitive concepts, and major divisions of the field will be provided in the next section.[3]

II. A BRIEF RÉSUME OF THE THEORY OF GAMES

As previously indicated, game theory is a theory of interdependent choice. Technically, the simplest type of game is a *one-person game*, sometimes called a *game against nature*, wherein a single player makes a decision in the face of an environment assumed to be either indifferent or neutral. One-

[1] Spatial models are discussed by Miller (this volume).

[2] For a recent overview of the experimental literature, see Colman (1982).

[3] For a more comprehensive summary or a literature review of some specialized topics, see either Riker and Ordeshook (1973) or Brams (1975).

person games, though, have received only limited attention from political scientists for the obvious reason that most intriguing political situations involve at least two players. *Two-person games*, therefore, are the most elementary interactive situations of general concern to students of politics. Games that involve more than two players are called *n-person games*.

In game theory, a *player* can be an individual, or a group of individuals, functioning as a decision-making unit. Individuals or groups become players when their decisions, coupled with the decision of at least one other actor, produce an *outcome*. Outcomes are, by necessity, content-dependent and therefore may range over the whole spectrum of possible societal states.

Players are assumed to be able to evaluate and compare the consequences associated with the set of possible outcomes, and to assign numbers, called *utilities*, to each outcome indicating a *preference* relationship among them. When these numbers are judged to reflect only a rank ordering of the outcomes, they are called *ordinal utilities;* when they indicate both order and intensity of preference, they are called *cardinal utilities.*[4]

The options available to players to bring about outcomes are called *strategies*. Strategies come in two types, *pure* and *mixed*. A pure strategy is a complete contingency plan that specifies a choice for a player in every situation that might arise in a game. A mixed strategy involves the use of a particular probability distribution to select one pure strategy from among a subset of a player's pure strategies.

Underlying the entire structure of game theory is the key assumption that players in a game are *rational* (or *utility maximizers*). As game theorists use this term, rationality simply means that a player in an interactive situation will act to bring about the most preferred of the possible outcomes, given the constraint that other players are also acting in the same way.

Games can be distinguished in a number of ways. In addition to the standard dichotomy between two-person and n-person games, games are sometimes categorized according to the extent to which the interests of the players diverge. Games in which conflict is total and the interests of the players diametrically opposed are known as *zero-sum* or *constant-sum games*. *Nonzero-sum games*, by contrast, are those games in which players have both competitive and complementary interests.

Games can also be divided according to the number of strategies available to each player. When each player has a finite number of strategies, the game is *finite*. If not, the game is *infinite*.

Finally, games can be classified according to the rules assumed to govern play. Games in which binding agreements are precluded are called *noncooperative games*. When binding agreements are possible, the game is termed *cooperative*.

---

[4] A nontechnical treatment of utility theory can be found in Davis (1983, chap. 4). A more formal discussion is given in Luce and Raiffa (1957, chap. 2).

From a theoretical perspective, the distinction between noncooperative and cooperative games is perhaps the most telling. For each of these two categories of games, a separate and not totally unified theory has evolved. For this reason, it seems appropriate to review the developments within each strand of theory individually. Recent advances in the noncooperative theory will be discussed first.

III. NONCOOPERATIVE GAME THEORY

A. Noncooperative Game Theory: A Theoretical Overview

Noncooperative game theory begins with the supposition that the players are unable to communicate or negotiate binding agreements with each other. Hence, this branch of game theory is particularly suited for analyzing situations where there are obstacles, such as antitrust laws, that block explicit player communication, or where institutional mechanisms for enforcing contracts do not exist, as might be the case in the international system.

In noncooperative game theory, the concept of an *equilibrium outcome* plays a central role. An outcome is defined to be an equilibrium outcome when no player has an incentive unilaterally to switch from his strategy associated with it.

Since equilibrium outcomes represent stable points in the set of possible societal states, they can be expected to be selected by rational agents on a regular basis. Consequently, the identification of these regularly occurring outcomes in both the model world of game theory and in the real world is a precondition to the discovery and specification of general laws of social behavior.

In classical noncooperative game theory, the standard notion of stability is due to Nash (1951). Briefly, Nash's equilibrium concept assumes that players consider only the immediate advantages and disadvantages of a unilateral strategy switch. If no player can benefit immediately by changing his strategy, the resulting outcome is a *Nash equilibrium.*

Nash's equilibrium concept possesses a number of attractive features that render it extremely appealing as the cornerstone of the theory of zero-sum games. First is *existence.* In 1928, von Neumann proved his famous *Minimax Theorem,* which established that all finite, two-person, zero-sum games have a (Nash) equilibrium in either pure or mixed strategies. Second is *equivalence.* Although several such equilibria may exist in a zero-sum game, each equilibrium outcome is equivalent or has the same value. This means that from the point of view of the players, there is no necessary tension among various equilibria when multiple equilibria exist in a zero-sum game. And finally, there is *interchangeability.* This means that an equilibrium outcome is found at the intersection of all equilibrium strategies. Therefore, regardless of the equilibrium strategy selected by a player in these games, the outcome will be a Nash equilibrium if the other player also selects some other equilibrium strategy.

Unfortunately, many of the attractive characteristics of Nash's equilibrium concept disappear when nonzero-sum games are examined. First, the significance of Nash's (1951) proof that all nonzero-sum games also have at least one pure or mixed strategy equilibrium outcome is mitigated by the dubious relevance of the concept of a mixed-strategy in nonzero-sum games (Shubik, 1982, pp. 250, 251). Moreover, when multiple Nash equilibria exist in a nonzero-sum game, they are not necessarily equivalent or interchangeable.

The limitations associated with Nash's equilibrium concept present important stumbling blocks for the development of a positive theory of politics within a noncooperative game-theoretic framework. More specifically, unless competing equilibria can be eliminated in situations wherein multiple equilibria are found, or unless specific equilibria can be discovered where none ostensibly exist, explanations and predictions derived from game-theoretic models will be weak and less than fully satisfying. It is partly for this reason that Ordeshook (1980, p. 450) has recently suggested that "the scientific task before us ... appears no different from the one von Neumann and Morgenstern confronted: generalizing and redefining the meaning of the word 'equilibrium.' "[5]

Related to, but distinct from, the difficulties implied by either nonexistent or nonequivalent and noninterchangeable Nash equilibria in nonzero-sum games is a puzzle that stems from the possibility that an equilibrium outcome, Nash or otherwise, may be less desireable than nonequilibria from the vantage point of all of the players in a game. This puzzle manifests itself most clearly in the game depicted in Figure 1, known as Prisoners' Dilemma, after a story,[6] attributed to A. W. Tucker, used to illustrate its structure.

In this nonzero-sum game, each of two players, A and B, is assumed to have two strategies, either to cooperate, C, with the other player, or to desist, D, from cooperation. These two strategies give rise to $2 \times 2 = 4$ possible outcomes. In Figure 1, these outcomes are represented by the ordered pair in each cell of the payoff matrix. For each player, the four outcomes are ranked from best to worst, with "4" assigned to each player's best outcome, "3" to each player's next-best outcome, and so on.[7]

[5] As will be seen, Ordeshook's observation applies equally well to cooperative game theory.

[6] For the original story, see Luce and Raiffa (1957), p. 95.

[7] Game theorists have developed a number of devices for abstracting the essential features of interactive situations. Games represented by a payoff matrix, as in Figure 1, are said to be represented in *normal form*. A normal form representation is to be contrasted with both the *extensive* (or *game tree*) *form* of representation in which a game tree is used to depict the sequence of moves available to each player, and the *characteristic function form* of representation which specifies a value, or the minimum payoff, that each player or coalition can guarantee itself in a game.

		Player B	
		C	D
Player A	C	(3,3)	(1,4)
	D	(4,1)	(2,2)

Figure 1 Prisoners' Dilemma.

By convention, the first number in each cell of the payoff matrix represents the row player's (i.e., A's) evaluation of the associated outcome while the second number represents the evaluation of the column player (i.e., B). For instance, if A chooses his strategy labeled "C," and B chooses his strategy labeled "D," the outcome (1,4) results. This represents A's worst outcome and B's best outcome.

Notice from Figure 1 that desisting, (D), *dominates* cooperating, (C), for both players, that is, each player does better in this game by selecting his D strategy, regardless of the strategy chosen by the other player. For instance, if B cooperates, A induces his best outcome by desisting and his next-best outcome by cooperating. And if B desists, A induces his next-worst outcome by also desisting and his worst outcome by cooperating. By symmetry, a similar logic faces B.

