# American Scientist

# Theory of Moves

*Adding a dynamic dimension to game theory allows players to look ahead before making moves, thereby creating a more realistic game*

Steven J. Brams

During the Cuban missile crisis in October 1962, the Kennedy administration demanded that the Soviet Union remove its missile bases from Cuba. The Soviets acquiesced, but only after the world teetered for days between peace and disaster. Theodore C. Sorenson, special counsel to President Kennedy, later recalled, "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."

The Cuban missile crisis is a classic, albeit high-stakes, example of strategic game-playing. Like chess players, world leaders in conflict situations make carefully considered moves and countermoves. But the outcomes are not always what the players or onlookers expect; in particular, it is sometimes hard to understand why players choose conflict over cooperation.

A body of theory, called game theory, has been developed and applied over the past half-century to analyze mathematically the strategic behavior of people in situations of conflict. The theory facilitates reconstruction of past situations and modeling of possible future ones, which can explain how rational decision makers arrive at outcomes that are often puzzling at first glance.

*Steven J. Brams is professor of politics at New York University, where he has taught since 1969. He is the author or coauthor of 11 books that involve applications of game theory or social-choice theory to voting and elections, international relations, the Bible and theology.* Theory of Moves, *the book on which this article is based, will be published by Cambridge University Press in January. Address: Department of Politics, New York University, New York, NY 10003.*

Game theory approaches conflicts by asking a question as old as games themselves: How do people make "optimal" choices when these are contingent on what other people do? The seminal work was done in the 1940s by mathematician John von Neumann and economist Oskar Morgenstern, both of Princeton University, who discovered that they held similar ideas about strategies in games. They realized, first, that strategies are interdependent: Players cannot make unilaterally optimal decisions, because one player's best choice depends on the choices of other players. Von Neumann was responsible for most of their theoretical work, whereas Morgenstern pushed the applications toward economic questions.

Their collaboration led to a monumental and difficult treatise, *Theory of Games and Economic Behavior* (1944), which was revised in 1947 and then again in 1953. Over the next several decades investigators applied game theory to strategic situations ranging from the evolution of animal behavior to the rationality of believing in God.

According to the classical theory, players choose strategies, or courses of action, that determine an outcome. Von Neumann and Morgenstern called their theory "thoroughly static" because it says little about the dynamic processes by which players' choices unfold to yield an outcome.

I have developed what I call the "theory of moves" to add a dynamic dimension to classical game theory. Like the classical theory, the theory of moves focuses on interdependent strategic situations in which the outcome depends on the choices that all players make. But it radically alters the rules of play, enabling players to look ahead—sometimes several steps—before making a move.

These modifications lead to different stable outcomes, or equilibria, from those of classical game theory and new concepts of power. In this article, I shall describe informally ideas underlying the theory of moves and illustrate some of its concepts in several games—the last being one that models the Iran hostage crisis that began in 1979.

## Making Moves

Before considering the theory of moves, it is worth noting some basic elements of classical game theory. Von Neumann and Morgenstern defined a game as "the totality of rules of play which describe it," which includes a starting point and a list of legal moves that players can make. The game of tic-tac-toe, for example, begins when one player makes a mark on a three-by-three board. The rules state that a player can mark either an X or an O, but only in an unmarked block. After the first player makes a mark, the second player does, with the players then alternating in making marks on the board. The game ends when one player gets three marks in a row or all the blocks are filled.

Most games can be described in two different ways. The "extensive form" is given by a game tree, with play beginning at the first fork in the tree. One player selects one side of this fork, which moves the game to another fork. Then the other player selects a side of that fork, and so on until the game ends. This form of a game provides a full description of its sequential moves.

