International Studies Perspectives (2000) **1**, 221–232.

VISIONS OF INTERNATIONAL STUDIES

Game Theory: Pitfalls and Opportunities in Applying It to International Relations

Steven J. Brams
New York University

Four problems plague game-theoretic models in international relations (IR): (1) misspecifying the rules, (2) confusing goals and rational choice, (3) arbitrarily reducing the multiplicity of equilibria, and (4) forsaking backward induction. An alternative approach, *theory of moves* (TOM), is discussed and applied to Prisoners' Dilemma and then, more prescriptively, to the Iran hostage crisis of 1979–80.

TOM incorporates into the framework of game theory an initial state in a payoff matrix, the moves and countermoves required to reach a "nonmyopic equilibrium," and threat, moving, and order power that reflect asymmetries in the capabilities of the players. It also allows for incomplete information, which in the Iran hostage crisis led to misperceptions and flawed play.

Two general lessons come out of the U.S. foreign-policy failure in the Iran hostage crisis: (1) know the game you are playing, and (2) make threats only if they are likely to be credible. In specific games, TOM provides detailed prescriptions for optimal play, depending on where play starts and the powers of the players, that could aid foreign-policy makers, especially in crises.

Keywords: game theory, theory of moves, Iran hostage crisis

In the last twenty or so years, there has been a surge of interest in modeling both national security and international political economy issues using the tools of game theory. But this greater attention to game theory and modeling has also come with growing questions about the policy relevance of the approach and the ease with which such concepts can be taught readily in the college classroom. The issue of policy relevance has most recently been broached in the popular press with articles in the *New Republic* (1999) and the *New York Times* (2000) that received widespread attention in the scholarly community. In the classroom, there is also a lack of guidance for finding ways of teaching these concepts and tools in an accessible and relevant manner to today's "hands-on" student.

In this brief article I attempt to address both of these issues by (1) outlining four major theoretical problems that have bedeviled various attempts at game-theoretic modeling of international relations (IR); and (2) proposing an alterna-

Author's note: I gratefully acknowledge the support of the C. V. Starr Center for Applied Economics at New York University.

Published by Blackwell Publishers, 350 Main Street, Malden, MA 02148, USA, and 108 Cowley Road, Oxford OX4 1JF, UK.

tive approach, grounded in game theory, which I and others have found attractive in (i) capturing the thinking of decision makers and (ii) realistically modeling their strategic choices in a parsimonious but historically informed way. I will then illustrate how this approach can be applied, prescriptively, to foreign-policy decision making by examining the failure of the United States to satisfy its goals in the Iran hostage crisis of 1979–80. This failure, I argue, might have been diminished, if not avoided, had the alternative modeling approach I recommend been used and certain "lessons" learned.

Modeling Problems

I start by outlining four major pitfalls that plague game-theoretic modeling of IR phenomena and processes:

1. Misspecifying the rules. A game is "the totality of rules that describe it" (von Neumann and Morgenstern, 1953: 49). To be sure, what constitutes rules is not a simple matter. Generally speaking, rules both prescribe and proscribe behavior. Any game-theoretic models, therefore, should propose rules of play that reflect how players think and act in the strategic situation being modeled.

Based on this reality check, standard game theory often misses the mark. Players do not usually

- choose strategies simultaneously or independently of each other, as assumed in the normal or strategic form of a game that can be represented by a payoff matrix;
- adhere to a specified sequence of choices, as assumed in the extensive form of a game that can be represented by a game tree.

Rather, play usually starts at some initial state, or status-quo point. Players look ahead to ascertain whether they can do better by moving or staying in this state. Their choices will vary, depending on the initial state and the rules that indicate what kinds of moves are possible and the means that best satisfy their goals.

Generally accepted rationality postulates, applied to the normal or extensive game forms, are frequently violated. For example, sometimes players choose dominant strategies in Prisoners' Dilemma, or they do not use backward induction, starting from the endpoints of a game tree, when Prisoners' Dilemma is repeated and the number of rounds is known. These violations, I suggest, are remediable, but they require different rules of play to render them rational and intelligible, as I will illustrate for Prisoners' Dilemma later.

2. Confusing goals with rational choice. Most people have goals, whether implicit or explicit, and they choose the best means to satisfy them, subject to environmental constraints like incomplete information. Emphatically, these constraints do not impair their rationality. Thus, if a person's goals are short-term, then he or she will be rational with respect to them; if long-term, this person's behavior may be quite different.