Dominant strategies are unconditionally best strategies. Therefore, D is each player's *optimal strategy.* The resulting outcome, (2,2), is the unique Nash equilibrium, and the next-worst for both players. Observe that if *both* players use their optimal strategy and desist, *both* are worse off than if *both* use their nonoptimal strategy and cooperate. Paradoxically, however, since each player has a dominant strategy, it remains true that each player is *individually* better off using it and desisting.

Technically, the unique equilibrium outcome of this game is said to be *non-Pareto-optimal* (or Pareto-deficient), that is, at least one player would do better and the other would do no worse by switching to another outcome. In this example, both players prefer (3,3) to (2,2). Conversely, the three nonequilibria are all *Pareto-optimal,* that is, each is preferred to any other outcome by *at least* one player. For instance, (3,3) is preferred to (1,4) by A, to (4,1) by B, and to (2,2) by both players. Among the 78 distinct 2 $\times$ 2 games identified by Rapoport and Guyer (1976), Prisoners' Dilemma alone is uniquely characterized by these two features, i.e., dominant strategies leading to a Pareto-deficient equilibrium outcome.

Prisoner's Dilemma is game theory's most famous game, and it has spawned an enormous amount of theoretical and experimental research. There are good reasons for this. First, the conflict between individual and collective interest, highlighted and neatly summarized in this game, lies at the heart of many important real-life situations with implications for political and other kinds of systems. For instance, a decision on whether or not to contribute toward the acquisition of a public good, to cheat on one's income tax, to join a labor union, or to conserve electricity in a power shortage are all essentially Prisoners' Dilemma-type choices. They pit an individual's narrow self-interest against the interests of a community.

Moreover, this perverse game demonstrates the fallaciousness of Adam Smith's argument concerning the ultimate effect of an individual pursuing "only his own gain." Players in a Prisoners' Dilemma game may, willy-nilly, find themselves caught in a "catch 22" situation in which they are both "done in" by their rational calculations. Even though they are both better off if they cooperate, the irrefutable logic of a dominant strategy dictates that each individual, in pursuing his own selfish ends, defects from cooperation. Put in a slightly different way, in many significant real-world situations, there may be "no invisible hand which brings the self-interest of one individual into harmony with the self-interest of another" (van den Doel, 1979, p. 55).

Finally, the paradoxical nature of the individually rational but collectively irrational solution to this game has important ramifications for nations in both their internal and external affairs. Internally, if in certain kinds of situations, individual members of a society are doomed to frustrate themselves and produce nonoptimal societal states, the case for increased government involvement in the private sector, and centralized political and economic control, would appear to be very strong. Indeed, the primary justification given by political philosophers like Hobbes, Rousseau, and Hume for the very existence of the state is the premise that individuals in Prisoners' Dilemma-type situations will not cooperate with one another.[8]

Externally, if nations are unable to devise mechanisms for cooperating with each other in areas of fundamental importance, we are condemned to live in a world in which conflict is the norm and in which international peace is but a respite for states preparing themselves for the next round of a prize fight without a final bell, except perhaps in the case of apocalypse. Thus, for a whole host of theoretical, philosophical and practical reasons, a specification of the conditions under which the dilemma of the prisoners can be overcome would be welcome.

B. Noncooperative Game Theory: Recent Advances

In the previous section, the most salient theoretical problems, and the most significant theoretical puzzle, of noncooperative game theory were identified. Since the mid-1970s, numerous efforts have been made to address one or both of these areas of concern. In this section, the most persuasive of these efforts are discussed.

[8] For a specific discussion of the connection between Prisoners' Dilemma and the work of Hobbes and Hume, see Taylor (1976).

1. The Fraser–Hipel Technique. Building upon earlier work by Howard (1971), the Fraser–Hipel (1979) technique extends Nash's notion of an equilibrium outcome and, in the process, provides a useful new methodology for analyzing complex conflicts. To illustrate both the richness and the empirical applicability of this new methodology, consider for now the Cuban missile crisis of 1962.

As is well-known, the missile crisis began, at least for the United States, on October 16, 1962, when President Kennedy was presented with the first hard evidence that the Soviet Union was installing medium- and intermediate-range ballistic missiles in Cuba. The most realistic options available to United States decision-makers at this time included:

1. Performing no aggressive action by either doing nothing or pursuing the matter through normal diplomatic channels.
2. Launching a "surgical" air strike to destroy the missile sites.
3. Blockading Cuba.

Once the missiles were discovered, Soviet decision-makers also had to make a choice from among three alternatives. They would either:

1. Withdraw the missiles.
2. Maintain the missiles.
3. Escalate the conflict.

Some of the options available to each superpower were mutually exclusive, while others could be selected concurrently. For instance, the Soviets could not simultaneously withdraw and maintain the missiles, although the United States could blockade and attack the missile sites at the same time.

After eliminating four mutually exclusive combinations of options for both players, twelve possible combinations, and hence, twelve different outcomes, remain. These are listed as columns in Figure 2. In the last row of this figure, the outcomes resulting from the various combinations of feasible op-

	Outcomes											
US												
Air Strike	0	1	0	1	0	1	0	1	0	1	0	1
Blockade	0	0	1	1	0	0	1	1	0	0	1	1
USSR												
Withdraw	0	0	0	0	1	1	1	1	0	0	0	0
Escalate	0	0	0	0	0	0	0	0	1	1	1	1
Decimalized	0	1	2	3	4	5	6	7	8	9	10	11

Figure 2 Players, options and outcomes for the Cuban missile crisis (from Fraser & Hipel, 1982).

tions are "decimalized" or converted into a binary number which allows convenient mathematical operations to be performed on the outcomes (for details, see Fraser & Hipel, 1979, 1982). For the purposes of this essay, however, these numbers will be used only to identify the various outcomes.

Each column in Figure 2 contains either a one (1) or a zero (0) next to each option. A "one" indicates that the associated option has been selected, while a "zero" indicates that the option has not been selected.

Notice that the "do nothing" option of the United States and the "maintain" option of the Soviet Union, although not listed, are implicit in Figure 2. The United States does nothing by not blockading or striking. The Soviet Union maintains by not withdrawing or escalating. Thus, in the first column, the United States does nothing (note the zeros next to its two options) and the Soviet Union maintains. In the second column (outcome 1), the United States strikes (indicated by a one next to this option) and the Soviets maintain. The options associated with the other ten outcomes are similarly interpreted.

In Figures 3 and 4, the various outcomes are arrayed in order of preference (as established by Fraser & Hipel, 1982) for the United States and the Soviet Union, respectively. For example, in this representation, the United States

	Outcomes											
US												
Air Strike	0	0	1	1	1	0	1	1	0	1	0	0
Blockade	0	1	0	1	1	0	1	0	1	0	1	0
USSR												
Withdraw	1	1	1	1	0	0	0	0	0	0	0	0
Escalate	0	0	0	0	0	0	0	0	1	1	1	1
Decimalized	4	6	5	7	2	1	3	0	11	9	10	8

Figure 3 Preference vector for the United States in the Cuban missile crisis (from Fraser & Hipel, 1982).

	Outcomes											
US												
Air Strike	0	0	0	0	1	1	1	1	1	1	0	0
Blockade	0	0	1	1	0	0	1	1	1	0	1	0
USSR												
Withdraw	0	1	1	0	1	0	1	0	0	0	0	0
Escalate	0	0	0	0	0	0	0	0	1	1	1	1
Decimalized	0	4	6	2	5	1	7	3	11	9	10	8

Figure 4 Preference vector for the USSR in the Cuban missile crisis (from Fraser & Hipel, 1982).

US												
E	E	X	X	X	X	X	X	X	X	X	X	overall stability
r	s	u	u	r	u	u	u	r	u	u	u	player stability
4	6	5	7	2	1	3	0	11	9	10	8	preference vector
	4	4	4		2	2	2		11	11	11	UIs
		6	6			1	1			9	9	
			5				3				10	
USSR												
r	s	r	u	r	u	r	u	u	u	u	u	player stability
0	4	6	2	5	1	7	3	11	9	10	8	preference vector
	0		6		5		7	7	5	6	0	UIs
								3	1	2	4	

Figure 5 Stability analysis tableau for the Cuban missile crisis (from Fraser & Hipel, 1982).

most prefers outcome 4—in which the Soviets withdraw the missiles without any overt American action—next most prefers outcome 6, and so on.

In Figure 5, the "unilateral improvements" (or "UIs") for each player from each outcome are listed beneath the preference vector of each player. A unilateral improvement is defined to be "an outcome to which a particular player can unilaterally move by changing his strategy, assuming the other player's strategy remains the same" (Fraser & Hipel, 1982). For example, there is a unilateral improvement from outcome 6 to 4 for the United States. From Figure 2 it can be seen that outcome 6 results when the United States blockades and the Soviets withdraw. Thus the United States can induce outcome 4 from 6 by switching to its do nothing option. And since the United States is assumed to prefer 4 to 6 (see Figure 3), a unilateral improvement from 6 to 4 is listed in Figure 5.

