CHAPTER 16

THEORY OF MOVES: A DYNAMIC APPROACH TO GAMES

The **theory of moves (TOM),** while based on game theory (see Chapter 15), makes major changes in its rules. These changes make game theory a more dynamic theory. In particular, by postulating that players think ahead not just to the immediate consequences of making moves but also to the consequences of countermoves to these moves, counter-countermoves, and so on, it extends the strategic analysis of conflict into the more distant future.

TOM also elucidates the role that different kinds of power can have in conflicts. In addition, it shows how misinformation, perhaps caused by misperceptions or deception, may affect player choices and game outcomes.

TOM has been applied to a number of different strategic situations in politics, economics, sociology, fiction, and the Bible, among other areas. In this chapter we will mention some of these that can be modeled by simple 2×2 matrix games of the kind already studied in Chapter 15. As was the case for several games analyzed in Chapter 15, we make the assumption that players can only rank outcomes from best to worst, not associate precise numerical payoffs with them. Thus, mixed strategies and expected-value calculations, which depend on quantitative values and probabilities, are not used in TOM. By comparison, TOM throws into bold relief what players can obtain through sequences of moves, based only on pure strategies.

Chess players apply the reasoning of TOM when they think ahead to consequences of moves, countermoves, and so on.

We will describe how TOM helps farsighted players resolve the dilemmas they face in games like Prisoners' Dilemma and Chicken. But first we develop the main ideas of TOM for a less well-known game, called Success, which we use to model the Samson and Delilah story in the Bible. We then consider other games—including games with more than two players or more than two strategies for each player—and ask what TOM says about their strategic characteristics. We relate the analysis at several points to real-life cases in politics and other areas of application.

GAME THEORY REVISITED

In Chapter 15 we used payoff matrices to describe *games in strategic form.* In these games, the row and column players' choices of strategies led to an outcome, from which each player received a payoff. These strategy choices were assumed to be simultaneous. In this chapter, however, we shall use a "game tree" to analyze the *sequential* choices players can make, as occurs when first you move, then I move, and so on, which are called *games in extensive form.*

Game trees can be drawn in different ways. For the three-person game that we illustrate at the end of the chapter, the tree is upside down, branching out at the bottom rather than at the top. But for the two-person games that make up most of this chapter, we indicate the sequential choices of players along a line, going from left to right. This sideways tree provides an economical representation of moves within a 2×2 payoff matrix, as does a related arrow diagram that illustrates cycling within the matrix.

Although TOM starts off with a payoff matrix, it does not assume that players choose strategies simultaneously. Rather, players look ahead and plan their moves, based on rules of play that enable them to make sequential choices.

We begin by applying the rules of standard game theory to a game called Success. Next we use the rules of TOM to show how the rational outcome of this game depends on where play starts and who, if anybody, has certain kinds of power.

EXAMPLE ▶ ***The Game of Success***

Consider the game shown in Figure 16.1, which we call Success. Notice that Row (R) has two strategies, s_1 and s_2, and Column (C) also has two strategies, t_1 and t_2, making Success, in appearance, a 2×2 game much like Prisoners' Dilemma and Chicken. Unlike the latter two games, however, R's two strategies do not lead to the same payoffs as C's two strategies, which makes Success an **asymmetric game.**

For example, if R chooses s_1, he obtains payoffs of either 2 or 4, depending on what C does. Because neither of C's two strategies, t_1 or t_2, leads to these payoffs (4 and 1 if she chooses t_1, 2 and 3 if she chooses t_2), the strategic consequences of play are different for the two players.

Like Prisoners' Dilemma and Chicken, we assume Success is an *ordinal game:* the payoffs indicate only an ordering of outcomes from best to worst ($4 =$ best, $3 =$ next best, $2 =$ next worst, and $1 =$ worst). As before, the higher the rank, the better the payoff, but the ranks do not indicate whether a player prefers, say, 4 to 3 more than 2 to 1, or vice versa.