Some theorists have suggested that short-term goals mirror a thin kind of rationality and long-term goals a thicker or deeper kind. But, in fact, the observable differences in the behavior of the thin and thick types usually arise from their different goals, not their rationality or lack thereof. Put another way, thin types are not being irrational but instead choosing (rationally) *not* to look deeply or far ahead, presumably because this is too costly.

Rationality is a concept appropriately applied to the efficiency or efficacy of the means, or instruments used, to attain desired ends. What are the costs and benefits of different means, and are people making efficacious choices to achieve their ends? Rationality does *not* concern the ends themselves, which are neither rational nor irrational. To be sure, it is an important developmental question how people come to harbor the goals that they do, but that question is not pertinent to any instrumental notion of rationality that game theory postulates.

3. Arbitrarily reducing the multiplicity of equilibria. Many games have multiple Nash equilibria; the 2×2 game of Chicken, for example, has two equilibria in pure strategies and one in mixed strategies. Worse, in repeated play, almost any outcome of any game can be supported as a Nash equilibrium under certain conditions. With this embarrassment of riches, a game-theoretic model may not be helpful in narrowing down a plethora of stable outcomes to a few precise predictions.

Some theorists argue that norms like reciprocity, when they take hold in a population, can induce cooperative outcomes in collective-action situations, thereby helping actors avoid inefficient equilibria. Indeed, there is a good deal of empirical evidence to support not only the reciprocity norm but also the build-up of trust, and the nurturing of reputations of honesty and probity, to stabilize cooperation in games.

Unfortunately, these explanations for cooperation are often *ad hoc:* they do not follow from tenets of rationality in the games being studied. Rather, reciprocity, reputation, and trust are imported as behavioral assumptions, in part because these characteristics of game play are actually observed. Institutions may also play a role in fostering cooperation, but they are not as robust in IR as in domestic politics.

An important theoretical direction of game theory has been to endogenize some of its more restrictive assumptions by coarsening information structures, relaxing common knowledge of rationality, offering refinements of Nash equilibria, and the like. In effect, the definition of a game is altered by changing what players know, how they apprehend others' calculations, and the nature of the outcomes they seek.

For example, there are now several models that demonstrate how cooperation can emerge from repeated play of Prisoners' Dilemma when information is incomplete. But repeated play is very much an artifact when modeling player choices over time. Do players really play a single stage game again and again, with no variation in its strategies, outcomes, or payoffs, accumulating payoffs at each stage?

In fact, strategies are often in flux, and possible outcomes are constantly changing, undermining the assumption of repeated play of a single stage game. Also problematic in the modeling of IR games, such as protracted wars, is that payoffs may not be realized in stages but only after considerable time has elapsed.

While evolutionary game theory and ecological models offer more sophisticated dynamics, they are complex, even if they narrow down the equilibria to a relatively few. In addition, few IR processes seem plausibly modeled by a dynamic involving the random matching of players—with possible mutations—as assumed in biological-evolutionary models.

4. Forsaking backward induction. On the one hand, backward induction is "the oldest idea in game theory" that has "maintained its centrality to this day" (Aumann, 1995: 6). On the other hand, real players do not painstakingly work backward from the endpoints of a game tree, with many levels and a multitude of branches, to try to determine what they should do at the start of play.

Rather, it seems, players develop heuristics and use rules of thumb that simplify this process, not because they are short-term thinkers and have abandoned their intelligence but because they have found good approximations to backward induction. But whence these rules, if not from robust and time-tested simplifications that approximate the calculations of backward induction?[1]

Thus, we should not dismiss backward induction too quickly. We use simple contingent strategies, like tit-for-tat, because they work well most of the time, even if they are not optimal in every situation. As I shall argue next, players make more exacting calculations in specific games, looking two, three, or four moves ahead, especially when the outcome is important to them.

TOM: An Alternative Approach

What I call the "theory of moves" (TOM) embeds extensive-form games within the normal form, deriving advantages of both forms: the nonmyopic thinking of the extensive form disciplined by the economy of the normal form (Brams, 1993, 1994). In my view, the rules of TOM, which allow players to move and countermove within a payoff matrix, capture the changing strategic nature of situations that evolve over time. TOM's ability, as well, to incorporate differences in power—moving, order, or threat—reflects an asymmetry of player capabilities in many IR games.