By contrast, the "normal form" is given by a payoff matrix, in which players choose strategies simultaneously or, if not, at least independently of each other. (A strategy gives a complete plan of possibly contingent choices—if you do this, I will do that, etc.) Thus, if a game
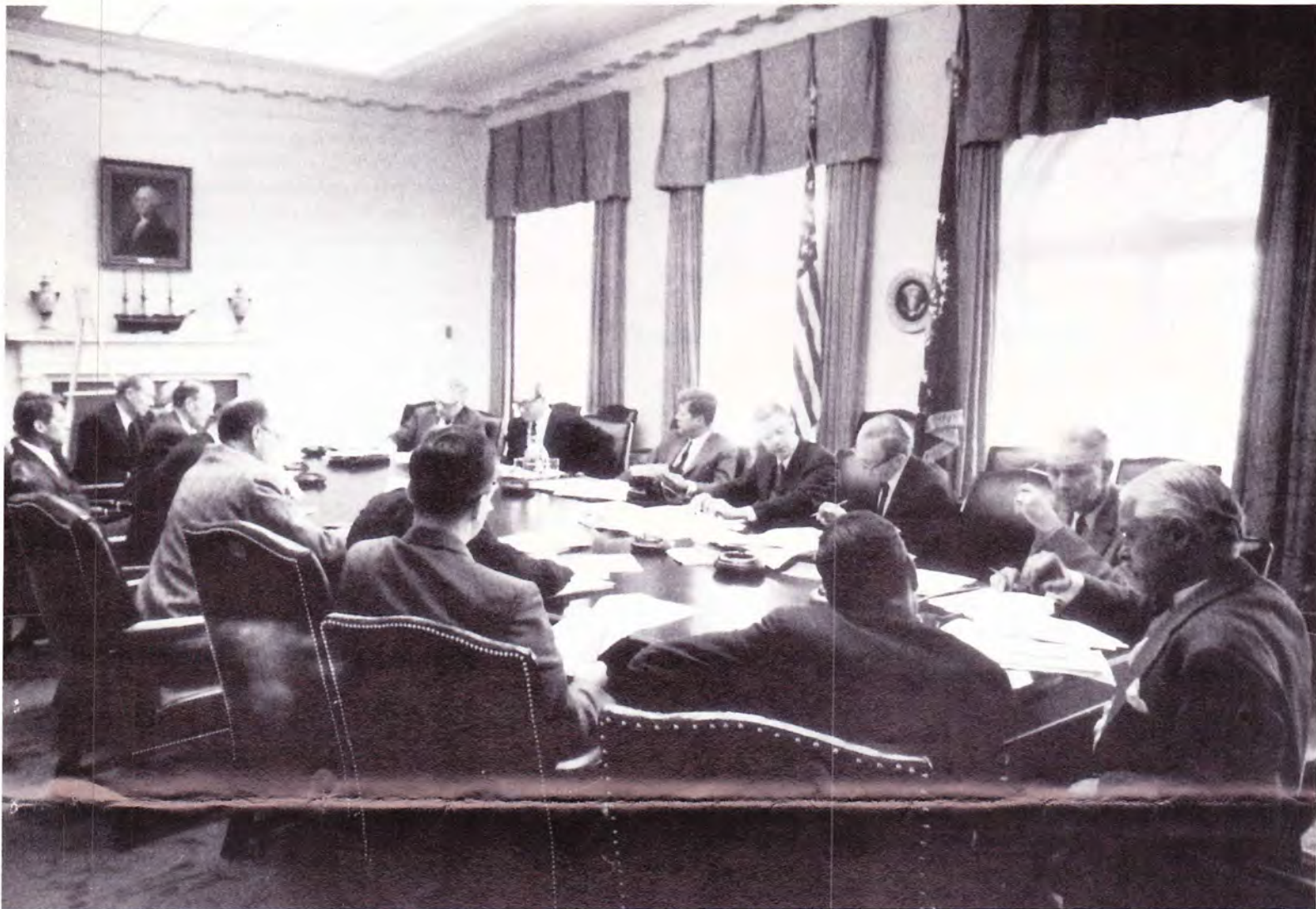
Figure 1. Game theory evaluates behavior in conflict situations. During the Cuban missile crisis, President Kennedy demanded that the Soviet Union remove its missiles from Cuba. Kennedy and his advisors, shown here in an Excom (Executive Committee) meeting, considered all their possible moves, the possible countermoves by the Soviets and the possible counter-countermoves by the United States. U.S. decision makers used their knowledge of the past and predicted future moves, which classical game theory tends to treat myopically. The author's theory of moves adds a dynamic component to classical game theory, enabling players to look ahead and select strategies that yield "nonmyopic equilibria." (Photograph courtesy of the John F. Kennedy Library.)

has two players, each with two possible strategies, it can be represented by a two-by-two payoff matrix. One player's strategy choices are given by the two rows; the other's are given by the two columns. Each row-and-column intersection defines an outcome, where payoffs are assigned to the two players.

The theory of moves combines the extensive and normal forms of classical game theory. A theory-of-moves game is played on a payoff matrix, like a normal-form game. The players, however, can move from one outcome in a payoff matrix to another, so the sequential moves of an extensive-form game are built into the more economical normal form. In large part, I shall concentrate on two-player games in which each player has two strategies; more complicated

games become quite intractable after just a few moves, although in principle the theory of moves is applicable to *n*-person games in which each of the *n* players has a finite number of strategies.

Beyond the structure of a game (normal or extensive form), one can make other modifications in the definition of what constitutes a rational choice. A rational choice depends on, among other things, how far players look ahead as they contemplate each other's possible moves and countermoves. In addition, moves are influenced by the capabilities of the players and their information about each other.

The theory of moves incorporates all these features. It is dynamic because players do not make choices *de novo*. Instead, their choices depend on the past

and present as well as the future, which players can anticipate at least in part and about which I assume they can make rational calculations.

In the theory of moves, I assume that players can rank the possible outcomes from best to worst. These payoffs, however, are only ordinal: They indicate an order of preference, but not the degree to which a player prefers one outcome over another. (Although other forms of decision-making theory indicate the degree of preference in payoffs, I have chosen ordinal payoffs to simplify the analysis and make it more applicable to real-life strategic situations.) In addition, the theory allows for power differences among players by assuming, for example, that one player may have the ability to carry out threats when necessary. Fi-
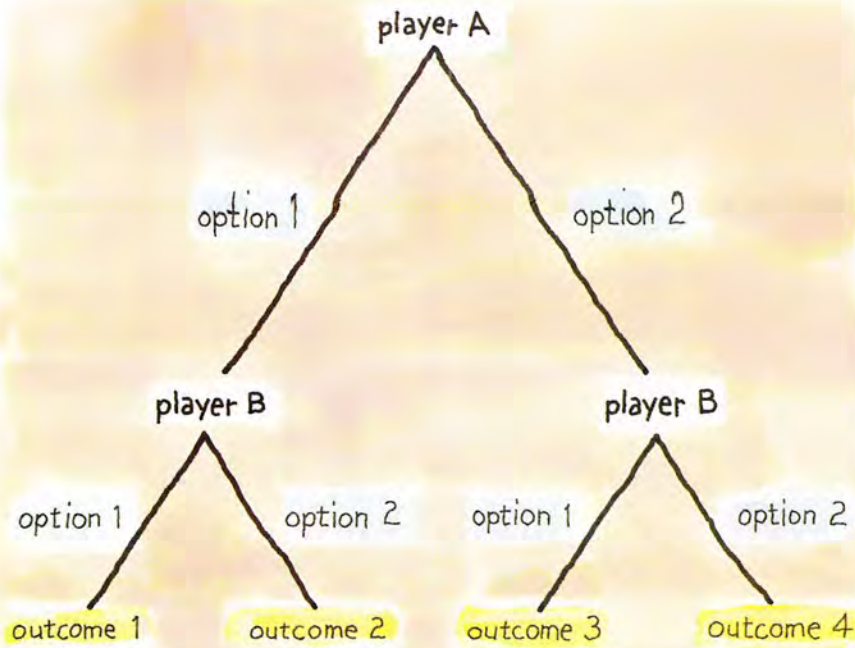
Figure 2. Extensive form of classical game theory is given by a game tree. This game involves two players, with each player able to choose one of two options. The game begins when Player A selects an option. Then Player B selects an option, which leads to one of four possible outcomes. This form highlights the sequential nature of moves.