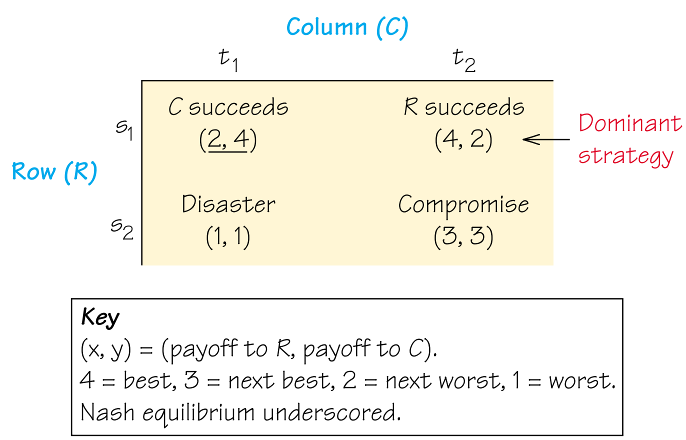


FIGURE 16.1 The game of Success according to standard game theory.

Assume R chooses s_1 and C chooses t_1 in the Figure 16.1 payoff matrix. The resulting outcome is that shown in the upper left-hand corner of the matrix, with a payoff of 2 to R and 4 to C, or next worst for R and best for C. As shorthand verbal descriptions of these outcomes, we call $(2, 4)$ "C succeeds," $(4, 2)$ "R succeeds," $(3, 3)$ "Compromise," and $(1, 1)$ "Disaster."

Consider what standard game theory, in which players are assumed to make simultaneous strategy choices, tells us about this game. (If the players' choices are not literally simultaneous, game theory assumes they are made independently of each other, so neither R nor C knows the other's choice when each makes his or her own choice.) First consider what strategy is rational for R to choose. If C selects t_1, R has a choice between $(2, 4)$ and $(1, 1)$ in the first column; his payoff will be 2 if he chooses s_1 and 1 if he chooses s_2. By comparison, if C chooses t_2, R has a choice between $(4, 2)$ and $3, 3)$ in the second column; his payoff will be 4 if he chooses s_1 and 3 if he chooses s_2.

Clearly, R is better off choosing s_1 regardless of the strategy that C chooses (t_1 or t_2), which, as we showed in Chapter 15, makes s_1 a *dominant* strategy over s_2. By contrast, R's strategy of s_2 is *dominated* by s_1, because it always leads to worse payoffs than s_1, whichever strategy C chooses.

C, on the other hand, does not have a dominant strategy in Success. Her better strategy depends on R's strategy choice: if R chooses s_1, C is better off choosing t_1 because she prefers $(2, 4)$ to $(4, 2)$ in the first row; but if R chooses s_2, C is better off choosing t_2 because she prefers $(3, 3)$ to $(1, 1)$ in the second row. In Chicken, recall from Chapter 15, neither player has a dominant strategy, whereas both players have dominant strategies in Prisoners' Dilemma.

We assume Success to be a game of **complete information,** meaning that both players have full knowledge of the rules and each other's payoffs as well as their own. Therefore, C will know that R's dominant strategy is s_1. Because s_1 is always better than s_2 for R, C can surmise that R will choose s_1. Given that R chooses s_1, it is rational for C to choose t_1, yielding $(2, 4)$ as the rational outcome of Success.

Curiously, this outcome is only R's next worst (2), though R is the player with the dominant strategy. C, the player without a dominant strategy, obtains her best outcome (4). Nevertheless, $(2, 4)$ has a strong claim to be called *the* solution of Success. Not only is it the product of one player's (R's) dominant strategy and the other player's (C's) best response to this dominant choice, but it is also the unique Nash equilibrium, as was $(2, 2)$ in Prisoners' Dilemma.

Recall from Chapter 15 that a *Nash equilibrium* is an outcome—or, more precisely, the strategies associated with this outcome, which are said to be "in equilibrium"—from which neither player would unilaterally depart because he or she would do worse doing so. Thus, if R chooses s_1 and C chooses t_1, giving $(2, 4)$, R will not switch to s_2 because he would do worse at $(1, 1)$; and C will not switch to t_2 because she would do worse at $(4, 2)$. Hence, $(2, 4)$ is stable in the sense that, once chosen, neither player would have an incentive to switch to a different strategy, given that the other player does not switch.

This is not true in the case of the other three outcomes in Success. From $(4, 2)$, C can do better by departing to $(2, 4)$; from $(3, 3)$, R can do better by departing to $(4, 2)$; and from $(1, 1)$, either player can do better by departing, R to $(2, 4)$ and C to $(3, 3)$.