Because Prisoners' Dilemma is a central focus of IR, let me illustrate its nonmyopic perspective by showing what constitutes rational play, according to TOM, in the simple 2×2 version of Prisoners' Dilemma:

- If the play starts at the *noncooperative outcome*, players are stuck, no matter how far ahead they look, because as soon as one player departs, the other player, enjoying its best outcome, will not move on. **Outcome:** The players stay at the noncooperative outcome, which makes it a "nonmyopic equilibrium."
- If play starts at the *cooperative outcome*, then neither player will defect, because if it does, the other player will also defect, and they both will end up worse off. Thinking ahead, therefore, neither player will defect. **Outcome:** The players stay at the cooperative outcome, which makes it a nonmyopic equilibrium.
- If play starts at one of the *win-lose outcomes* (best for one player, worst for the other), the player doing best will know that if it is not magnanimous, and consequently does not move to the cooperative outcome, its opponent will move to the noncooperative outcome, inflicting on the best-off player its next-worst outcome. Therefore, it is in the best-off player's interest, as well as its opponent's, that the former act magnanimously, anticipating that if it does not, the noncooperative outcome (next-to-worst for both), rather than the cooperative outcome (next-to-best for both), will be chosen. **Outcome:** The best-off player will move to the cooperative outcome, where play will stop.

These rational moves, grounded in backward induction, are not beyond the pale of most players. They are frequently made by those who look beyond the immediate consequences of their own choices.[2] Such far-sighted players *can* escape the dilemma in Prisoners' Dilemma—as well as poor outcomes in other

[1] Backward induction itself is not always robust; it may be sensitive to the parity of the number of rounds of play (i.e., whether this number is even or odd) and the boundedness of the game (i.e., whether it will end after some finite number of rounds) (Brams and Kilgour, 1998).

[2] As a case in point, consider the following statement by Theodore C. Sorensen about the deliberations of the Executive Committee (Excom) during the October 1962 Cuban missile crisis (Holsti, Brody, and North, 1964: 188): "We discussed what the Soviet reaction would be to any possible move by the United States, what our reaction with them would have to be to that Soviet reaction, and so on, trying to follow each of those roads to their ultimate conclusion."

games—provided play begins at a state other than the noncooperative one.[3] Hence, TOM does not predict unconditional cooperation in this game but, instead, makes it a function of the starting point of play.[4]

TOM's history-dependent approach has been applied to the 56 other 2×2 strict ordinal *conflict games*—in which players can rank outcomes, but there is no outcome that is mutually best—and its predictions empirically tested in several IR studies.[5] Not only have its predictions been generally supported, but it has also explained well the dynamics of these conflicts.

I turn next to a real-life application of TOM that gives predictions different from classical game theory. Prescriptively, it suggests two lessons in applying TOM to foreign-policy making, both of which are, in a sense, obvious. Less obvious is what TOM says, specifically, about the opportunities the United States had to influence (or not influence) the situation and, in particular, how it might have prevented a major foreign-policy failure.

The Iran Hostage Crisis

When Iranian students, with Iranian government support, seized the American embassy and took embassy officials hostage in November 1979, the military capabilities of the two sides were almost irrelevant.[6] Although an attempt was made in April 1980 to rescue the hostages in an aborted U.S. military operation that cost eight American lives, the conflict was never really a military one. It can best be represented as a game in which President Jimmy Carter misperceived the preferences of Ayatollah Ruholla Khomeini and attempted, in desperation, to find a solution in the wrong game, suggesting

Lesson 1. Know what game you are playing.

Why did Khomeini sanction the takeover of the American embassy by militant students? Doing so had two advantages. First, by creating a confrontation with the United States, Khomeini was able progressively to sever the many links that remained with this "Great Satan" from the days of the shah. Second, the takeover mobilized support for extremist revolutionary objectives just at the moment when moderate secular elements in Iran were challenging the principles of the theocratic state that Khomeini had installed.

President Carter's primary goal was immediate release of the hostages. His secondary goal was to hold discussions with Iranian religious authorities on resolving the differences that had severely strained U.S.–Iranian relations. Of course, if the hostages were killed, the United States would likely have defended their honor, probably by a military strike against Iran.