To determine which of the twelve outcomes are equilibria, a four step search procedure is necessary. The first three steps involve characterizing the stability, or lack thereof, of each outcome from the point of view of each player. In this regard, Fraser and Hipel (1979) identify three types of outcomes: *rational* (r), *sanctioned* (s), and *unstable* (u).

Rational outcomes are defined as outcomes from which a player has no unilateral improvement. Thus they represent a player's best response to a particular strategy of the other player. In Figure 5, all rational outcomes are designated by an "r" above the preference vector of each player. For example, the best outcomes of both the United States and the Soviet Union are judged rational. Clearly, since these outcomes are best, neither player can induce a better outcome by switching to another strategy, given the associated strategy choice of the other.

Sanctioned outcomes, designated by an "s" in Figure 5, are those outcomes for which the other player can credibly induce a worse outcome for the player who acts upon a unilateral improvement and switches to another strategy. A *credible action* is one which results in a more preferred outcome for the player taking the action. When a player has at least one unilateral improvement for which the other player has no credible sanction, the outcome is unstable, and is designated by a "u."

For example, in Figure 5, a unilateral improvement for the United States exists from outcome 6 to 4. But the Soviet Unon has a unilateral improvement from 4 to outcome 0, which is worse for the United States than 6. Hence not only can the Soviet Union induce a worse outcome for the United States if the United States switches from 6 to 4, but it also has an incentive to do so. Hence outcome 6 is sanctioned for the United States.

By contrast, outcome 5 for the United States is not sanctioned, and hence is unstable. As was just demonstrated, the Soviets can sanction a move to 4, but not to 6. Since 6 is rational for the Soviets, they have no credible response to a United States move to 6. Thus, because the United States cannot credibly be deterred from moving from 5 to 6, and also because it has an incentive to so move, outcome 5 is listed as unstable for the United States.

After designating each outcome as either rational, sanctioned, or unstable, a test for *stability by simultaneity* is required to identify equilibrium outcomes. An outcome unstable for both players is rendered stable by simultaneity if, when both players move from a particular outcome simultaneously, *all* resulting outcomes are worse for both players.

For example, consider outcome 1 in Figure 5, which is unstable for both players. The United States has a unilateral improvement from 1 to 2, and the Soviet Union has a unilateral improvement from 1 to 5. To move from 1 to 2, the United States must switch from its strike option to its blockade option. To move from 1 to 5, the Soviet Union must switch from its maintain option to its withdraw option. If both change at the same time, i.e., the United States blockades and the Soviet Union withdraws, outcome 6 is induced. Since outcome 6 is preferred by both players to outcome 1, outcome 1 is not rendered stable by simultaneity, that is, it remains unstable. Each of the four other outcomes in this game that are unstable for both players are also not stable by simultaneity.

After testing for stability by simultaneity, it is easy to determine which outcomes are equilibria in this game. Any outcome that is not unstable for either player is an equilibrium outcome. Only 4 and 6 pass this test, and in Figure 5 they are designated by an "E." Nonequilibria are denoted by an "X."

Fraser and Hipel (1979, pp. 810,-811) have proved that at least one such equilibrium outcome exists in every game. To distinguish which of several such equilibria will be selected when multiple equilibria exist, they suggest that the status quo outcome, which was outcome 0 when the missiles were discovered, be examined for clues. Although outcome 0 is rational for the Soviets, it is unstable for the United States which has a unilateral improvement to three outcomes, of which its unilateral improvement to 2 is its most preferred. Outcome 2, in turn, is unstable for the Soviets who have a unilateral improvement from 2 to 6. Thus, by moving from 0 to 2, the United States can induce outcome 6, which it also prefers to the original status quo. Of course, 6 is one of the two equilibria in this game. Given the foregoing scenario, one might expect it to evolve, which it did. Such considerations, incidentally, introduce a dynamic element into the Fraser–Hipel methodology.

Notice that Fraser and Hipel have subtly extended the notion of a Nash equilibrium[9] by assuming that players not only assess the immediate consequences of a strategy switch (this is the Nash criterion, i.e., outcomes that are stable because they are rational for both players) but also take into account both the rational response of the other player to a unilateral strategy switch (i.e., outcomes that are rendered stable by sanction), and the ramifications of simultaneous strategy switches by all of the players (i.e., outcomes that are rendered stable by simultaneity). Moreover, the extended notion of an equilibrium outcome they have developed is not merely an exercise in definition. The numerous empirical applications of their new methodology, including the 1956 invasion (Wright et al., 1980) and subsequent nationalization (Shupe et al., 1980) of the Suez Canal, the fall of France (Bennett & Dando, 1977, 1979), the Watergate tapes conflict (Meleskie, Hipel, & Fraser, 1982), the 1979 Zimbabwe conflict (Kuhn, Hipel, & Fraser, 1983), and the Alaskan gas pipeline conflict (Savich, Hipel, & Fraser, 1983), demonstrate that their equilibrium concept has both predictive and explanatory potential.

On the other hand, in some ways, Fraser and Hipel's extension of Nash's stability criterion seems unnecessarily restrictive. More specifically, an arbitrary limit is imposed on the ability of the players to calculate the consequences of moves and countermoves. More specifically, although their equilibrium concept considers the possibility of a sanction being levied against a player's unilateral strategy switch, it does not take into account the possibility of countersanctions, subsequent countersanctions, and so on. Moreover, when strategy switches by both players are considered, only the consequences of simultaneous switches are examined, thus ignoring, again, the possibility of sequential or alternating moves by the players. Put in a slightly different way, the criteria for stability advanced by Fraser and Hipel, while less myopic than Nash's, do not seem farsighted enough.

Nevertheless, the Fraser–Hipel methodology does offer exciting possibilities for analyzing a wide range of real-world conflicts. It is applicable to both two-person and n-person games. Moreover, Fraser and Hipel have developed mathematics that permit a computer analysis, using their basic algorithm, of exceptionally large and complex conflicts. And finally, they have devised

[9] It is worth noting that the cooperative outcome, (3,3), in Prisoners' Dilemma is rendered stable if the criteria suggested by Fraser and Hipel are adopted.

methods for dealing with games in which some players might have erroneous or partial information, called *hypergames* (Takahashi, Fraser, & Hipel, in press), as well as for studying the effects of shifting or changing preferences, thus adding an additional dynamic element to their model. Further refinements and extensions of their new methodology can only be eagerly awaited.

2. Nonmyopic Equilibria. The concept of *nonmyopic equilibrium,* recently developed by Brams and Wittman (1981), suggests one way to extend the stability criteria of both Nash and Fraser and Hipel. Unlike both of these equilibrium notions, the concept of a nonmyopic equilibrium places no arbitrary limitation on the number of moves and countermoves players can make. More specifically, it assumes that the following rules of play operate in 2×2 ordinal games:

1. Both players simultaneously choose strategies, thereby defining an *initial outcome* of the game.
2. Once at an initial outcome, either player can unilaterally switch his strategy and change that outcome to a subsequent outcome.
3. The other player can respond by unilaterally switching his strategy, thereby moving the game to a new outcome.
4. These strictly alternating moves continue until the player with the next move chooses not to switch his strategy. When this happens, the game terminates, and the outcome reached is the *final outcome* (Brams & Hessel, 1982).

The concept of a nonmyopic equilibrium also assumes that players are able to anticipate the consequences of strategy choices made in games governed by these rules. Put another way, this equilibrium concept is a look-ahead idea that assumes that a player will evaluate the consequences of departing from an initial outcome, taking into account the probable response of the other player, his own counterresponse, subsequent counterresponses, and so on. If, for *both* players, the starting outcome is preferred to the outcome each player calculates he will end up at if he makes an initial departure, the starting outcome is a nonmyopic equilibrium.[10]

Nonmyopic equilibria exist in 37 of the 78 distinct 2×2 ordinal games identified by Rapoport and Guyer (1966). Of these 37 games, only two have nonmyopic equilibria that are not also Nash equilibria. One of these is Prisoners' Dilemma. Significantly, the other is "Chicken," another notorious game that has received considerable attention in the literature. In both of

[10] It is important to point out that in addition to surviving the backward induction process on the tree for both players, for an outcome to be a nonmyopic equilibrium the process must terminate, that is, it must not cycle back to the original outcome. Brams and Wittman assume that there will be no cycle if a node exists on the game tree whereby the player with the next move can ensure his best outcome by staying at it.

these games, the cooperative outcome is stable in the nonmyopic sense but not in the sense of Nash. Thus this new equilibrium concept provides a rationale for cooperation in precisely the two games for which the question of cooperative behavior is the most salient and problematic.