In the latter case, if *both* players switched their strategies in an effort to scramble away from the mutually worst outcome of $(1, 1)$, they would end up at $(4, 2)$, which also is better for both. Indeed, because $(4, 2)$ is R's best outcome, R would be the player who would most welcome a double departure; next most welcome would be a departure by C alone to $(3, 3)$; and least welcome a departure by just himself to $(2, 4)$.

C would not particularly welcome a double departure, obtaining only her next-worst payoff of 2. Like R, she would prefer that her adversary make the first move from $(1, 1)$, because R's departure would yield $(2, 4)$, whereas C's departure yields $(3, 3)$. There are other games, as we shall see later, in which the opposite is true: each player would prefer to be the *first* to depart from an inferior outcome rather than wait for his or her adversary to make the first move. ◆

Standard game theory, by assuming that the players choose strategies simultaneously, does not raise questions about the rationality of moving or departing from outcomes (beyond an immediate departure). In fact, however, most real-life games do not start *with* simultaneous strategy choices and then ask whether the resulting outcome is stable in the sense of being a Nash equilibrium. Rather, play begins with the players already *in* some state. The question then becomes whether or not they would benefit from staying in this state or moving, given the possibility that a move will set off a series of subsequent moves by the players. In the next section we discuss how the rules of TOM alter the players' calculations of stable outcomes in Success.

To get some perspective on our analysis of Success and other games, we note here that there are 78 distinct 2×2 ordinal games that are structurally distinct in the sense that no interchange of the players, their strategies, or any combination of these can transform one of these games into any other. These games represent *all* the different configurations of ordinal payoffs in which two players, each with two strategies, may find themselves embedded. (See Rapoport et al. in Suggested Readings for how the number 78 was determined.)

Success is only one such configuration; Prisoners' Dilemma and Chicken are two others. The rules of play we describe next apply to all 78 2×2 games. These rules can be extended to larger games, as we will illustrate later.

RULES OF TOM

The founders of game theory, John von Neumann and Oskar Morgenstern, defined a **game** to be "the totality of rules of play which describe it." The rules of TOM apply to all ordinal games between two players, each of whom has two strategies. The first four **rules of play** of TOM are as follows:

1. Play starts at an **initial state,** given at the intersection of the row and column of a payoff matrix (i.e., one of the four entries in a 2×2 payoff matrix).
2. Either player can unilaterally switch his or her strategy (i.e., make a **move**), thereby changing the initial state into a new state, in the same row or column as the initial state. The player who switches is called player 1 (*P1*).
3. Player 2 (*P2*) can respond by unilaterally switching his or her strategy, thereby moving the game to a new state.
4. The alternating responses continue until the player (*P1* or *P2*) whose turn it is to move next chooses not to switch his or her strategy. When this happens, the game terminates in a **final state,** which is the **outcome** of the game.

Note that the sequence of moves and countermoves is *strictly alternating.* First, say, *R* moves, then *C* moves, and so on, until one player stops, at which point the state reached is final and therefore the outcome of the game. We assume that no payoffs accrue to players from being in a state unless it becomes the outcome (which could be the initial state if the players choose not to move from it).

To assume otherwise would require that payoffs be numerical values, rather than ordinal ranks, which players can accumulate as they pass through states. But in most real-life games, payoffs cannot easily be quantified and summed across the states visited; moreover, the big reward in many games depends overwhelmingly on the final state reached, not on how it was reached. In politics, for example, the payoff for most politicians is not in campaigning, which is arduous and costly, but in winning.

Rule 1 differs radically from the corresponding rule of play in standard game theory, in which players simultaneously choose strategies in a matrix game that determines the outcome. Instead of starting with strategy choices, TOM assumes that players are already *in* some state at the start of play and receive payoffs from this state *only if they stay.* Based on these payoffs, they decide, individually, whether or not to change this state in order to try to do better.

To be sure, some decisions are made collectively by players, in which case it would be reasonable to say that they choose strategies from scratch, either simultaneously or by coordinating their choices. But if, say, two countries are coordinating their choices, as when they agree to sign a treaty, the important strategic question is what individualistic calculations led them to this point. The formality of jointly signing the treaty is simply the culmination of their negotiations and does not reveal the move-countermove process that preceded the signing. It is precisely these negotiations, and the calculations underlying them, that TOM is designed to uncover.

To continue this example, the parties who sign the treaty were in some prior state, from which both desired to move—or, perhaps, only one desired to move and the other could not prevent this move from happening. Eventually they may arrive at a new state (e.g., after treaty negotiations) in which it is rational for both countries to sign the treaty that has been negotiated.