[3] However, Willson's (1998) nonmyopic rules, which allow for backtracking, provide an avenue of escape by rendering even a sequence of moves that carry the players from the noncooperative to the cooperative state in Prisoners' Dilemma rational. Alternatively, I show how credible threats can induce cooperation in Prisoners' Dilemma and related games (Brams, 1994: ch. 5).

[4] Under the TOM rules of play that allow players to move and countermove within the payoff matrix, Prisoners' Dilemma is, of course, no longer the classical version of this game. Because there is almost always some status quo in most IR games, however, I believe it more realistic to start play at states rather than have players commence by making simultaneous or independent strategy choices. Contrast the cooperation that emerges from this framework, based on backward induction, with that which emerges from a framework that postulates *ad hoc* behavioral assumptions, like reciprocity.

[5] Brams (1994) includes analyses of the Cuban missile crisis in 1962 and the U.S. bombing of Vietnam in 1972. Other IR applications of TOM can be found in worked by Brams (1997, 1999), Brams and Togman (1998, 2000), Maoz and Mor (1996), Massoud (1998), Mor (1993, 1997), Mor and Maoz (1999), Simon (1996), Zeager and Bascom (1996), and Zeager (1998, 1999).

[6] This case is adapted from Brams and Mattli, 1993, which was later used in Brams, 1993 and 1994. Here it is developed to show, prescriptively, how TOM might be applied to a past crisis in order to draw lessons from it relevant to better crisis decision making in future crises.

Carter considered two strategies:

1. **Negotiate (N).** With diplomatic relations broken after the seizure, negotiations could be pursued through the U.N. Security Council, the World Court, or informal diplomatic channels; the negotiations might include the use of economic sanctions.
2. **Intervene militarily (I).** Military action could include a rescue mission to extract the hostages or punitive strikes against selected targets (e.g., refineries, rail facilities, or power stations).

Khomeini also had two strategies:

1. **Negotiate (N).** Negotiations would involve demanding a return of the shah's assets and an end to U.S. interference in Iran's affairs.
2. **Obstruct (O).** Obstructing a resolution of the crisis could be combined with feigning to negotiate.

The two players and their two strategies generate a 2×2 payoff matrix. Each cell in the payoff matrix has an associated payoff for each player. I assume that Carter and Khomeini can rank the four resulting outcomes from best (4) to worst (1).

Carter's rankings are shown as the first numbers, and Khomeini's the second numbers, in each ordered pair of the two payoff matrices in Figure 1. I begin with game 50, the "game as misperceived by Carter."[7] As shown, Carter most preferred that Khomeini choose N (4 and 3) rather than O (2 and l), but in any case he preferred N to O, given the difficulties of military intervention.

These difficulties were compounded in December 1979 by the Soviet invasion of Afghanistan, which eliminated the Soviet Union as a possible ally in seeking concerted action for release of the hostages through the United Nations. With Soviet troops next door in Afghanistan, the strategic environment was anything but favorable for military intervention.

As for Khomeini, Carter *thought* that he faced serious problems within Iran because of a critical lack of qualified people, demonstrations by the unemployed, internal war with the Kurds, Iraqi incursions across Iran's western border, and a continuing power struggle at the top (though his own authority remained unchallenged). Consequently, Carter believed that negotiations would give Khomeini a dignified way out of the impasse (Carter, 1982: 459–489).

One implication of this view is that while Carter thought that Khomeini most preferred a U.S. surrender at NO (4), he believed Khomeini would next most prefer the compromise of NN (3). Thus, Khomeini's two worst states (1 and 2), in Carter's view, were associated with the U.S.'s strategy of I.

Carter's imputation of these preferences to Khomeini turned out to be a major misperception of the strategic situation. Khomeini wanted the total Islamization of Iranian society; the United States was "a global Shah—a personification of evil" (quoted in Saunders, 1985: 102) that had to be cut off from any contact with Iran. Khomeini abjured his nation never to "compromise with any power [and] to topple from the position of power anyone in any position who is inclined to compromise with the East and West" (Sick, 1985a: 237).

If Iran's leaders should negotiate the release of the hostages, this would weaken their uncompromising position. Those who tried, including President Bani-Sadr

[7] Game 50 is one of the 57 2×2 conflict games, mentioned earlier, all of which are listed in the Appendix of Brams, 1994. My treatment of TOM here will be quite informal, but unlike the analysis of Prisoners' Dilemma in the previous section, I will begin by justifying two games as models of the empirical situation under study before turning to their analysis.