Figure 3. Normal form of classical game theory is given by a payoff matrix. In a two-player game in which each player has two strategies, the matrix is two-by-two. Player A's strategies are represented by the two rows, and Player B's are represented by the two columns. Players independently select strategies that lead to an outcome. Each outcome is assigned payoffs, which are given in x-y combinations such that $x_i$ is the payoff to the row player (Player A) and $y_j$ is the payoff to the column player (Player B), where $i$ and $j$ are given by the players' strategies (either 1 or 2).

nally, the theory is information-dependent, meaning that players do not always share the same information, making misperception and deception possible.

The theory of moves includes six basic rules. Rule 1 states that a game starts at an "initial state," which is a row-and-column intersection of a payoff matrix. Rule 2 says that either player can switch to a new strategy, thereby generating a new outcome; the first player to move is called Player 1. According to Rule 3, the other player, Player 2, can then move. A game's end is determined by Rule 4: The players respond alternately until neither switches strategies. The resulting outcome is the "final state," which is the only point at which the players accrue payoffs.

The remaining rules, which I call rationality rules, explain the reasons for moving or not moving. Rule 5 states that a player will not make a move unless it leads to a preferred outcome, based on his or her anticipation of the final state. Rule 6, which I call the two-sidedness rule, says that a player considers the rational calculations of the other players before moving, taking into account their possible moves, the possible countermoves of the other players, their own counter-countermoves, and so on. Thus, a player may do immediately better by moving first according to Rule 5; but if this player can do even better by letting the other player move first, and it is rational for that player to do so, then the first player will await this move, according to Rule 6.

Truels

Some of the differences between classical game theory and the theory of moves arise in an imaginary confrontation situation called a truel. A truel is like a duel, except there are three players. It illustrates nicely the applicability of the theory of moves to games with more than two players.

In the truel I posit, a player has two choices: either to fire or not to fire at one of the other two players. Each player has one bullet and is a perfect shot. The players cannot communicate, which prevents the selection of a common target. I assume that a player's primary goal is to survive, and his or her secondary goal is to survive with as few other players as possible.

In this rather gruesome situation, the theory of moves suggests a different outcome than does classical game theory. In fact, the theory of moves provides a resolution that is more satisfactory for all the players.

If the players must make simultaneous strategy choices, they will all fire at each other according to classical game theory. They do so because their own survival does not depend one iota on what they do. Since they cannot affect what happens to themselves but can affect only how many others survive (the fewer the better, according to the postulated secondary goal), they should all fire at each other.

Such a scenario generates two possible results: Either one player survives or no players survive. Players A and B might both fire at Player C, who fires at

one of them, say Player A. This leaves a single survivor, Player B. On the other hand, each player may fire at a different player, leaving all players dead.

If each player has an equal probability of firing at one of the other two players, there is only a 25 percent chance that any player will survive. The reason is that Player A will be killed if fired at by Player B, Player C or both (three cases); the only case in which Player A will survive is if Players B and C fire at each other, which gives Player A one chance in four of surviving. Although this calculation implies a 75 percent chance that some player will survive, an individual player will be more concerned with his or her own low chance (25 percent) of survival.

The theory of moves offers a different perspective. Instead of assuming simultaneous strategy choices, it asks each player: Given your present situation and the situation that you anticipate will ensue if you fire first, should you fire? At the start of a truel, all the players are alive, which satisfies their primary goal of survival but not their secondary goal of surviving with as few others as possible. Player A now contemplates shooting Player B to reduce the number of survivors. By looking ahead, however, Player A realizes that firing at Player B will cause Player C subsequently to fire at him or her (Player A). This would be in Player C's interest, because it would make C the sole survivor.

Instead of firing, therefore, Player A will, thinking ahead, not shoot at anybody. By symmetry, the other players will choose the same strategy, so all will survive. This longer-term perspective leads to a better outcome than that provided by classical game theory, in which each player's primary goal is satisfied only 25 percent of the time when players make simultaneous strategy choices without looking ahead.

The purpose of the theory of moves, however, is not to generate a better outcome but to provide a more plausible model of a strategic situation that mimics what people might think and do. The players in a truel, artificial as such a shoot-out might be, would be motivated to look ahead, given the dire consequences of their actions. To be sure, classical game theory can also provide this outcome if one player (say, A) were designated to move first. Then Player A would rationally choose not to fire, lest he or she be killed subsequently by the sole surviving player (either B or C).
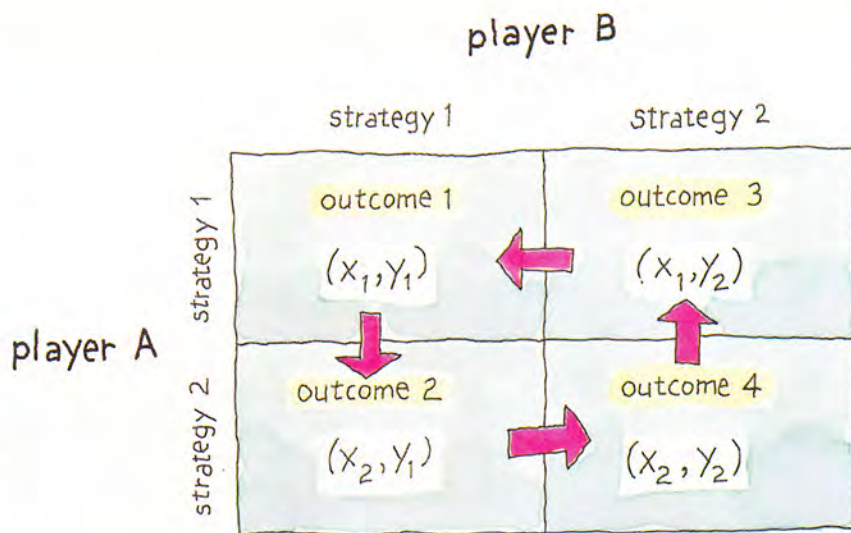


Figure 4. Theory of moves embeds a game tree in a payoff matrix. Players can move from an initial outcome or state to another one by either moving vertically (Player A) or horizontally (Player B). The matrix shows the strategies of the players and their payoffs at various outcomes. The arrows within the matrix reveal how players might move—in this case counterclockwise—between different outcomes through a sequence of moves.