To see this, and to illustrate the calculation of this new equilibrium concept, consider the game tree depicted in Figure 6, which lists the sequence of moves and countermoves implied by a departure of Player A away from (3,3) in the Prisoners' Dilemma game of Figure 1. A's incentive to move from this outcome can be determined simply by working backwards up the game tree and asking what the rational choice of each player is at each node or decision point. If the outcome that is implied by this process is inferior to (3,3) for the player postulated to have the first move on the tree—in this case, A—then this outcome is stable in the nonmyopic sense for this player. If a similar calculation also reveals that this outcome is stable in the nonmyopic sense for B,

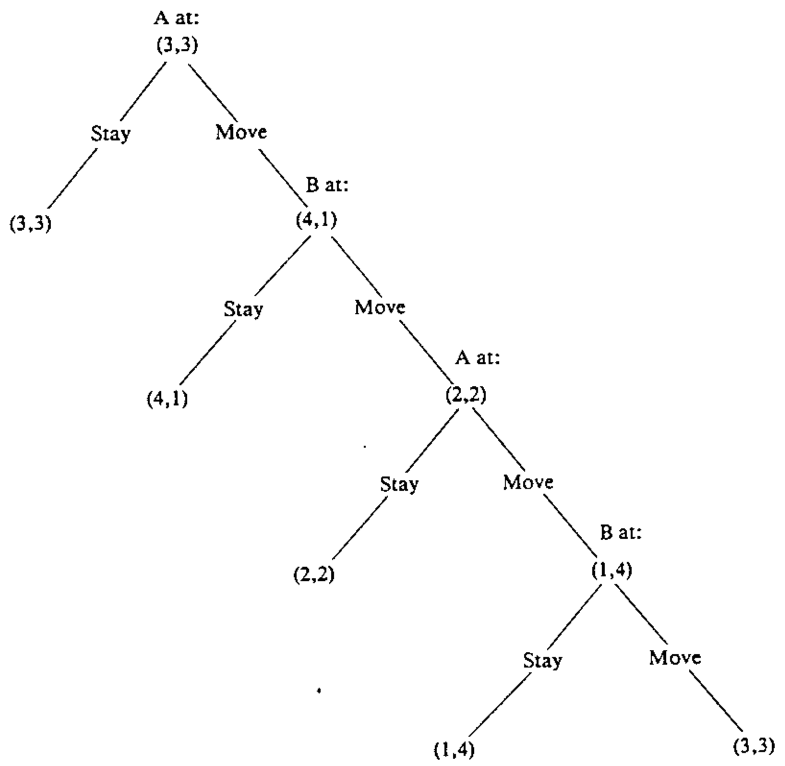

Figure 6 Game Tree representation of moves in Prisoners' Dilemma, starting with A at (3,3).

then it is a nonmyopic equilibrium outcome. On the other hand, if the outcome implied by a departure from an initial outcome by either player, when he is postulated to have the first move on the tree, is superior to the initial outcome, then this outcome is not stable in the nonmyopic or long-term sense.

At the last node on the tree, B must choose between staying at (1,4)—his best outcome—or moving to (3,3)—his next-best outcome. Clearly, should this node be reached in a sequence of moves and countermoves, B will rationally choose to stay at (1,4).

But would such a sequence ever rationally get to this point if (3,3) is the initial outcome? To determine this, consider A's choice at (2,2). At this node, A is faced with a choice of staying at (2,2)—his next-worst outcome—or moving to (1,4)—his worst outcome. Given this choice, A would not switch strategies, thereby terminating the sequence of moves before B can choose at (1,4).

Given A's rational choice at (2,2), what should B do at the preceding node? Here B can decide to stay at (4,1)—his worst outcome—or move to (2,2)—his next-worst outcome—which, because of the expected choice of A, would become the final outcome. For B, the rational choice is to move to (2,2).

What should A do at the previous node, (3,3)? A can either stay at (3,3)—his next-best outcome—or move to his best outcome at (4,1). However, as was just illustrated, a move to (4,1) implies (2,2) as the final outcome in a sequence of moves and countermoves. Since A prefers (3,3) to (2,2), he should rationally choose to stay at the original status quo. And since A has no long-term incentive to depart from (3,3), the initial outcome is a nonmyopic equilibrium *for A*.

It is also a nonmyopic equilibrium for B. By symmetry, the calculus facing B at (3,3) is identical to that of A. And since neither player has a long-term incentive to move away from (3,3), this outcome is a nonmyopic equilibrium in this game.

It should be pointed out, however, that (3,3) is not the only nonmyopic equilibrium outcome in Prisoners' Dilemma. The noncooperative outcome (2,2)—the unique Nash equilibrium and the conventional solution to this game—is also stable in the nonmyopic sense. If this outcome were the initial outcome, neither player would have an incentive to change his strategy because the player with the subsequent move would immediately terminate the process at the outcome best for him and worst for the departing player—either (4,1) or (1,4).

More significant, though, is that (2,2) "absorbs" all other outcomes, including itself, except for (3,3). This means that should either (1,4) or (4,1) be reached in a sequence of alternating strategy choices, the outcome that would be chosen would be (2,2), not (3,3). Thus the very calculations that enhance the stability of the compromise outcome can also undermine it.

Nevertheless, the cooperative outcome remains a nonmyopic equilibrium in this game. It thus provides a solution of sorts to Prisoners' Dilemma. Provided that the cooperative outcome is the initial outcome and that players can make, and evaluate the consequences of, an unlimited number of moves and countermoves, a strategy supporting the compromise outcome in a Prisoners' Dilemma game is both farsightedly rational and stable in the long-term.

In addition to providing a new rationale for cooperative behavior in Prisoners' Dilemma (and Chicken), the concept of a nonmyopic equilibrium possesses a number of other features that render it both theoretically attractive and empirically relevant. Perhaps the most appealing theoretical characteristic of this new equilibrium concept is that it introduces a dynamic element into a normal-form, game-theoretic analysis. In addition, it is based on assumptions that closely match the actuality of many real-life political games; it is readily calculable and interpretable; and since it is based only on ordinal utilities, it requires fewer "heroic assumptions" than other decision-making models that rest upon the notion of cardinal utility.

It is also worth pointing out that the underlying dynamic approach postulated by Brams and Wittman has been extended in a number of interesting ways. For example, Brams and Hessel (1982) have explored the absorbing properties of outcomes in the 41 2 × 2 games without nonmyopic equilibria. Zagare (1984b) has examined the consequences of limitations on the ability of players to make all of the logically possible moves and countermoves in a 2 × 2 game. And Kilgour (1984) has defined an extended nonmyopic equilibrium for players who are able to predict the ramifications of an unlimited number of strategy changes, including those that result in "cycling back" to outcomes that have already been reached. In addition, several intuitively satisfying notions of power have been defined and the implications of these definitions within a sequential framework explored. To wit: Brams (1982a, 1982b) has studied the effect on outcome stability of "moving power," or the ability of one player to move indefinitely after the other player is forced to stop after some finite number of moves. Brams and Hessel (1983, 1984) have examined the consequences of both "staying power," wherein one player can make his initial strategy choice after the other and stay at this outcome until after the other player moves from it, and "threat power," where in an iterated game, one player can threaten to move to a Pareto-inferior outcome to deter or compel the other player from making undesired choices in subsequent plays of the game. (A comparison and synthesis of these various notions of power is given in Brams, 1983). And Zagare (1985, in press) has explored the implications of this new conceptual framework for deterrence in one-shot games. Finally, like the methodology devised by Fraser and Hipel, the predictive and explanatory potential of the Brams–Wittman framework has been demonstrated in several empirical applications, including the Polish strategic situation of 1980–81 (Brams & Hessel, 1984), and the Middle East conflicts of 1967 and 1973 (Zagare, 1981, 1983).

Though the concept of a nonmyopic equilibrium has been extended in several directions relevant to the study of politics, some significant issues remain unanswered, and many important theoretical questions have yet to be examined in a systematic way. For example, this new equilibrium concept has only been fully developed for 2×2 games and its extension to larger games remains problematic (but see Kilgour, 1984 for one possible extension). In addition, not all 2×2 games have outcomes stable in the nonmyopic sense, although, with somewhat stronger assumptions, a similar type of stability can be induced in these games (Brams & Hessel, 1982). And the effects of backtracking on outcome stability have not yet been completely studied. Still, the dynamic framework developed by Brams and Wittman, and extended by others, represents a maturing methodology and provides a promising starting point for examining the conditions leading to stability (or the lack thereof) in many open-ended political games in which decision-makers think seriously about the long-term as well as the short-term ramifications of their actions. As such, it is certain to inspire further theoretical modification and refinement and empirical testing and application.