Game as Misperceived by Carter (Game 50)

		Khomeini			
		Negotiate (N)		Obstruct (O)	
	Negotiate (N)	Compromise (4,3)	→	Carter surrenders (2,4)	← Dominant strategy
Carter		↑		↓	
	Intervene militarily (I)	Khomeini surrenders (3,2)	←	Disaster (1,1)	

Real Game (Game 5)

		Khomeini			
		Negotiate (N)		Obstruct (O)	
	Negotiate (N)	Carter succeeds (4,2)	→	Khomeini succeeds (2,4)	← Dominant strategy
Carter		↑		↓	
	Intervene militarily (I)	Carter adamant (3,1)	←	Khomeini adamant (1,3)	
				↑ Dominant strategy	

Key: (x, y) = (payoff to Carter, payoff to Khomeini)

4 = best; 3 = next best; 2 = next worst; 1 = worst

Nash equilibria underscored

Arrows within matrices indicate the direction of cycling

FIG. 1. Iran Hostage Crisis (Games 50 and 5)

and Foreign Minister Ghotbzadeh, lost in the power struggle. Bani-Sadr was forced to flee for his life to Paris, and Ghotbzadeh was arrested and later executed.

What Carter was unable to grasp was that Khomeini most preferred O (4 and 3), independent of what the United States did. Doubtless, Khomeini also preferred that the United States choose N, whatever his own strategy choice was, giving him the preferences shown in the second payoff matrix of Figure 1, which is game 5 and what I call the "real game."

Rational Play in the Two Games

What does classical game theory say about the rational choices of the players in the misperceived game (game 50) and the real game (game 5)? In both games, Carter's *dominant*, or unconditionally better, strategy is N. Regardless of what Khomeini chooses, Carter's payoff from N is better than his payoff from I. In the misperceived game, for example, if Khomeini chooses N, Carter receives a payoff of 4 by choosing N and a payoff of 3 by choosing I; if Khomeini chooses O, Carter receives a payoff of 2 by choosing N and a payoff of 1 by choosing I.

Although Carter's dominant strategy of N in both games is independent of Khomeini's choice, Khomeini's best choice in the misperceived game depends on what Carter selects. If Carter chooses N, which he should because it is dominant, Khomeini, anticipating this, does better by choosing O, which gives him a payoff of 4, rather than choosing N, which gives him a payoff of 3. So in the misperceived game, Khomeini should choose O, leading to the NO outcome. That outcome, which I call "Carter surrenders," gives Carter a payoff of 2 and Khomeini a payoff of 4.

Game theory calls NO a *Nash equilibrium*, because neither player would have an incentive unilaterally to depart from it: Carter would go from 2 to 1 if he switched from N to I, and Khomeini would go from 4 to 3 if he switched from O to N. Likewise in the real game, NO is a Nash equilibrium, because both players have dominant strategies associated with it and so would do worse by deviating from them.

In the real game, I call the NO outcome "Khomeini succeeds." (The other three outcomes in the real game are ranked differently by Khomeini from those in the misperceived game, which is why I give them different shorthand descriptions.) In the real game, the rationality of the (2,4) outcome is reinforced by Khomeini's dominant strategy of O associated with it; O is not dominant in the misperceived game but, instead, Khomeini's best response if Carter chooses his own dominant strategy of N.

Given that Carter always does better in both games by choosing N, why would he consider, much less try, to choose I? Classical game theory does not give a reason, but TOM suggests the basis for his miscalculation. Carter might have thought—with some justification in the misperceived game—that by threatening Khomeini with I, he could induce him to choose N, giving Carter the opportunity, by choosing N himself, to obtain his best payoff.

The reasoning underlying this calculation goes as follows. In the misperceived game, the NN outcome gives Carter his best payoff of 4 and Khomeini his next-best payoff of 3. A threat by Carter to choose I, if carried out, would inflict upon Khomeini his two worst outcomes in the misperceived game: a payoff of 2 if he chose N and a payoff of 1 if he chose O. Since Khomeini would prefer a payoff of 2 over 1, he would choose N, given Carter's threat was credible. But because both players do better by choosing "compromise" at (4,3) rather than "Khomeini surrenders" at (3,2), Khomeini should choose N when Carter does, assuming that he takes seriously Carter's threat of I.