What classical game theory does not ask is whether it is rational for one player, if afforded the opportunity, to move first. This is specified by the rules in the classical theory instead of being made endogenous—that is, incorporated into the theory as a question to be answered—as in the theory of moves.

Changing the rules of play may generate still different outcomes. For example, permit the players of a truel the additional option of firing in the air, thereby disarming themselves, and specify the order of play, such as Player A goes first, followed by Players B and C going simultaneously. Given these rules, Player A will fire in the air, and then Players B and C will shoot each other. The disarmed Player A is, after all, no threat, so he or she would not be shot by Player B or Player C. On the other hand, Players B and C will fire immediately at each other; otherwise, they will have no chance of surviving to get in the last shot. In the end, Player A will be the sole survivor under these rules that give a player the option of firing in the air.

## Prisoners' Dilemma

Game theory's most famous game is called Prisoners' Dilemma. It starts with the following scenario: Two persons, suspected of being partners in a crime,
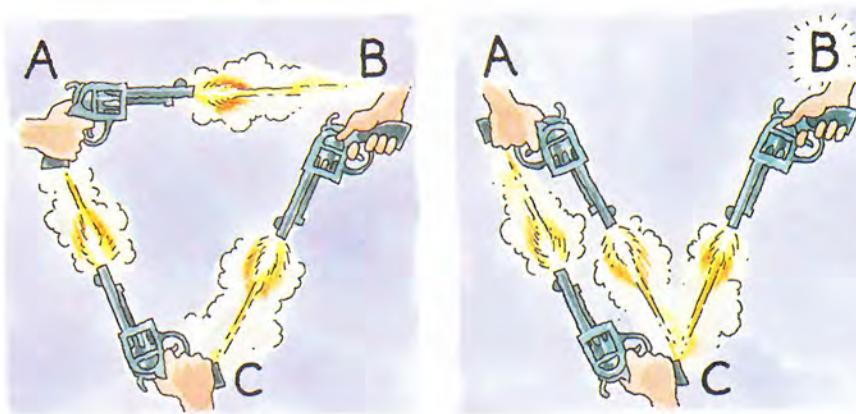


Figure 5. Truel is a three-person duel. Here each player is assumed to have a single bullet and be a perfect shot. A player's primary goal is surviving, and the secondary goal is surviving with as few others as possible. If the players must make simultaneous choices, a player cannot affect his or her own survival but can affect how many others survive; consequently, each player is motivated to shoot another. Classical game theory indicates two possible outcomes: either no one survives (left), or one player—Player B in this case—survives (right).
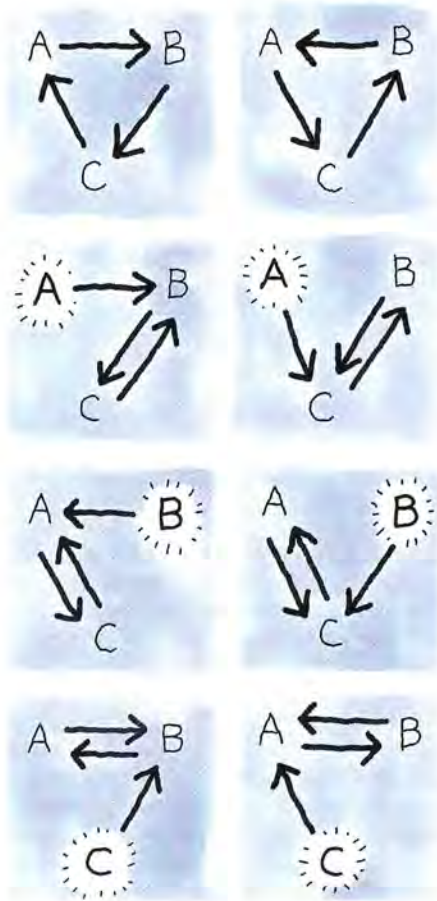
Figure 6. Truel, according to classical game theory, produces a single survivor in 75 percent of the cases. There are eight different firing combinations, and the arrows above reflect who shoots at whom. In two cases (*top row*), no one survives. The other six combinations leave either Player A, Player B or Player C as the sole survivor.
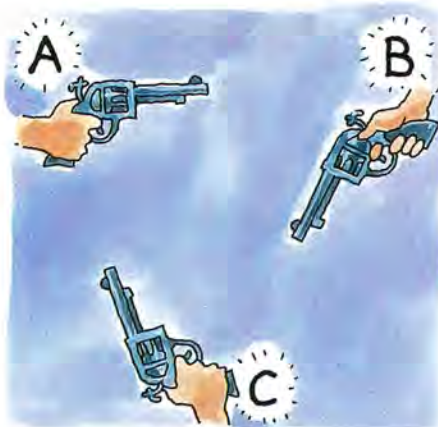


Figure 7. Truel, according to the theory of moves, leaves three survivors. The theory of moves has each player ask: What will happen if I shoot first? If Player A shoots at Player B, Player C will subsequently shoot at Player A, so Player A decides not to shoot. The other players reach the same decision, and they all hold their fire.