3. **Supergames.** Through the concept of a nonmyopic equilibrium provides one way of resolving the tension between individual and collective interests, the conditions underlying the stability of the compromise outcome in Prisoners' Dilemma are somewhat restrictive and may not always be satisfied. Are there any other conditions under which players might cooperate in games of this type?

Almost from the time that the paradoxical nature of Prisoners' Dilemma was first recognized, game theorists have speculated that cooperation *is* rational when the players are faced with repeated plays (called *supergames*) of this game with the same opponent. For instance, despite presenting evidence that the use of any strategy that is an equilibrium strategy in a Prisoners' Dilemma supergame of finite and known length will result in the repeated selection of the noncooperative outcome, (2,2), throughout the sequence of games, Luce and Raiffa (1957, p. 101) asserted that they would not choose (D) at every move if this game were played more than once, but would try to teach the other player to cooperate by rewarding him if he does and punishing him if he does not. And Davis (1983, p. 113) has suggested, though without proof, that Luce and Raiffa's argument has merit, but only when the number of times the game is iterated is not known.

Is this intuition justified? Taylor (1976) has provided a provocative way of addressing this question. He begins, first, by assuming that the number of times the game will be repeated is not fixed.[11] Following an earlier suggestion by Shubik (1970), he also assumes that players discount the value of their

[11] Iterated games in which the last game occurs with some random probability known to the players are usually referred to as "stochastic games." For a discussion of Prisoners' Dilemma played under these conditions, see Hill (1975).

future payoffs, that is, that they value a present payoff more than a payoff received at some unspecified time in the future. This does not mean that future payoffs are seen as worthless, just worth less.[12] And finally, since the number of strategies in an iterated game is potentially very large, Taylor assumes that only the following five strategies are available to the players:

D^∞: (D) is chosen every time the game is repeated.
C^∞: (C) is chosen every time the game is repeated.
A_k: (C) is chosen in the first game and in every subsequent game as long as the other player chooses (C). If the other player chooses (D), (D) is selected for the next k games regardless of what the other player does. Then (C) is chosen until the other player chooses (D), in which case (D) is chosen for the next $k+1$ games and so on.
B: (C) is chosen in the first game and the choice of the other player is chosen in the next and subsequent games.
B′: Same as B except that (D) is chosen in the first game. (B and B′ are two variations of a tit-for-tat strategy).

Arguing that these five strategies are the ones most likely to be considered by players in an iterated Prisoners' Dilemma game, at least on a conscious level, Taylor (1976, pp. 32–33) asked whether any of the $5 \times 5 = 25$ outcomes resulting from these five strategies are (Nash) equilibria in the Prisoners' Dilemma supergame.

Not surprisingly, Taylor proves that (D^∞,D^∞) will always be an equilibrium in the iterated game, since a single player cannot do better by unilaterally switching to one of the four other strategies. Conversely, (C^∞,C^∞) is never an equilibrium since one player can always improve his payoff by switching to (D) for the remaining plays of the game. In addition, as Taylor (1976, pp. 31–43) shows, depending on the rate at which the players discount future payoffs and the value of the payoffs in the component games, the following are sometimes equilibria:

1. The four pairs in which each player uses either A_k or B. In each case, the outcome is mutual cooperation in every ordinary game throughout the supergame.
2. The three pairs in which each player chooses B′ and the other player B′ or D. In each case, the outcome is mutual defection throughout the supergame.
3. The two strategy pairs (B,B′) and (B′,B). Here the outcome is an alteration throughout the supergame of.(C,D) in one ordinary game and (D,C) in the

[12] The analysis of the payoffs to players in a repeated game requires that the sum of the payoffs of each component game be finite. Such is the case if discounting is assumed (Taylor, 1976, pp. 29–30). For a discussion of other assumptions that imply a finite payoff to players in a repeated game situation, see Harris (1969) and Rapoport (1967).

next, beginning with (C,D) in the first game in the case of (B,B′) and with (D,C) in the case of (B′,B).

In an empirical sense, equilibria of category (3) do not seem relevant for many situations (e.g., an arms race) involving repeated plays of a Prisoners' Dilemma game. Because minor complications are introduced when category (3) equilibria are admitted,[13] they will be rejected on empirical grounds to facilitate the subsequent discussion.

If category (3) equilibria are eliminated from consideration, then it is possible for some or all four of the mutual defection equilibria to coexist with some or all of the mutual cooperation equilibria of category (1). Basically, this occurs if each player's discounting of future payoffs is "sufficiently low," that is, if each player does not prefer the payoff resulting from unilateral defection in the first component game and mutual defection in all subsequent component games to the payoff resulting from mutual cooperation throughout the repeated game (Taylor, 1976, p. 89). Since each mutual cooperation equilibrium is both equivalent and interchangeable with every other one, and in addition is Pareto-optimal under these conditions, it is clear that if at least one equilibrium of category (1) exists, the outcome of the iterated Prisoners' Dilemma game will result in the repeated selection of the cooperative outcome on every move of the game!

Before euphoria takes over, however, it is important to remember that this result applies only when the limited number of strategies considered by Taylor are available to the players. But as two recent computer tournaments (Axelrod, 1980a, 1980b) demonstrate, these five strategies hardly exhaust the set of reasonable strategies. Hence, the relevance of Taylor's results to a wide variety of empirical situations must remain suspect.

By contrast, Axelrod's (1981) "evolutionary approach" to the iterated Prisoner's Dilemma game is not restricted by arbitrary limits placed upon the number of strategies available to the players. This approach, which considers all possible strategies, posits "the existence of a whole population of individuals employing a certain strategy, B, and a single mutant individual employing another strategy, A" (Axelrod, 1981, p. 310).

If the players in this game interact with each other one at a time, then it is possible that the expected payoff of an individual using strategy A is higher than the expected payoff of a member of the general population. In this case, strategy A is said to *invade* strategy B. But if the converse is true, and if no other strategy can invade B, B is said to be *collectively stable.*[14]

Are there any conditions under which a cooperative strategy can invade a

---

[13] For a brief discussion, see Taylor (1976, pp. 88–89).

[14] Axelrod's conception of a collectively stable strategy is based upon the idea of an *evolutionary stable strategy* developed by Maynard Smith and Price (1973) to study problems of biological evolution. For a discussion of this concept and subsequent developments in this field, see Maynard Smith (1982).

noncooperative strategy in an iterated Prisoners' Dilemma game? And are there any conditions under which a cooperative strategy is collectively stable in such a game?

To answer this question, Axelrod, like Taylor, assumes a discount parameter, w – where $0 \leq w \leq 1$ – which can be interpreted as before, or as an estimate that an individual player makes of the probability of encountering the same opponent in a future game. Thus, the smaller w is, the less important future payoffs become.

Interestingly, Axelrod proves that a tit-for-tat strategy that cooperates in the first game and then reciprocates the previous choice of the other player in each subsequent game – Taylor's (B) strategy – is a collectively stable strategy, provided that w is sufficiently large. Unfortunately, a strategy of selecting (D) on every play of the iterated Prisoners' Dilemma game is *always* a collectively stable strategy, regardless of the value of w. This means that a population of unconditionally noncooperative players (called *meanies*) cannot be invaded by conditionally cooperative players arriving in the population one at a time.

To some extent, however, this dismal conclusion can be mitigated if newcomers arrive in clusters, rather than individually, and if the newcomers interact with each other more than they interact with members of the general population. As Axelrod shows, under these conditions, a number of strategies *can* invade a world of meanies. Of those strategies that can invade a population of noncooperative players, *maximally discriminating strategies* require the smallest amount of interaction among members of the invading cluster. "A strategy is *maximally discriminating* if it will eventually cooperate even if the other has never cooperated yet, and once it cooperates it will never cooperate again with ALL D but will always cooperate with another player using the same strategy" (Axelrod, 1981, p. 316). Significantly, tit-for-tat is one such maximally discriminating strategy.