There are two problems, however, with this reasoning. First, it is not clear that Carter had what TOM calls "threat power"—specifically in game 50, the ability to induce the compromise outcome in this misperceived game—suggesting

Lesson 2. Make threats only if you have threat power and your threats are, therefore, perceived as credible by your opponent.

To be sure, the game as misperceived by Carter was not the real game being played (Lesson 1). In the real game, Khomeini had no reason to accede to threats by Carter, because his political position was stronger choosing O, regardless of Carter's choice.

Nonetheless, Carter tried threats. He dispatched the aircraft carrier USS *Kitty Hawk* and its supporting battle group from the Pacific to the Arabian Sea. The carrier USS *Midway* and its battle group were already present in the area. Sick (1985b: 147) reported:

> With the arrival of Kitty Hawk, the United States had at its disposal the largest naval force to be assembled in the Indian Ocean since at least World War II and the most impressive array of firepower ever deployed to those waters.

But this threat, like those preceding it, did not lead to any change in Khomeini's strategy because of Carter's fateful underestimation of Khomeini's willingness and ability to absorb economic, political, and military punishment in the pursuit of his revolutionary goals. In the real game, I leads to the (1,3) outcome when Khomeini chooses obstruct O. Because of the possible execution of the hostages that an attack might provoke—the threat of which was "taken with deadly seriousness in Washington" (Sick, 1985b: 147)—I rank it as worst for the United States.

After negotiations faltered and then collapsed in April 1980, Carter was forced to move to his I strategy. Naturally, if the rescue operation had succeeded and the hostages had been freed, the game would have been in state (3,1) because Khomeini could in that situation no longer use the hostages as a weapon and choose O.

The rescue's failure kept the situation in state (2,4) for another nine months. But the Iranian leadership had already concluded in August 1980, after the installation of an Islamic government consistent with Khomeini's theocratic vision, that the continued retention of the hostages was a net liability (Saunders, 1985: 44–45). Further complicating Iran's position was the attack by Iraqi forces in September 1980. It was surely no accident that on the day of Carter's departure from the White House on January 20, 1981—444 days after the capture of the hostages—they were set free.[8]

Although Carter's strategic acumen in this crisis can be questioned, it was less his rationality that was at fault than his misperception of Khomeini's preferences. Within a week of the embassy seizure, analysts in the State Department had reached the conclusion that

> diplomatic action had almost no prospect of being successful in liberating the hostages and that no economic or other U.S. pressure on the Iranian regime, including military action, was likely to be any more successful in securing their safe release. Consequently, they concluded, the detention of the hostages could continue for some months. (Sick, 1985a: 246)

In the first few months of the crisis, U.S. Secretary of State Cyrus Vance counseled that "we continue to exercise restraint" (Vance, 1993: 408). Privately, he was vehemently opposed to any military action; and after the military rescue operation failed, he resigned.

But others voiced different views, including secular politicians in Iran who claimed to speak for Khomeini. There was an abundance, rather than a dearth, of information, but the question, as always, was what was accurate.

[8] That the hostages were not released before the November 1980 presidential election, which clearly would have benefited Carter's bid for reelection, Sick (1991) attributes to a secret deal Iran made with Reagan supporters. But this allegation, at least as far as George Bush's involvement is concerned, is disputed by a bipartisan October Surprise Task Force of the U.S. House of Representatives chaired by Lee H. Hamilton (Lewis, 1992, 1993). The arguments for a secret deal (Sick) and against (Hamilton) are joined in "Last Word on the October Surprise" (1993).

Carter, perhaps, should not be judged too harshly for misjudging the situation. In fact, even if he had foreseen the real game from the start (Lesson 1), this analysis suggests that there was little that he could have done to move the outcome away from (2,4), given that the military power of the United States could not readily be translated into a credible threat (Lesson 2).

Evidently, it was Carter's misperception that gave him the hope that he could implement the compromise outcome of (3,4) in the misperceived game. Not only can it be induced by threat power, but it can also be induced by "moving power," which is another kind of power, mirroring the stamina or endurance of a player, that TOM admits.

To illustrate how it works, assume that the players move and countermove in a clockwise direction in the misperceived game, as shown by the arrows in Figure 1. In that direction, neither player ever moves from its best outcome (Carter vertically or Khomeini horizontally). In a counterclockwise direction, by contrast, players do move from their best outcomes: Khomeini moves from a payoff of 4 at (2,4) when he switches from O to N, and Carter moves from a payoff of 4 at (4,3) when he switches from N to I. So if there is cycling, it must be in a clockwise direction.