Figure 8. Third option of shooting in the air modifies the results of a truel under both classical game theory and the theory of moves. If Player A goes first and shoots in the air, thereby being disarmed, Players B and C must shoot at each other if they want any chance of survival. That leaves Player A as the sole survivor.

are arrested and placed in separate cells so that they cannot communicate with each other. Although they are in fact guilty, without a confession from at least one suspect, the district attorney does not have sufficient evidence to convict them of the crime.

In an attempt to extract a confession, the district attorney explains to each suspect the following consequences of their joint actions. If one suspect confesses and the other does not, the one who confesses can go free for cooperating with the state, but the other gets a 10-year sentence. If both suspects confess, they get reduced sentences of five years. If both remain silent, each goes to prison for one year on a lesser charge of carrying a concealed weapon.

Prisoners' Dilemma can be reduced to a two-by-two payoff matrix, in which 4 is the best payoff, 3 is the next-best, 2 the next-worst and 1 the worst. Each prisoner has a choice of two strategies, confession or silence, giving four possible outcomes for each prisoner: no sentence (4), a one-year sentence (3), a five-year sentence (2) and a 10-year sentence (1). The four possible outcomes that may result are compromise (both remain silent, giving each prisoner a payoff of 3), conflict (both confess, giving each prisoner a payoff of 2) and one or the other "wins" (one confesses and gets a payoff of 4, and the other remains silent and gets a payoff of 1).

The silence-silence outcome is a compromise, because it requires that both players forgo their best payoffs and instead get their next-best payoffs at (3, 3). The confession-confession outcome is a conflict, because if both players try for their best payoffs at (4, 1) or (1, 4) by confessing when the other stays silent, they both get their next-worst payoffs at (2, 2).

The outcome depends on the choices of both players. If both remain silent, they get the compromise, or (3, 3), outcome. Nevertheless, each prisoner has an incentive to defect from that outcome to obtain his or her best payoff at (4, 1) or (1, 4) by confessing when the other remains silent. But if both choose to confess, they bring upon themselves the conflict, or (2, 2), outcome, which is obviously worse for both than (3, 3).

The dilemma, according to classical game theory, is that both players have dominant strategies of confessing, because these choices give each suspect a better payoff regardless of the other suspect's strategy choice. But if both suspects confess, they receive the conflict outcome, which gives them their next-worst payoffs at (2, 2). Moreover, conflict is the unique equilibrium outcome: Neither suspect has an incentive to depart unilaterally from confession, because switching to silence would give that player its worst payoff of 1.

What should the suspects do to save their skins, assuming that neither has any compunction about squealing on the other? If either suspect confesses, it is advantageous for the other to confess as well to avoid the worst payoff, a 10-year sentence. The rub is that confessing when the other suspect does, and receiving a five-year sentence, is not all that appealing, even though neither suspect can ensure a better outcome. Finally, if one suspect remains silent, it is better for the other to turn state's evidence and go free.

There have been many attempts to justify the compromise outcome in which both suspects keep silent. Most of these attempts involve changing the rules of play to allow for repetition. Then, the argument goes, the suspects would recognize the folly of suffering the conflict outcome again and again, come to their senses and both choose silence. In *The Evolution of Cooperation* (1984), however, Robert Axelrod of the University of Michigan argues that the best route to this end will not generally be the unconditional choice of silence. Instead, it will be a tit-for-tat strategy, wherein each player chooses silence

conditional on the other's choice of silence in the previous round.

The theory of moves gives the same result, but in a one-shot play of Prisoners' Dilemma, if the players are sufficiently farsighted to look ahead more than one step and do not start at the conflict outcome. Assume that both suspects are initially silent, which gives the compromise outcome. If one suspect changes from silence to confession, that player also recognizes that the other suspect will, according to the theory of moves, respond by confessing as well. Such a move and countermove would induce the conflict outcome. Therefore, the players will remain silent and stay at the compromise outcome, making this what I call a "nonmyopic equilibrium."

Proceeding from an initial state at which one or the other suspects wins (one suspect remains silent and the other confesses) requires more subtle reasoning. If Suspect 1 remains silent and Suspect 2 confesses, they would get payoffs of 1 and 4, respectively, at (1, 4). Although at this point Suspect 1 would have an immediate incentive to confess, both players would realize only payoffs of 2 from this action at (2, 2). On the other hand, both would realize that if Suspect 2 instead selected silence, they would achieve the compromise outcome at (3, 3). Thus, from the initial states of (3, 3), (4, 1) and (1, 4), the players, thinking ahead, will be motivated to stay at (3, 3) (if already there) or to move to that outcome from (4, 1) or (1, 4). Regrettably, even the theory of moves does not enable the players to escape from (2, 2), so this conflict outcome is also a nonmyopic equilibrium.

Prisoners' Dilemmas have been used to model a great variety of strategic situations, from arms races to price wars to the overpopulation problem. The players in such situations—countries, companies or couples—would prefer that no one buy more arms, reduce prices or have more children. Nevertheless, each of these players does better, whatever the other players do, by engaging in such noncooperative behavior. The consequence is that they all end up worse off than if they had cooperated.