From this and related results, Axelrod (1981) concludes that

> cooperation can emerge even in a world of unconditional defection. The development cannot take place if it is tried only by scattered individuals who have no chance to interact with each other. But cooperation can emerge from small clusters of discriminating individuals, as long as these individuals have even a small proportion of their interactions with each other. Moreover, if nice strategies (those which are never the first to defect) eventually come to be adopted by virtually everyone, then those individuals can afford to be generous in dealing with any others. The population of nice rules can also protect themselves against clusters of individuals using any other strategy just as well as they can protect themselves against single individuals....So mutual cooperation can emerge in a world of egoists without central control, by starting with a cluster of individuals who rely on reciprocity. (p. 317)

This is a long way from the pessimistic assessment usually associated with this game, and confirms what was suspected all along, namely that cooperation *is* rational in the indefinitely repeated Prisoners' Dilemma game. Nevertheless, because this result—and the results suggested by Brams and Wittman (1981) and Taylor (1976)—depend upon highly specific sets of assumptions, they should sensitize us to the fact that cooperation is not automatic and may be difficult to achieve in the real world. Stated optimistically, these findings suggest that mutual cooperation in a Prisoners' Dilemma situation is possible if players take a long-run view of things and consider the future consequences of their present actions. Unfortunately, in many important political games, it seems as if players take too seriously John Maynard Keynes's well-known dictum that in the long-run we will all be dead.

IV. COOPERATIVE GAME THEORY

A. Cooperative Game Theory: A Theoretical Overview

The cooperative branch of game theory differs in both spirit and appearance from its noncooperative cousin. Most of these differences stem from the assumption, missing in the noncooperative theory, that there are rules that allow the players to negotiate binding agreements with each other.

The assumption about the possibility of binding agreements that distinguishes cooperative from noncooperative game theory has important implications for the outcomes that each branch of theory predicts are likely to be selected by rational players. As already indicated, since the noncooperative theory assumes that binding agreements are not possible, attention is focused almost exclusively on outcomes that are, in effect, self-enforcing, that is, are equilibrium outcomes. By contrast, in the cooperative theory, no special status is afforded to equilibrium outcomes—as previously defined—since players are assumed to be able to commit themselves to any outcome. Instead, cooperative game theory attempts to single out those outcomes from among the set of possible outcomes that rational players would agree to. Minimally, it is assumed that such an outcome must satisfy both the conditions of *individual rationality* and *group rationality*.

Individually rational outcomes guarantee each player a payoff at least as good as the payoff he can ensure himself without the assistance of any other player. Similarly, outcomes that satisfy the condition of group rationality (or Pareto-optimality) provide the set of all players with a payoff equal to the payoff that the group can guarantee itself.

The set of outcomes that satisfy the conditions of individual and group rationality in a two-person game define the *von Neumann-Morgenstern solution* (also called the *negotiation set*) for this category of games (von Neumann & Morgenstern, 1953). In n-person games, outcomes that satisfy these two conditions are termed *imputations*.

Since the von Neumann-Morgenstern solution and the set of imputations typically contain an infinitely large number of outcomes, they do not provide particularly satisfying solutions to their respective categories of cooperative games. Consequently, in order to induce more determinate results, game theorists have imposed more restrictive conditions on the set of possible outcomes. For a variety of reasons, most of the resultant theory is either not compelling or not of immediate relevance to students of politics.

For example, in two-person cooperative game theory, most formal work begins with the imposition of normative conditions that theorists feel a rational agreement should satisfy. (For this reason, Luce and Raiffa (1957) refer to these models as "arbitration schemes.") Hence most of two-person cooperative game theory is totally devoid of a descriptive interpretation. And although several solution concepts in n-person cooperative theory admit both a descriptive and normative interpretation, existent solution concepts are also deficient because for many games they are either empty (e.g., the core), or too inclusive (e.g., the von Neumann–Morgenstern V-solution), or because they are defined by assumptions that are not particularly germane to the study of politics.[15]

B. Cooperative Game Theory: Recent Advances

In this section, two possible exceptions to the foregoing characterization of cooperative game theory are discussed, a solution to finite two-person cooperative games proposed by Rice (1979), and a solution to n-person cooperative games, called the *competitive solution,* put forth by McKelvey, Ordeshook, and Winer (1978).

1. The Rice Solution for Two-Person Games. Rice's (1979) attempt to define a solution for a two-person cooperative game is comprised of both a description of a negotiation scenario and an examination of the rational consequences of the process he postulates. The inherent plausibility of this scenario—which will not be fully described here—as well as the fact that it leads to a unique outcome that satisfies several properties characteristic of the traditional, axiomatic solution concepts, including both individual and group rationality, renders Rice's solution worthy of attention.

In Rice's model, players are assumed to have the option of choosing to negotiate. They will only negotiate when they can guarantee a better outcome than they can ensure by not negotiating. If they both choose to negotiate, Rice assumes that their behavior is characterized by threats and promises rendered credible by some "mechanism external to the game" (Rice, 1979, p. 565). In other words, Rice assumes that each player is able to commit himself to any available strategy.

[15] For example, the assumption of "transferable utility," which is a standard simplifying assumption, renders most of existent cooperative game theory inappropriate for examining those segments of the political world (e.g., legislatures) that are characterized by indivisible outcomes (McKelvey, Ordeshook, and Winer, 1978).

In the negotiation scenario posited by Rice, one player is assumed to move first. In some games, the players will not be indifferent towards the order of play, since the Rice solution may be different when one player, rather than the other, makes the first move. But in the real world, Rice argues, this issue is oftentimes resolved by circumstances. And when this question is not resolved, negotiations might break down. Negotiations will also break down if the player who makes the second move anticipates a lower payoff than the payoff he anticipates if he does not negotiate. The player with the first move will always do better by negotiatiing than by not negotiating.

If both players agree to negotiate, Rice assumes that the player with the first move begins by specifying what he will do in response to each possible choice of the other player. The problem for this player, then, is to determine that combination of threats and promises that produces his highest payoff.

For example, in the game depicted in Figure 7, if Player A is assumed to make the first move, he can guarantee a payoff of 15 by "promising" to select a_2 if Player B selects b_1, and "threatening" to select a_4 if Player B selects any other strategy. If Player A can commit himself to this combination of threats and promises, Player B will do best and receive a payoff of 3 if he selects b_1, and will receive a payoff of less than 3 (either 2, 1, or 0) if he selects b_2, b_3, or b_4, respectively. No other combination of threats and promises induces a better outcome for Player A. Similarly, if Player B moved first, he could also ensure a payoff of 15 by promising to select b_2 in response to a_1, and threatening to select b_4 in response to any other strategy selected by Player A.

Will this game be negotiated? Rice says "no" since either player will receive a payoff of only 3 if he makes the second move in the negotiated game, but will receive payoff of 7 [associated with the unique Nash equilibrium, (7,7)] if the game is played noncooperatively and negotiations do not take place.

By contrast, Rice argues that the game in Figure 8 would be negotiated. In the negotiated game, the Rice solution is (9,4) or (4,9), depending on whether Player A or Player B moves first, while the unique Nash equilibrium of the

		Player B			
		b_1	b_2	b_3	b_4
Player A	a_1	(14,14)	(3,15)	(4,9)	(13,11)
	a_2	(15,3)	(10,10)	(5,8)	(2,12)
	a_3	(9,4)	(8,5)	(7,7)	(1,6)
	a_4	(11,13)	(12,2)	(6,1)	(0,0)

Figure 7 (From Rice, 1979, p. 571, reprinted by permission of Sage Publications, Inc.).

		Player B		
		b_1	b_2	b_3
	a_1	(8,8)	(4,9)	(1,7)
Player A	a_2	(9,4)	(5,5)	(2,6)
	a_3	(1,1)	(6,2)	(3,3)

Figure 8 (From Rice, 1979, p. 574, reprinted by permission of Sage Publications, Inc.).

nonnegotiated game, (3,3), is less than the payoff for either player in the negotiated game.

Given this view of the negotation process, it is clear that the external enforcement mechanism posited by Rice must be activated by the consent of both players. This suggests that each player, regardless of whether he moves first or second, would have a certain amount of bargaining leverage over the other. Yet, curiously, the Rice solution would seem to apply only to those situations wherein the player with the second move is completely unable to use this leverage in any meaningful way.

To see this, consider again the game of Figure 8. Recall that the unique Nash equilibrium in this game, and presumably the solution if the game is not negotiated, is (3,3). To sustain this supposition, Rice must assume that only (Nash) equilibrium strategies are credible in the nonnegotiated game. This means that neither player can be said to possess what Brams and Hessel (1984) call "threat power," or the ability to induce a better outcome by credibly threatening to move to a Pareto-inferior outcome. Put in a slightly different way, Rice implicitly assumes that the nonnegotiated game is played between players of approximately equal power.

But what if this game is negotiated and Player A is assumed to make the first move. As already indicated, Player A can propose (9,4) with the proper combination of threats and promises and, according to Rice, Player B would rationally accept this proposal, since it is better than the payoff he would expect in the nonnegotiated game. Similar reasoning would suggest, however, that Player B could make a counteroffer – say, (4,9) or even (8,8). Would not Player A also rationally accept either of these two offers? Since neither player is assumed to have superior negotiating power (see the foregoing) the answer must be in the affirmative. Of course, Player A would prefer to make a counteroffer to Player B's counteroffer, and Player B would prefer to make a subsequent counteroffer, and so on. Consequently, when the order in which the players move is not fixed, the Rice solution is essentially indeterminate.