Put another way, almost all outcomes of games that we observe have a history. TOM seeks to explain strategically the progression of (temporary) states that lead to a (more permanent) outcome. Consequently, play of a game starts in an initial state, at which players accrue payoffs only if they remain in that state so that it becomes the final state, or outcome, of the game.

If they do not remain in the initial state, they still know what payoffs they would have accrued had they stayed; hence, they can make a rational calculation of the advantages of staying or moving. They move precisely because they calculate that they can do better by switching states, anticipating a better outcome when the move-countermove process finally comes to rest. The game is different, but not the payoff matrix, when play starts in a different state.

Rules 1–4 say nothing about what *causes* a game to end, but only when: termination occurs when a "player whose turn it is to move next chooses not to switch his or her strategy" (rule 4). But when is it rational not to continue moving, or not to move at all from the initial state?

To answer this question partially, TOM postulates a **termination rule:**

5. If play returns to the initial state, the initial state becomes the outcome.

We will illustrate shortly how rational players, starting from some initial state, can predict what the outcome will be. The initial state will *not* be the outcome if the players find it rational to terminate play before returning to it. On the other hand, if after *P1* moves, it is rational for play of the game to cycle back to the initial state, then rule 5 says that there will be no further movement. After all, what is the point of continuing the move-countermove process if play will, once again, return to "square one," given that the players receive no payoffs along the way (i.e., before the outcome is reached)?

At this point, we make rule 5 only provisional. An alternative rule (5′) that allows for cycling will be considered later (along with "moving power" as a way to break cycles).

A final rule of TOM is needed to ensure that *both* players take into account each other's calculations before deciding to move from the initial state. We call this rule the **two-sidedness rule:**

6. Each player takes into account the consequences of the other player's **rational choices,** as well as his or her own, in deciding whether or not to move from the initial state or any subsequent state. If it is rational for one player to move and the other player not to move from the initial state, then the player who moves takes **precedence:** his or her move overrides the player who stays, so the outcome will be induced by the player who moves.

Later we will show that if both players want to move first, or both want the other player to move first, this conflict can be resolved by "order power."

EXAMPLE ▶ ***Applying TOM to Success***

If players have complete information, they can look ahead and anticipate the consequences of their moves and thereby decide whether or not to move from the initial state or any subsequent states reached. We next show how they can use backward induction to make this decision.

Backward induction is a reasoning process in which players, working backward from the last possible move in a game, anticipate each other's rational choices.

To illustrate backward induction, consider again Success in Figure 16.2 (the payoffs in brackets, just below the payoffs in parentheses, will be defined shortly). We show next the progression of moves, starting from each of the four possible initial states of Success and cycling back to this state, and indicate where rational players will terminate play:

1. Initial state (2, 4). If R moves first, the counterclockwise progression of moves from $(2,4)$ back to $(2,4)$—with the player (R or C) who makes the next move shown below each state in the alternating sequence—is as follows (see Figure 16.2):

	State 1 R		State 2 C		State 3 R		State 4 C		State 1
R starts:	(2, 4)	→	(1, 1)	→	$\underline{(3, 3)}$	→|	(4, 2)	→	(2, 4)
Survivor:	(3, 3)		(3, 3)		(3, 3)		(2, 4)		

Below the progression of states we indicate the survivor at each state.

The **survivor** is the payoff selected at each state as the result of backward induction. It is determined by working backward, after a cycle has been completed and play returns to the initial state (state 1).

Assume the players' alternating moves have taken them counterclockwise in Success from (2, 4) to (1, 1) to (3, 3) to (4, 2), at which point C must decide whether to stop at (4, 2) or complete the cycle and return to (2, 4). Clearly, C prefers (2, 4) to (4, 2), so (2, 4) is listed as the survivor below (4, 2): because C *would* move the process back to (2, 4) should she reach (4, 2), the players know that if the move-countermove process reaches this state, the outcome will be (2, 4).