The theory of moves shows that cooperation does not depend necessarily on signing an enforceable contract or playing a repeated game. Instead, cooperation depends on calculating that compromise is the nonmyopic equilibrium, or rational outcome, starting from any state except (2, 2).
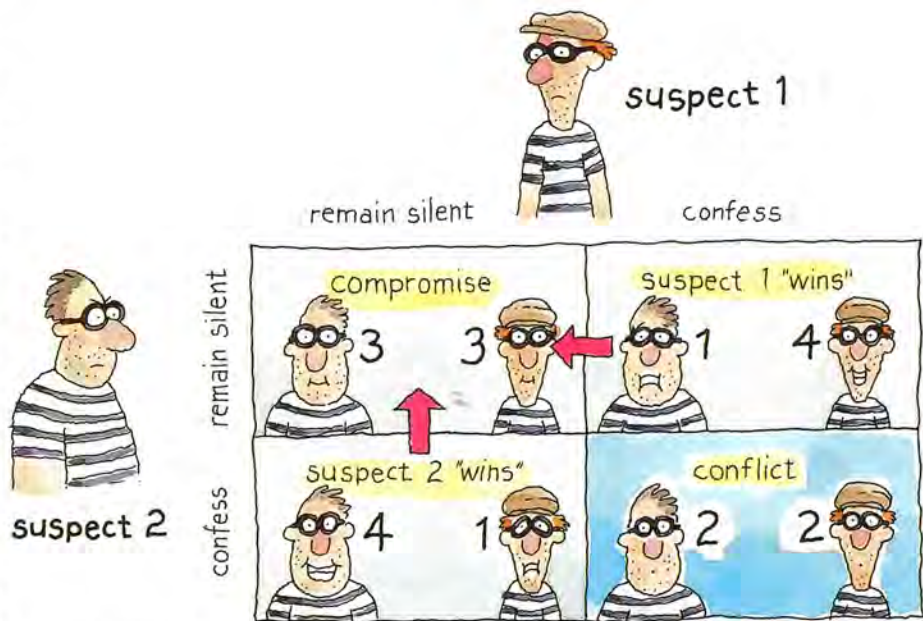


Figure 9. Prisoners' Dilemma involves two suspected criminals. The suspects face the following consequences: If one confesses and the other remains silent, the confessor goes free (the best payoff of 4) and the silent suspect gets a 10-year sentence (the worst payoff of 1); if both confess, they get five-year sentences (the next-worst payoff of 2); and if both remain silent, they get one-year sentences (the next-best payoff of 3). The payoff matrix reveals four possible outcomes: compromise, conflict and one or the other suspect "wins." In classical game theory, both suspects have dominant strategies of confessing. That produces the conflict outcome, which is an equilibrium outcome *(blue)* because neither suspect would move unilaterally from it. The theory of moves can induce the compromise outcome. If the dilemma begins at this outcome (silence), both suspects can look ahead and see that confessing will induce the other suspect to confess as well, so both suspects remain silent. If the dilemma begins with one or the other suspect winning, the winning suspect would move from confession to silence—generating the compromise outcome—to prevent the conflict outcome from occurring when the losing suspect, to do better, decides to confess as well. If the dilemma begins at the conflict outcome, the theory of moves provides no reason to move from it.



Figure 10. Iranian militants seized U.S. embassy personnel in November 1979, initiating the Iran hostage crisis. Iranian students assembled outside the embassy for an anti-American protest on Novermber 5, 1979. By assessing the news reports of that time and later writings of government officials, the author and Walter Mattli recreated the possible moves of that "game." President Carter considered two strategies: negotiation or military intervention. Khomeini also had two strategies: negotiation or obstruction.
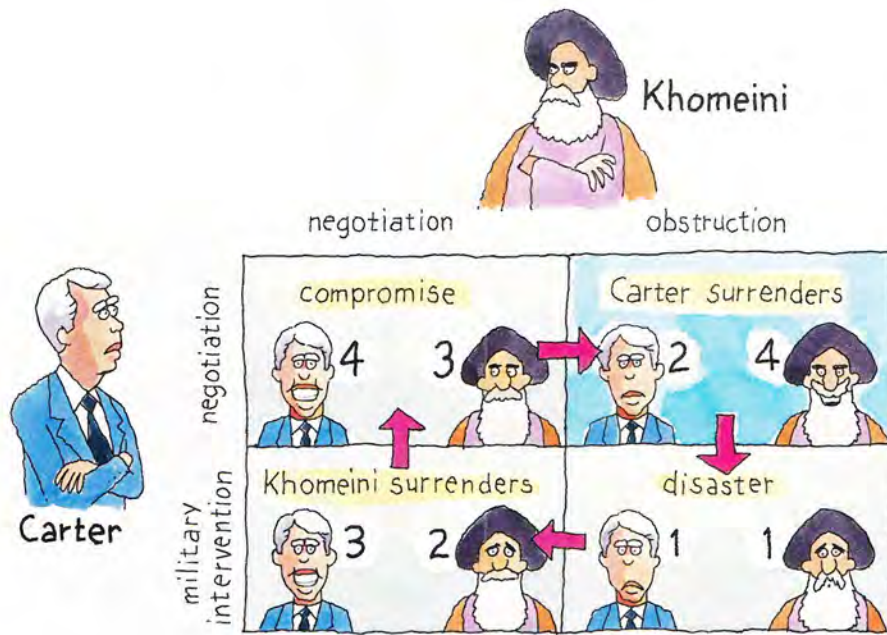
Figure 11. Carter apparently misperceived the structure of the Iran hostage crisis by believing that Khomeini preferred compromise to a confrontation that Carter thought might end in a disaster. Carter's misperceived payoff matrix shows that he gets a better payoff by selecting negotiation, regardless of Khomeini's choice. According to this payoff matrix, Khomeini's best strategy depends on Carter's selection. If Carter selects military intervention, Khomeini should select negotiation. If Carter selects negotiation, Khomeini should select obstruction, resulting in the outcome called "Carter surrenders," which is the equilibrium outcome (blue). If the players moved and countermoved around the matrix, the moves would be clockwise, because in that direction no player ever moves from his best payoff.
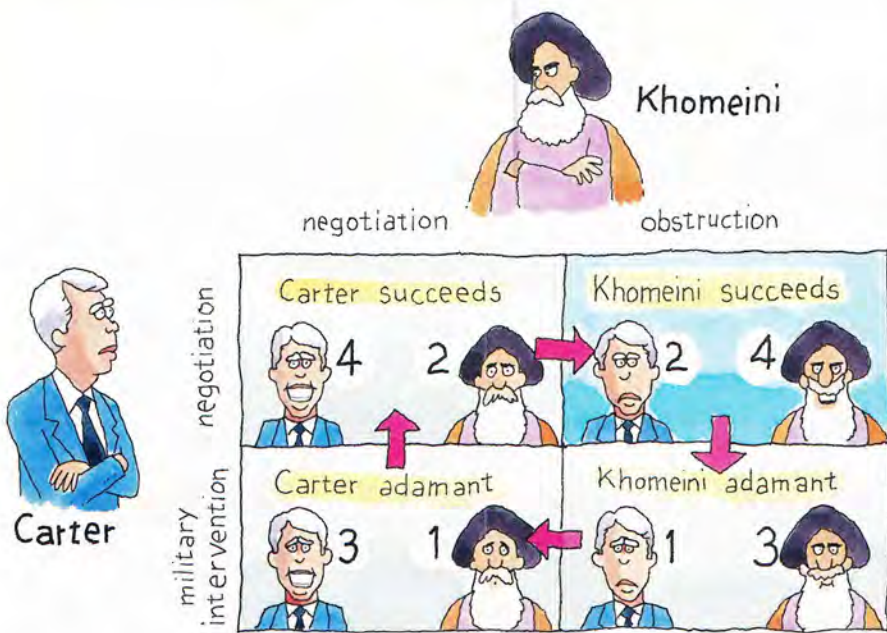