If Carter believed he had moving power—the ability to force Khomeini to stop in the move-countermove process—he could force Khomeini to stop at either the NN outcome or the IO outcome, which are the two outcomes where Khomeini has the next move. Khomeini would prefer the former, which gives him a payoff of 3, to the latter, which gives him a payoff of 1. In the real game, however, these outcomes give Khomeini payoffs of 2 and 3, respectively, so he would choose to stay at the IO outcome. As a consequence, Carter's hoped-for NN outcome in the misperceived game became, in April 1980, an IO outcome in the real game.

Finally, it is worth noting that (3,4) is a nonmyopic equilbrium, just as the cooperative outcome in Prisoners' Dilemma is: starting from any other outcome in the misperceived game, it is rational for the players to move and countermove to (3,4), at least if Carter has what TOM calls "order power," or the ability to determine the order of play in making moves from certain outcomes. But he almost surely did not have this power in the hostage crisis, despite the overwhelming military strength of the United States, which simply was not an effective weapon in extracting the hostages.

This contrast between the two games is obscured by the classical theory, which shows (2,4), in which Carter surrenders, to be the unique Nash equilibrium, associated with Carter's dominant strategy of N, in each game. Thus, the classical theory makes Carter's actions inexplicable in terms of rational choice—whether he misperceived Khomeini's preferences or not—whereas TOM shows that Carter's actions were not ill-founded, given his misperception of both the game being played and his ability to influence the choice of outcomes in it.

Conclusions

TOM incorporates into the framework of game theory an initial state in a payoff matrix, the moves and countermoves required to reach a nonmyopic equilibrium, along with *threat*, *moving*, and *order power* that reflect asymmetries in the capabilities of the players. It also allows for incomplete information, which in the Iran hostage crisis led to misperceptions and flawed play.

Because TOM postulates that players rank outcomes but do not attach cardinal utilities to them, it seems eminently applicable to strategic situations in which it is hard to estimate players' precise valuations of the outcomes. In addition, it overcomes some problems of classical game theory by providing realistic rules for dynamic play, restricting nonmyopic equilibria to those that can be reached from where play commences, and using backward induction that enables players to make far-sighted calculations.

Two general lessons come out of U.S. foreign-policy failures in the Iran hostage crisis: (1) know the game you are playing, and (2) make threats only if they are likely to be credible. In specific games, TOM provides detailed prescriptions for optimal play, depending on where play starts and the powers of the players, that could aid foreign-policy decision makers, especially in crises.

But TOM is no panacea. It has been mostly developed for two-person games, whereas strategic situations in IR, including alliances and collective-action situations more generally, are often best modeled as *n*-person games.[9]

While TOM can be extended to more complex games, more general theoretical structures need to be built, based on plausible rules of play. In my view, these need to take into account (1) players' strategic choices of coalition partners, and (2) the courses of actions that flow from these choices (Brams, Jones, and Kilgour, 2000). Power differences among those coalitions that form, and the history of play that delimits equilibria—by ruling out later choices that are precluded by earlier ones—should provide foundations for a more general dynamic theory.

References

Aumann, R. J. (1995) Backward Induction and Common Knowledge of Rationality. *Games and Economic Behavior* **8**(1):6–19.

Brams, S. J. (1993) Theory of Moves. *American Scientist* **81**(6):562–570.

Brams, S. J. (1994) *Theory of Moves*. Cambridge: Cambridge University Press.

Brams, S. J. (1997) "The Rationality of Surprise: Unstable Nash Equilibria and the Theory of Moves." In *Decisionmaking on War and Peace: The Cognitive-Rational Debate*, edited by A. Mintz and N. Geva, pp. 103–129. Boulder, CO: Lynne Rienner. Updated version, with a response and rebuttal, in *Mathematica Japonica* **49**(2)(1999):293–319.

Brams, S. J. (1999) To Mobilize or Not to Mobilize: Catch-22s in International Crises. *International Studies Quarterly* **43**(4):621–640.

Brams, S. J., M. A. Jones, and D. M. Kilgour (2000) "The Paradox of Disconnected Coalitions." Preprint, Department of Politics, New York University.

Brams, S. J., and D. M. Kilgour (1998) Backward Induction Is Not Robust: The Parity Problem and the Uncertainty Problem. *Theory and Decision* **45**(3):263–289.