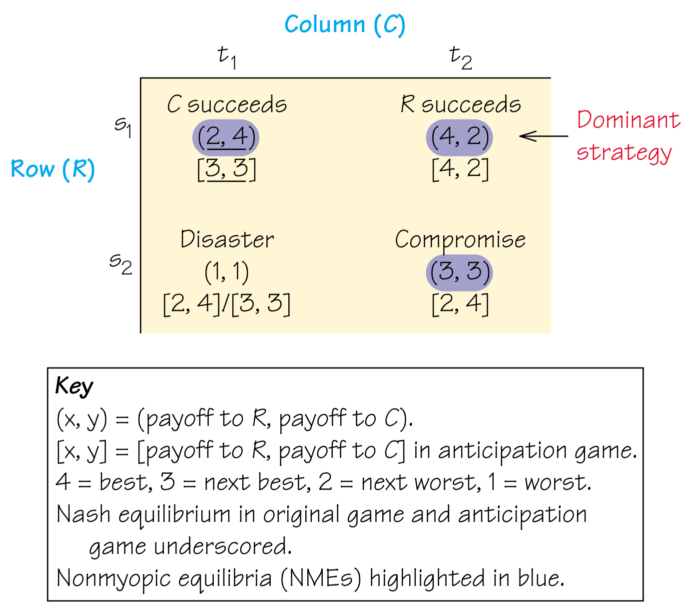


FIGURE 16.2 The game of Success according to the theory of moves.

Knowing this, would *R* at the prior state, (3, 3), move to (4, 2)? Because *R* prefers (3, 3) to the survivor at (4, 2)—namely, (2, 4)—the answer is no. Hence, (3, 3) becomes the survivor when *R* must choose between stopping at (3, 3) and moving to (4, 2)—which, as we just showed, would become (2, 4) once (4, 2) is reached.

At the prior state, (1, 1), *C* would prefer moving to (3, 3) than stopping at (1, 1), so (3, 3) again is the survivor if the process reaches (1, 1). Similarly, at the initial state, (2, 4), because *R* prefers the previous survivor, (3, 3), to (2, 4), (3, 3) is the survivor at this state as well.

The fact that (3, 3) is the survivor at initial state (2, 4) means that it is rational for *R* initially to move to (1, 1), and *C* subsequently to move to (3, 3), where the process will stop, making (3, 3) the rational choice if *R* has the opportunity to move first from initial state (2, 4). That is, after working *backward* from *C*'s choice of completing the cycle or not from (4, 2), the players can reverse the process and, looking *forward,* determine that it is rational for *R* to move from (2, 4) to (1, 1), and *C* to move from (1, 1) to (3, 3), at which point *R* will stop the move-countermove process at (3, 3).

Notice that *R* does better at (3, 3) than at (2, 4), where he could have terminated play at the outset, and *C* does better at (3, 3) than at (1, 1), where she could have terminated play, given that *R* is the first to move. We indicate that (3, 3) is the consequence of backward induction by underscoring this state in the progression; it is the state at which **stoppage** of the process occurs. In addition, we indicate that it is not rational for *R* to move on from (3, 3) by the vertical line blocking the arrow emanating from (3, 3), which we refer to as **blockage:** a player will always stop at a blocked state, wherever it is in the progression. Stoppage occurs when blockage occurs for the *first* time from some initial state, as we illustrate next.

If *C* can move first from (2, 4), backward induction shows that (2, 4) is the last survivor, so (2, 4) is underscored when *C* starts. Consequently, *C* would *not* move from the initial state, where there is blockage (and stoppage), which is hardly surprising since C receives her best payoff in this state:

	C		*R*		*C*		*R*		
C starts:	$\underline{(2,4)}$	$\rightarrow\vert$	(4, 2)	$\rightarrow\vert$	(3, 3)	$\rightarrow$	(1, 1)	$\rightarrow$	(2, 4)
Survivor:	(2, 4)		(4, 2)		(2, 4)		(2, 4)		

As when *R* has the first move, (2, 4) is the first survivor, working backward from the end of the progression, and is also preferred by *C* at (3, 3). But then, because *R* at (4, 2) prefers this state to (2, 4), (2, 4) is temporarily displaced as the survivor. It returns as the last survivor, however, because *C* at (2, 4) prefers it to (4, 2). Nonetheless, there are circumstances in which a player does better by departing from a best state, paradoxical as this may seem (see Exercises 7–9).

Thus, the first blockage and, therefore, stoppage occur at (2, 4), but blockage occurs subsequently at (4, 2) if, for any reason, stoppage does not terminate moves at the start. In other words, if *C* moved initially, *R* would then be blocked. Hence, blockage occurs at two states when *C* starts the move-countermove process, whereas it occurs only once when *R* has the first move.