Figure 12. Real-game payoff matrix, taking into account Iran's internal politics as revealed by events and analysis, shows that Khomeini has a dominant strategy of selecting obstruction, which is better for him regardless of Carter's strategy. Like the misperceived-game payoff matrix (Figure 11), Carter has a dominant strategy of selecting negotiation. These strategies lead again to a negotiation-obstruction outcome, which is an equilibrium outcome (blue), now called "Khomeini succeeds" because Khomeini ranks the other outcomes differently than in the misperceived-game payoff matrix. Cycling in this matrix would be clockwise.

## Iran Hostage Crisis

Although a rational player should look ahead before acting, that advice works well only if the players have complete information about their opponents. The United States apparently lacked such information about Iran in November 1979, when Iranian militants seized personnel at the U.S. embassy. By analyzing the news reports of the time and the later writings of some government officials, Walter Mattli of the University of Chicago and I reconstructed the strategic thinking of decision makers in this crisis. As I shall show, it explains well why the crisis took so long to resolve.

During the crisis, the military capabilities of the two opponents were almost irrelevant. In April 1980 the United States attempted a rescue that cost eight American lives and freed no hostages, but the conflict was never really a military one. The crisis can be best represented as a game in which President Jimmy Carter misperceived the preferences of Ayatollah Ruholla Khomeini. In desperation, Carter sought a solution in the wrong game.

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

Carter's primary goal was immediate release of the hostages. His secondary goal was holding discussions with Iranian religious authorities about resolving the differences that had strained relations between the United States and Iran. Of course, if the hostages were killed, the United States would likely defend its honor, probably through a military strike on Iran.

Carter considered two strategies: negotiation and military intervention. Because the seizure of the embassy had led to a severing of diplomatic relations, negotiation could be pursued only through the United Nations Security Council, the World Court or informal diplomatic channels. Military intervention could have taken the form of a rescue mission, as it did, or punitive strikes

against selected targets, such as refineries, rail facilities or power stations.

Khomeini also had two strategies: negotiation or obstruction. His negotiating demands included a return of the shah's assets and ending U.S. interference in Iran's affairs. On the other hand, a refusal to negotiate was sure to block a resolution of the crisis.

The two players and their two strategies generate a two-by-two payoff matrix. Each cell in the matrix has an associated payoff for each player. As in Prisoners' Dilemma, I assume that Carter and Khomeini can rank the four outcomes from best (4) to worst (1).

Carter obtains a better payoff by choosing negotiation, which would save him from the overwhelming difficulties of military intervention, whatever Khomeini does. In December 1979 those difficulties were compounded by the Soviet invasion of Afghanistan, which eliminated the Soviet Union as a possible ally in seeking concerted action for the release of the hostages. Moreover, the Soviet troops next door in Afghanistan made the strategic environment for military intervention anything but favorable.

Carter initially believed that his selection of negotiation would appeal to Khomeini as well. The president perceived that Khomeini faced serious problems in Iran, such as demonstrations by the unemployed and Iraqi incursions across Iran's western border. In Carter's 1982 memoir, *Keeping Faith*, he reported his belief that a U.S. choice of negotiation would give Khomeini a dignified way out of the impasse.

The president also believed that Khomeini preferred a U.S. surrender that would result from the obstruction of negotiations. That result, Carter thought, would give Khomeini his best payoff of 4, whereas Khomeini would get his next best payoff of 3 if both sides selected negotiation. And finally, Carter saw Khomeini getting inferior payoffs of 2 and 1 if the United States selected military intervention.

## Carter's Miscalculations

Unfortunately for Carter, he misperceived the strategic situation and, hence, played the wrong game. Khomeini wanted the total Islamization of Iranian society; he viewed the United States as "a global Shah—a personification of evil" that had to be cut off from any contact with Iran. Khomeini abjured his nation never to "compromise with any

Figure 13. Carter tried military intervention even though negotiation was his dominant strategy in both games. The attempted rescue mission left one U.S. helicopter destroyed and another abandoned. Classical game theory makes such a strategy appear irrational in both game matrices, whereas the theory of moves offers a rational explanation for Carter's action. In the misperceived-game matrix *(Figure 11)*, Carter believes that selecting military intervention—or threatening its use—will force Khomeini to select negotiation in order to improve his payoff, at which point Carter can also choose negotiation to obtain his best payoff at the "compromise" outcome, which is also better for Khomeini. In the real-game payoff matrix *(Figure 12)*, Khomeini cannot be swayed from obstruction. The crisis, in fact, remained at a negotiation-obstruction outcome until the hostages were released on January 20, 1981.

power ... [and] to topple from the position of power anyone in any position who is inclined to compromise with the East and West."