As Rice argues, however, this indeterminacy is eliminated if and when the negotiation order is set and is itself not negotiable. Moreover, this order must be exogenously determined since, as just argued, neither player can be assumed to have a power advantage. But this leads to a small problem. How can the negotiation order be both fixed and exogenously determined on the one hand, and triggered by the consent of the players on the other? Either the order in which the players make their moves is fixed or it is not. If it is fixed in the negotiated (cooperative) version of this game, then one would expect it to be fixed in the nonnegotiated (noncooperative) variant, though Rice implicitly assumes that it is not. And if it is not fixed, then one would expect that it is negotiable, in which case the Rice solution may be indeterminate.

There are, of course, games like Prisoners' Dilemma for which the order in which the players move is inconsequential since the Rice solution is the same for each player, regardless of whether he moves first or second. Still, this should not obscure the fact that for the Rice solution to be determinate and, hence empirically meaningful in games other than these, both players must be able and willing to activate an external enforcement device that is (a) predetermined, nonnegotiable, and unalterable, and (b) may confer a distinct advantage on one of them. Such restrictive requirements, though, seriously limit the real-world relevance of Rice's solution concept. Thus one must conclude, unfortunately, that Rice's proposed solution, while theoretically elegant, is limited, by its assumptions, in its empirical applicability.

2. The Competitive Solution for *N*-person Games. Implicit in the notion of the *competitive solution* recently developed by McKelvey, Ordeshook, and Winer (1978) is the supposition that *protocoalitions,* or potential winning coalitions in an n-person game, compete with each other for pivotal players who can render a protocoalition a winning coalition. The competition is assumed to take the form of an auction in which the bids are outcomes (called *proposals*) that can be enforced if and when the protocoalition reaches winning size.

Clearly, the set of proposals is identical to the set of possible outcomes. Some proposals, however, are not *viable.* Specifically, if the set of players who are pivotal between two protocoalitions strictly prefer one outcome, say A, to another outcome, say B, B is not considered to be a viable proposal with respect to the two protocoalitions and the set of pivotal players. Proposals that are not viable will (presumably) be rejected by rational players and need not be considered as likely societal states.

To define a competitive solution for an n-person cooperative game, McKelvey, Ordeshook, and Winer introduced the notion of a *balanced set* of proposals. A set of proposals is balanced if (a) each distinct proposal is associated with a different coalition, and (b) each proposal in the set is viable against all other proposals in the set.

In some games, many balanced sets of proposals may exist, but only proposals from those balanced sets that are unable to be *upset* are likely to be considered by pivotal players. A balanced set of proposals is upset if a proposal exists that is viable against every proposal in the set and is strictly preferred by the pivotal players associated with at least one proposal in the set. Balanced sets of proposals that cannot be upset are said to constitute a competitive solution to an *n*-person game.[16]

To illustrate these concepts with an example, consider a legislature composed of three players (or three factions) that must decide which of three bills to pass. The utilities that each player is assumed to attach to each of the eight possible combinations of these three bills, including passing none or all of them, are listed in Figure 9.

Since this legislative body is also assumed to operate by majority rule, each player needs the support of at least one other player to constitute a winning coalition. Suppose, then, that legislator 1 and legislator 3 are competing for the support of legislator 2. Legislator 1 could propose to pass all three bills if legislator 2 joins him in a coalition, thereby inducing a payoff of 3 and 1 to legislators 1 and 2, respectively. But this offer is not viable since legislator 3 can make a counteroffer (i.e., to pass bills 2 and 3) that is preferred by legislator 2 to the offer of legislator 1.

By contrast, legislator 3's offer to legislator 2 is viable since legislator 2 is indifferent between it and the best counteroffer that legislator 1 can make (i.e., to pass bills 1 and 2). For similar reasons, legislator 1's counteroffer is also viable.

What if legislators 1 and 2 were competing for legislator 3? In this case, both legislator 1's proposal to pass bills 1 and 3, and legislator 2's proposal to pass bills 2 and 3, constitute viable proposals. And finally, if legislators 2 and 3 were competing for legislator 1, both legislator 2's proposal to pass bills 1 and 2, and legislator 3's proposal to pass bills 1 and 3 are viable.

Given these considerations, the following set of proposals constitute a (strong) competitive solution for the game depicted in Figure 9:

MAJORITY DECISION	PROPOSAL	
pass only bills 1 and 2	(4,3, −6; [1,2])	
pass only bills 1 and 3	(4, −4,3; [1,3])	≡K
pass only bills 2 and 3	(−2,3,3; [2,3])	

First note that this set is composed only of proposals that are viable against each other, thereby rendering this a balanced set. And second, observe that there is no proposal in this set that can be upset by any proposal not in the set,

[16] Slightly more stringent requirements define a *strong competitive solution*. For the details, see McKelvey, Ordeshook, and Winer (1978).

Legislator	Bills Passed by the Committee							
	None	1	2	3	1 & 2	1 & 3	2 & 3	All
1	0	5	−1	−1	4	4	−2	3
2	0	−2	5	−2	3	−4	3	1
3	0	−3	−3	6	−6	3	3	0

Figure 9 Three-Person, three-bill logrolling example (from McKelvey, Ordeshook, & Winer, 1978, p. 601).

that is, there is no alternate proposal that is preferred by each member of any winning coalition. Thus the requirements that define a competitive solution are met.

The competitive solution possesses a number of characteristics that render it extremely attractive as a solution for n-person cooperative games. First, it is based on an intuitively satisfying and behaviorally meaningful notion of coalition competition. In addition, it is a general solution that does not depend upon restrictive assumptions (e.g., transferable utility) that divorces it from many meaningful political situations, although, as McKelvey, Ordeshook, and Winer, (1978, p. 614) point out, the competitive solution seems especially applicable when bargaining among the players is unrestricted. Moreover, since the competitive solution is equivalent to both the core and the (main-simple) von Neumann-Morgenstern V-solution when they exist, "it is a natural extension of these classical solution concepts" (McKelvey, Ordeshook, & Winer, 1978, p. 600). Existant experimental support for these two solution concepts, therefore, can also be interpreted as corroborating evidence for the empirical validity of the competitive solution.[17]

On the other hand, crucial theorems concerning both the existence and the uniqueness of the competitive solution do not exist. And since the competitive solution has yet to be applied to a real-world political problem, further theoretical and empirical research is required before a final judgment can be made about its explanatory and predictive power.

V. SUMMARY AND CONCLUSIONS

It is not easy to offer a summary evaluation of the advances made in the game theory and politics literature during the last decade. The primary reason for this is that game theory is not really a single theory—at least the way most political scientists think of the term "theory"—but is rather a collection of different theories that are interrelated and connected, but not fully inte-

[17] For contrary evidence, see McKelvey and Ordeshook (1983).

grated. And, as one might expect in a maturing science, developments within the various strands of theory are uneven and do not necessarily mirror one another.

Still, some general observations can be gleaned from the pockets of research reviewed in this essay. First, it seems clear that the achievements of the noncooperative branch of game theory have outstripped those of the cooperative theory. This is perhaps because the problems normally addressed by the noncooperative theory (e.g., two-person versus n-person games) are inherently more tractable. Nevertheless, the research examined herein does indicate that noncooperative game-theoretic models are emerging from the normative-theoretical realm into a form whereby they can provide more compelling explanations and sounder predictions about political activity. This development stems from the more realistic assumptions that underlie these models and from a greater appreciation by theorists of the need to adapt and modify basic, unadorned constructs to the exigencies of the real world.

Second, the notions of rational behavior and of outcome stability have undergone considerable refinement during the past ten years.[18] Coincidentally or not, all of the efforts revived in this essay that have attempted to modify Nash's equilibrium concept have sought to extend it by incorporating into each player's calculus the long-term consequences of their present actions. For instance, both Taylor's and Axelrod's work examine the rational implications of particular strategy choices for player payoffs in future plays of a game with the same, or similar opponents, whereas Fraser and Hipel's technique for analyzing complex conflicts, and Brams and Wittman's notion of a nonmyopic equilibrium, consider these ramifications within the confines of a single game. Significantly, in some way, each of these efforts also provides a resolution of the Prisoners' Dilemma game, and thereby augments our understanding of the conditions that are conducive to player cooperation when there is a conflict between individual and collective interests. This latter achievement is perhaps the single most impressive addition to the literature in recent years.