The fact that the rational choice depends on which player has the first move—(3, 3) is rational if *R* starts, (2, 4) if *C* starts—leads to a conflict over what outcome will be selected when the process starts at (2, 4). However, because it is not rational for *C* to move from the initial state, *R*'s move takes precedence, according to rule 6, and overrides *C*'s decision to stay. Consequently, when the initial state is (2, 4), the outcome will be (3, 3).

2. Initial state (4, 2). The progressions, survivors, stoppages, other blockages, and outcome from this state are as follows:

	R		*C*		*R*		*C*		
R starts:	<u>(4, 2)</u>	→|	(3, 3)	→	(1, 1)	→	(2, 4)	→|	(4, 2)
Survivor:	(4, 2)		(2, 4)		(2, 4)		(2, 4)		

	C		*R*		*C*		*R*		
C starts:	<u>(4, 2)</u>	→|*c*	(2, 4)	→	(1, 1)	→	(3, 3)	→	(4, 2)
Survivor:	(4, 2)		(4, 2)		(4, 2)		(4, 2)		
Outcome:	(4, 2)								

Clearly, when (4, 2) is the initial state, there is no conflict between *R* and *C* about staying there. Yet, while neither player has an incentive to move from (4, 2), each player's reasons for stoppage are different. If *R* starts, there is blockage at the start, whereas if *C* starts, there will be cycling back to (4, 2).

Because cycling is no better for *C* than not moving, we assume that *C* will stay at (4, 2), which we indicate by *c* (for "cycling") following the arrow at (4, 2). This might be interpreted as a special kind of blockage: while rule 5 allows play to return to the initial state after one complete cycle, whence it terminates, there is no benefit to *C* from doing so. Consequently, we assume that *C* will not move initially from (4, 2), simply to cycle once. Because *R* also will not move, there will be a consensus on the part of both players of staying at (4, 2).

3. Initial state (3, 3). The progressions, survivors, stoppages, other blockages, and outcome from this state are as follows:

	R		C		R		C		
R starts:	$\underline{(3,3)}$	$\rightarrow\mid_C$	(4, 2)	→	(2, 4)	→	(1, 1)	→	(3, 3)
Survivor:	(3, 3)		(3, 3)		(3, 3)		(3, 3)		

	C		R		C		R		
C starts:	(3, 3)	→	(1, 1)	→	$\underline{(2,4)}$	$\rightarrow\mid$	(4, 2)	$\rightarrow\mid$	(3, 3)
Survivor:	(2, 4)		(2, 4)		(2, 4)		(4, 2)		
Outcome:	(2, 4)								

As from initial state (2, 4), there is a conflict. If *R* starts, (3, 3) is the rational choice, but if *C* starts, (2, 4) is. But because *C*'s move takes precedence over *R*'s staying, the outcome is that which *C* can induce—namely, (2, 4).

4. Initial state (1, 1). The progressions, survivors, stoppages, other blockages, and outcome from this state are as follows:

	R		C		R		C		
R starts:	(1, 1)	→	$\underline{(2,4)}$	$\rightarrow\mid$	(4, 2)	$\rightarrow\mid$	(3, 3)	$\rightarrow\mid$	(1, 1)
Survivor:	(2, 4)		(2, 4)		(4, 2)		(3, 3)		

	C		R		C		R		
C starts:	(1, 1)	→	$\underline{(3,3)}$	$\rightarrow\mid$	(4, 2)	→	(2, 4)	→	(1, 1)
Survivor:	(3, 3)		(3, 3)		(2, 4)		(2, 4)		

Outcome: Indeterminate—(2, 4)/(3, 3), depending on whether *R* or *C* starts.

Unlike the conflicts from initial states (2, 4) and (3, 3), it is rational for *both* players to move from initial state (1, 1). But, as we showed earlier, each player would prefer that the other player be *P1*, because

- *R*'s initial move induces (2, 4), *C*'s preferred state; and
- *C*'s initial move induces (3, 3), *R*'s preferred state.

Presumably, each player will try to hold out longer at (1, 1), hoping that the other will move first. Because neither player's move takes precedence according to the rules of play, neither rational choice can be singled out as *the* outcome. Hence, when play starts at $(1, 1)$, we classify the state as **indeterminate**—either $(2, 4)$ or $(3, 3)$ can occur, depending on which player *P1* is; we write this state as $(2, 4)/(3, 3)$. Because the choice of first mover is not specified by the rules of play, indeterminacy is a consequence of TOM.