Brams, S. J., and D. M. Kilgour (2001) Fallback Bargaining. *Group Decision and Negotiation*. Forthcoming.

Brams, S. J., and W. Mattli (1993) Theory of Moves: Overview and Examples. *Conflict Management and Peace Science* **12**(2):1–39.

Brams, S. J., and J. M. Togman (1998) Cooperation Through Threats: The Northern Ireland Case. *PS: Political Science & Politics* **30**(1):32–39.

Brams, S. J., and J. M. Togman (2000) "Agreement Through Threats: The Northern Ireland Case." In *Being Useful: Policy Relevance and International Relations Theory*, edited by M. Nincic and J. Lepgold. Ann Arbor: University of Michigan Press.

Carter, J. (1982) *Keeping Faith: Memoirs of a President*. New York: Bantam.

Holsti, O. R., R. A. Brody, and R. C. North (1964) Measuring Affect and Action in International Reaction Models: Empirical Materials from the 1962 Cuban Missile Crisis. *Journal of Peace Research* **1**:170–189.

"Last Word on the October Surprise" (1993) *New York Times*, January 24, p. E17.

Lewis, N. A. (1992) "Panel Rejects Theory Bush Met Iranians in Paris in 80." *New York Times*, July 2, p. A16.

Lewis, N. A. (1993) "House Inquiry Finds No Evidence of Deal on Hostages in 1980." *New York Times*, January 13, pp. A1, A19.

Maoz, Z., and B. D. Mor (1996) Enduring Rivalries: The Early Years. *International Political Science Review* **17**(2):141–160.

[9] Brams and Kilgour (2001) and Willson (1998, 2000), among others, have offered extensions to *n*-person games.

Massoud, T. G. (1998) "Theory of Moves and the Persian Gulf War." In *The Political Economy of War and Peace*, edited by M. Wolfson, pp. 247–266. Amsterdam: Kluwer Academic.
Mor, B. D. (1993) *Decision and Interaction in Crisis: A Model of International Crisis Behavior*. Westport, CT: Praeger.
Mor, B. D. (1997) The Middle East Peace Process and Regional Security. *Journal of Strategic Studies* **20**(1):172–202.
Mor, B. D., and Z. Maoz (1999) Learning and the Evolution of Enduring International Rivalries: A Strategic Approach. *Conflict Management and Peace Science* **17**(1):1–48.
New Republic (1999) www.tnr.com/archive/1099/102599/coverstory 102599
New York Times (2000) February 26, Arts and Ideas section, A15, A17.
Saunders, H. (1985) "The Crisis Begins." In *American Hostages in Iran: The Conduct of a Crisis*, edited by W. Christopher, pp. 35–71. New Haven, CT: Yale University Press.
Sick, G. (1985a) *All Fall Down*. New York: Penguin.
Sick, G. (1985b) "Military Options and Constraints." In *American Hostages in Iran: The Conduct of a Crisis*, edited by W. Christopher, pp. 144–172. New Haven, CT: Yale University Press.
Sick, G. (1991) *October Surprise: America's Hostages in Iran and the Election of Ronald Reagan*. New York: Random House.
Simon, M. V. (1996) When Sanctions Can Work: Economic Sanctions and the Theory of Moves. *International Interactions* **21**(3):203–228.
Vance, C. (1993) *Hard Choices: Critical Years in America's Foreign Policy*. New York: Simon and Schuster.
von Neumann, J., and O. Morgenstern (1953) *Theory of Games and Economic Behavior*, 3d ed. Princeton, NJ: Princeton University Press.
Zeager, L. A. (1998) Negotiations for Refugee Repatriation or Local Settlement: A Game-Theoretic Analysis. *International Studies Quarterly* **42**(2):367–384.
Zeager, L. A. (1999) "The Role of Threat Power in Refugee Resettlement: The Indochinese Crisis of 1979." Preprint, Department of Economics, East Carolina University.
Zeager, L. A., and J. B. Bascom (1996) Strategic Behavior in Refugee Repatriation: A Game-Theoretic Analysis. *Journal of Conflict Resolution* **40**(3):460–485.
Willson, S. J. (1998) Long-Term Behavior in the Theory of Moves. *Theory and Decision* **45**(32): 201–240.
Willson, S. J. (2000) Axioms for the Outcomes of Negotiation in Matrix Games. *Mathematical Social Sciences* **39**(3):323–348.