All of which is not to say that there have been no achievements within the cooperative branch of the theory. Rice's efforts to evolve a solution for two-person cooperative games is noteworthy beyond the areas for which it is immediately applicable, because it represents the first attempt of consequence to study these games when the strategies available to the players are finite in number, and when the outcomes produced by these strategies are not divisible. Similarly, by also avoiding the assumption of transferable utility, McKelvey, Ordeshook, and Winer's definition of a competitive solution for

[18] For a mathematical analysis and comparison of several noncooperative equilibrium concepts, see Kilgour, Hipel, and Fraser (1984).

n-person games avoids many of the limitations the more standard solution concepts possess when they are applied to political games. Both of these efforts, therefore, are encouraging because they constitute fundamental theoretical refinements of a form that should enhance the relevance of this strand of game theory to political scientists.

In conclusion, it appears that the game theory and politics literature has reached a mature stage. Concepts and models that have immediate import to the study of politics are being developed. Moreover, political scientists themselves are for the first time sharing in this development of the theory. This is a healthy trend because it portends the evolution of the paradigm in ways that will make more likely the realization of a positive theory of politics within a game-theoretic perspective.

REFERENCES

Axelrod, R. (1980a). Effective choice in the Prisoner's Dilemma. *Journal of Conflict Resolution, 24,* 3–25.

Axelrod, R. (1980b). More effective choice in the Prisoner's Dilemma. *Journal of Conflict Resolution, 24,* 379–403.

Axelrod, R. (1981). The emergence of cooperation among egoists. *American Political Science Review, 75,* 306–318.

Bennett, P.G., & Dando, M.R. (1977). "Fall Gelb" and other games: A hypergame perspective of the fall of France, 1940. *Journal of the Conflict Research Society, 1,* 1–32.

Bennett, P.G., & Dando, M.R. (1979). Complex strategic analysis: A hypergame perspective of the fall of France. *Journal of the Operational Research Society, 30,* 23–32.

Borel, E. (1921). La Theorie du jeu et les équations intégrales á noyau symétrique. *Comptes Rendus de l'Academie des Sciences, 173,* 1304–1308.

Borel, E. (1924). "Sur les jeux on interviennent l'hasard et l'habilité dese joueurs." In J. Herman (Ed.), *Theorie des probabilitiés,* (pp. 204–241) Paris: Librairie Scientifique.

Brams, S.J. (1975). *Game theory and politics.* New York: Free Press.

Brams, S.J. (1982a). Omniscience and Omnipotence: How They May Help—or Hurt—in a Game. *Inquiry, 25,* 217–231.

Brams, S.J. (1982b). A resolution of the paradox of omniscience. In M. Bradie and K. Sayre (Eds.), *Reason and decision.* Bowling Green, OH: Applied Philosophy Program, Bowling Green State University.

Brams, S.J. (1983). *Superior beings: If they exist, How would we know?* New York: Springer-Verlag.

Brams, S.J., & Hessel, M. (1982). Absorbing outcomes in 2 × 2 games. *Behavioral Science, 27,* 393–401.

Brams, S.J., & Hessel, M. (1983). Staying power in sequential games. *Theory and Decision, 15,* 279–302.

Brams, S.J., & Hessel, M. (1984). Threat power in sequential games. *International Studies Quarterly, 28,* 23–44.

Brams, S.J., & Wittman, D. (1981). Nonmyopic equilibria in 2 × 2 games. *Conflict Management and Peace Science, 6,* 39–62.

Colman, A. (1982). *Game theory and experimental games: The study of strategic interaction.* Oxford: Pergamon Press.

Davis, M. (1983). *Game theory: A nontechnical introduction* (rev. ed.). New York: Basic Books.

Fraser, N.M., & Hipel, K.W. (1979). Solving complex conflicts. *IEEE Transactions on Systems, Man, and Cybernetics* (SCM-9) *12*, 805–816.

Fraser, N.M., & Hipel, K.W. (1982). Dynamic modeling of the Cuban missile crisis. *Conflict Management and Peace Science, 6*, 1–18.

Harris, R.J. (1969). Note on "Optimal policies for the Prisoner's Dilemma." *Psychological Review, 76*, 363–375.

Howard, N. (1971). *Paradoxes of rationality: Theory of metagames and political behavior.* Cambridge, MA: MIT Press.

Kilgour, D.M. (1984). Equilibria for far-sighted players. *Theory and Decision, 16*, 135–157.

Kilgour, D.M., Hipel, K. W., & Fraser, N.M. (1984). Solution concepts in non-cooperative games. *Large Scale Systems, 6*, 49–71.

Kuhn, J.R.D., Hipel, K.W., & Fraser, N. (1983). A coalition analysis algorithm with application to the Zimbabwe conflict. *IEEE Transactions on Systems, Man, and Cybernetics* (SMC-13) *3*, 338–352.

Luce, R., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey.* New York: Wiley.

Maynard Smith, J. (1982). *Evolution and the theory of games.* London: Cambridge University Press.

Maynard Smith, J., & Price, G.R. (1973). The logic of animal conflict. *Nature, 246*, 15–18.

McKelvey, R.D., Ordeshook, P., & Winer, M.D. (1978). The competitive solution for *N*-person games without transferable utility, with an application to committee games. *American Political Science Review, 72*, 599–615.

McKelvey, R.D., & Ordeshook, P.C. (1983). Some experimental results that fail to support the competitive solution. *Public Choice, 40*, 281–291.

Meleskie, M.F., Hipel, K.W., & Fraser, N.M. (1982). The Watergate tapes conflict: A metagame analysis. *Political Methodology, 8*, 1–23.

Nash, J. (1951). Non-cooperative Games. *Annals of Mathematics, 54*, 286–295.

Ordeshook, P.C. (1980). Political disequilibrium and scientific inquiry: A comment on William Riker's "Implications from the disequilibrium of majority rule for the study of institutions." *American Political Science Review, 74*, 447–450.

Rapoport, A. (1967). Optimal policies for the Prisoner's Dilemma. *Psychological Review, 74*, 136–148.

Rapoport, A., & Guyer, M. (1966). A taxonomy of 2 × 2 games. *General Systems: Yearbook of the Society for General Systems Research, 11*, 203–214.

Rice, P. (1979). The finite negotiation problem: A solution theory. *Journal of Conflict Resolution, 23*, 561–576.

Riker, W.H. & Ordeshook, P.C. (1973). *An introduction to positive political theory.* Englewood Cliffs, NJ: Prentice-Hall.

Savich, P., Hipel, K.W., & Fraser, N.M. (1983). The Alaskan gas pipeline conflict. *Energy, 8*, 213–224.

Schotter, A., & Schwödiauer, G. (1980). Economics and the theory of games: A survey. *Journal of Economic Literature, 18*, 479–527.

Shubik, M. (1970). Game theory, behavior and the paradox of the Prisoner's Dilemma: Three solutions. *Journal of Conflict Resolution, 14*, 181–193.

Shubik, M. (1982). *Game theory in the social sciences: Concepts and solutions.* Cambridge, MA: MIT Press.

Shupe, M.C., Wright, W., Hipel, K., & Fraser, N. (1980). Nationalization of the Suez Canal. *Journal of Conflict Resolution, 24*, 477–493.

Takahashi, M.A., Fraser, N.M., & Hipel, K.W. (in press). A procedure for analyzing hypergames. *The European Journal of Operations Research.*

Taylor, M. (1976). *Anarchy and cooperation.* London: Wiley.

van den Doel, H. (1979). *Democracy and welfare economics*. Cambridge: Cambridge University Press.

von Neumann, J. (1928). Zur theorie der gesellschaftsspiele. *Mathematische Annalen, 100,* 295–320.

von Neumann, J., & Morgenstern, O. (1953). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Wright, W.M., Shupe, M., Fraser, N., & Hipel, K. (1980). A conflict analysis of the Suez Canal invasion of 1956. *Conflict Management and Peace Science, 5,* 27–40.

Zagare, F.C. (1981). Nonmyopic equilibria and the Middle East crisis of 1967. *Conflict Management and Peace Science, 5,* 139–162.

Zagare, F.C. (1983). A game-theoretic evaluation of the cease-fire alert decision of 1973. *Journal of Peace Research, 20,* 73–86.

Zagare, F.C. (1984a). *Game theory: The essential ideas with social science applications*. Beverly Hills, CA: Sage Publications.

Zagare, F.C. (1984b). Limited move equilibria in 2 × 2 games. *Theory and Decision, 16,* 1–19.

Zagare, F.C. (1985). Toward a reformulation of the theory of mutual deterrence. *International Studies Quarterly, 29,* 155–169.

Zagare, F.C. (in press). The dynamics of deterrence. Chicago, IL: University of Chicago Press.