Typically, this kind of indeterminacy is characterized by bargaining, wherein each player tries to hold off being the first to make concessions. Although both players would benefit at either $(2, 4)$ or $(3, 3)$ over $(1, 1)$, there is greater benefit to each in having the other player move first.

A player has **order power** if he or she can determine the order of moves from an indeterminate state.

Thus, if R has order power, he can force C to depart first from $(1, 1)$ and induce his preferred state of $(3, 3)$; if C has order power, she can induce $(2, 4)$ as the state. Note, however, that the state that R most prefers, $(4, 2)$, is unattainable from $(1, 1)$—it can occur only if the process starts at $(4, 2)$.

To summarize, each of the initial states goes into the following final determinate states, or outcomes—except when there is a conflict, as there is from (1, 1), and neither player's move takes precedence, according to rule 6, and neither player wants to move first from $(1, 1)$:

$$(2, 4) \rightarrow (3, 3); \quad (4, 2) \rightarrow (4, 2); \quad (3, 3) \rightarrow (2, 4); \quad (1, 1) \rightarrow (2, 4)/(3, 3).$$ ◆

The outcomes into which each state goes are **nonmyopic equilibria (NMEs).** They are the consequence of both players' looking ahead and anticipating where, from each of the initial states, the move-countermove process will end up.

We hasten to add that once reached, a player may have an incentive to leave an NME. Thus, NMEs might be better thought of as "reachable outcomes" from a state, rather than equilibria, because they may not be stable once reached. Only $(4, 2)$ is stable in the sense that, if it is the initial state, neither player would depart from it. Although this is an argument for considering $(4, 2)$ to be "more stable" than the other NMEs in Success, we shall not distinguish among the NMEs in a game. Perhaps the best way to think about NMEs is as states (i) at which players would stay or (ii) to which they would migrate, but if the latter, not necessarily stay there.

In Success the players can end up at every state except $(1, 1)$, which therefore is not an NME. Note that from each initial state except $(1, 1)$ the NMEs are unique. From (1, 1) the NMEs are (2, 4) or (3, 3), depending on whether *C* or *R* has order power and can thereby dictate who moves first from this initial state (recall that each player would like to go second).

The move-countermove process can be interpreted as a bargaining process in which, starting at the initial state, a player can choose not to move (i.e., to accept an offer) or to move (i.e., reject an offer). If a player chooses to move, the other player can then terminate the game by accepting the offer, or continue it by moving to an adjacent offer, which may in turn be accepted or rejected, and so on.

If this process did not stabilize, the initial offer (i.e., first move) would not be worth making. But every 2 × 2 game contains at least one NME, because from each initial state there is an outcome (perhaps indeterminate) of the move-countermove process. If this outcome is both determinate and the same from every initial state, then it is the only NME; otherwise, there is more than one NME (Success has three).

In the offer-counteroffer interpretation, then, there is at least one offer that will always be accepted, so the process always stabilizes. But this may occur at more than one outcome if there is more than one NME, as in Success. In this situation, there might be a kind of positioning game played over the choice of an initial state, which we can analyze in terms of an anticipation game.

An **anticipation game** is simply the game resulting from the substitution of the NMEs into which each of these states goes (in brackets in Figure 16.2) for the original payoffs at each of the four states (in parentheses in Figure 16.2).

When we apply standard game theory to the anticipation game, s_1 is a dominant strategy for *R*: if *C* chooses t_1, [3, 3] is at least as good for *R* as [2, 4]/[3, 3]; if *C* chooses t_2, [4, 2] is better for *R* than [2, 4]. By contrast, *C*'s strategies are undominated, but anticipating that *R* will choose s_1, *C*'s best response is t_1, yielding [3, 3]. Thus, if the players believe that they are choosing only initial states rather than outcomes when they choose their strategies, their choices of s_1 and t_1 will start them out at (2, 4), whence a move by *R* and a countermove by *C* will, according to TOM, bring them to the NME of (3, 3).

Players in most real-life games, whether they do or do not anticipate making moves from some initial state, rarely choose strategies simultaneously, as we argued earlier. Rather, they find themselves in some state, or status quo point, from which they consider moving. In Success, as we have seen, moves and countermoves can lead the players to three different NMEs, which is the maximum number that can occur in a 2 × 2 strict ordinal game; the minimum, as already noted, is one. Most 2×2 games have either one or two NMEs; in fact, Chicken and Success are the *only* 2×2 games (of the 78) that have three NMEs.