For Khomeini to have selected negotiation would have weakened his uncompromising position. Iranian leaders who tried negotiating, including President Abolhassan Bani-Sadr and Foreign Minister Sagdegh Ghotbzadeh, lost in the power struggle. Bani-Sadr was forced to flee for his life to Paris, and Ghotbzadeh was arrested and later executed.

In the "real game"—the actual strategic situation—Khomeini most preferred obstruction (4 and 3), regardless of the U.S. strategy choice. Doubtless, he preferred that the United States choose negotiation (4) over military intervention (3).

What does classical game theory say about the rational choices of the players in the misperceived game and the real game? In both games, Carter's dominant, or unconditionally best, strategy is negotiation. Regardless of what Khomeini chooses, Carter's payoff from negotiation is better than his payoff from military intervention. In the misperceived game, for example, if Khomeini chooses negotiation, Carter gets a

payoff of 4 by choosing negotiation and a payoff of 3 by choosing military intervention; if Khomeini chooses obstruction, Carter receives a payoff of 2 by choosing negotiation and a payoff of 1 by choosing military intervention.

Although Carter's dominant strategy 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 negotiation, which he should because it is dominant, Khomeini, anticipating this, does better by choosing obstruction, which gives him a payoff of 4, rather than choosing negotiation, which gives him a payoff of 3. So in the misperceived game, Khomeini should choose obstruction, leading to the negotiation-obstruction outcome. That outcome, which I call "Carter surrenders," gives Carter a payoff of 2 and Khomeini a payoff of 4.

Game theory calls this outcome—Carter chooses negotiation and Khomeini chooses obstruction—rational in the real game as well, because both players have dominant strategies associated with it. In the real game, I call this 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 obstruction associated with it; obstruction is not dominant in the real game but, instead, Khomeini's best response if Carter chooses his own dominant strategy of negotiation.

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

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

There are two problems with this reasoning. First, it is not clear that Carter had what I call the "threat power" needed to induce a compromise outcome in the misperceived game. More important, that was not the game being played. In the real game, Khomeini had no reason to accede to a threat from Carter, because his political position was stronger if he refused to compromise. Regardless of Carter's choice, Khomeini does better by selecting obstruction in the real game.

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 in the area. Those two battle groups created the largest U.S. naval force in the Indian Ocean since World War II. But this vast array of firepower proved useless, at least for the purpose of inducing Khomeini to select negotiation.

The failed rescue operation in April 1980 kept the situation at the negotiation-obstruction outcome for another nine months. This was so despite the fact that Iranian leaders had concluded in August 1980—after the installation of an Islamic government consistent with Khomeini's theocratic vision—that keeping the hostages was a net liability.

Further complicating Iran's position was the attack by Iraqi forces in September 1980. It was surely no accident that the hostages were set free on the day of Carter's departure from the White House on January 20, 1981. Although Gary Sick claims in *October Surprise* (1991) that the hostages were not released before the November 1980 presidential election because of a secret deal that Iran made with Ronald Reagan's supporters, later congressional investigations disputed Sick's claim, at least regarding the involvement of George Bush.

Perhaps Carter should not be judged too harshly for misperceiving the strategic situation. If he had correctly foreseen the real game from the start, both game theory and the theory of moves agree that he could not have moved away from an outcome that gave him a payoff of 2 and Khomeini a payoff of 4. What the theory of moves explains, and game theory does not, is why Carter might have thought that he could implement the compromise outcome through the exercise of threats.

The theory of moves also shows how a series of moves and countermoves in the misperceived game can induce this outcome if Carter has what is called "moving power." Assume that the players move and countermove in a clockwise direction on the misperceived-game payoff matrix. In that direction, neither player ever moves from his 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 obstruction to negotiation, and Carter moves from a payoff of 4 at (4, 3) when he switches from negotiation to military intervention. So, if there is cycling, it must be in a clockwise direction.

If Carter believed that he had moving power—the ability to force Khomeini to stop in the move-countermove process—Carter could force Khomeini to stop at the negotiation-negotiation outcome or the military intervention-obstruction outcome, which are the two outcomes where Khomeini has the next move. Khomeini would prefer the former, which gives him a payoff of 3, rather than 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 military intervention-obstruction outcome. As a consequence, Carter's hoped-for negotiation-negotiation outcome in the misperceived game became, in April 1980, a military intervention-obstruction outcome in the real game.

The theory of moves formally incorporates into the framework of game theory an initial state in a payoff matrix, possible moves and countermoves from it to try to reach a nonmyopic equilibrium, and threat and cycling to wear down an opponent. It also allows for the possibility that players possess only incomplete information, as I illustrated in the case of the Iran hostage crisis, which can lead to misperception. As a theory that assumes that players can rank outcomes but not necessarily attach utilities to them, it is eminently applicable to the way we contemplate the strategic choices of others as we try to make our own best choices in a dynamic environment.

## Bibliography

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

Brams, Steven J. 1990. *Negotiation Games: Applying Game Theory to Bargaining and Arbitration*. New York: Routledge.

Brams, Steven J. in press. *Theory of Moves*. Cambridge, U.K.: Cambridge University Press.

Brams, Steven J., and Walter Mattli. 1993. Theory of moves: overview and examples. *Conflict Management and Peace Science* 12(2):1–39.

Carter, Jimmy. 1982. *Keeping Faith: Memoirs of a President*. New York: Bantam Books.

Christopher, Warren (ed.). 1985. *American Hostages in Iran: The Conduct of a Crisis*. New Haven, CT: Yale University Press.

Sick, Gary. 1991. *October Surprise: America's Hostages in Iran and the Election of Ronald Reagan*. New York: Random House.

Von Neumann, John, and Oskar Morgenstern. 1953. *Theory of Games and Economic Behavior*. 3rd ed. Princeton, NJ: Princeton University Press.