In these games, in particular, *where* play starts matters, which the unique $(2, 4)$ Nash equilibrium in Success masks. To show how TOM gives insight into player choices in such games, consider the story of Samson and Delilah from the Book of Judges in the Old Testament.

INTERPRETING TOM

In using the game of Success to model the conflict between Samson and Delilah, we seek to show that Samson's behavior was *not* irrational, despite his ample later troubles. On the contrary, we will argue that his moves and Delilah's were entirely rational in Success.

To provide evidence that Samson and Delilah had the preferences of the players in Success, we quote extensively from the Bible. We suggest in the end, however, that it is possible that Delilah's preferences were different from those of the row player in Success—but this does not make a difference for rational play of the game.

Samson and Delilah, played by Victor Mature and Hedy Lamarr, 1949.

EXAMPLE ▶ ***Samson and Delilah***

The background to the story is as follows: After aiding the flight of the Israelites from Egypt and delivering them into the promised land of Canaan, God became extremely upset by their recalcitrant ways and punished them severely:

> The Israelites again did what was offensive to the LORD, and the LORD delivered them into the hands of the Philistines for forty years. (Judg. 13:1)

But a new dawn appears at the birth of Samson, which is attended to by God and whose angel predicts: "He shall be the first to deliver Israel from the Philistines" (Judg. 13:5).

Samson developed a reputation as a ferocious warrior of inhuman strength. This served him well as judge (leader) of Israel for 20 years, but then he fell in love with a Philistine woman named Delilah. Apparently, Samson's love for Delilah was not reciprocated. Rather, Delilah was more receptive to serving as bait for Samson for a suitable payment. The lords of the Philistines made her a proposition:

> Coax him and find out what makes him so strong, and how we can overpower him, tie him up, and make him helpless; and we'll each give you eleven hundred shekels of silver. (Judg. 16:5)

After agreeing, Delilah asked Samson: "Tell me, what makes you so strong? And how could you be tied up and be made helpless?" (Judg. 16:6). Samson replied: "If I were to be tied with seven fresh tendons that had not been dried, I should become as weak as an ordinary man" (Judg. 16:7).

After Delilah bound Samson as he had instructed her, she hid men in the inner room and cried, "Samson, the Philistines are upon you!" (Judg. 16:9). Samson's lie quickly became apparent:

> Whereat he pulled the tendons apart, as a strand of tow [flax] comes apart at the touch of fire. So the secret of his strength remained unknown.
>
> Then Delilah said to Samson, "Oh, you deceived me; you lied to me! Do tell me how you could be tied up." (Judg. 16:9–10)

Twice more Samson lied to Delilah about the source of his strength, and she became progressively more frustrated by his deception. In exasperation, Delilah exclaimed:

> "How can you say you love me, when you don't confide in me? This makes three times that you've deceived me and haven't told me what makes you so strong." Finally, after she had nagged him and pressed him constantly, he was wearied to death and he confided everything to her. (Judg. 16:15–17)

The secret, of course, was Samson's long hair. When he told his secret to Delilah, she had his hair shaved off while he slept. The jig was then up when he was awakened:

> For he did not know that the LORD had departed him. The Philistines seized him and gouged out his eyes. They brought him down to Gaza and shackled him in bronze fetters, and he became a mill slave in the prison. After his hair was cut off, it began to grow back. (Judg. 16:20–22)

Thus is a slow time bomb set ticking. The climax approaches when Samson is summoned by the Philistines to a great celebration:

> They put him between the pillars. And Samson said to the boy who was leading him by the hand, "Let go of me and let me feel the pillars that the temple rests upon, that I may lean on them." Now the temple was full of men and women; all the lords of the Philistines were there, and there were some three thousand men and women on the roof watching Samson dance. Then Samson called to the LORD, "O Lord GOD! Please remember me, and give me strength just this once, O God, to take revenge of the Philistines if only for one of my two eyes." (Judg. 16:25–28)

Samson, his strength now restored, avenged himself on his captors in an unprecedented biblical reprisal (by a human being, not God) that sealed both his doom and the Philistines':

> He embraced the two middle pillars that the temple rested upon, one with his right arm and one with his left, and leaned against them; Samson cried, "Let me die with the Philistines!" and he pulled with all his might. The temple came crashing down on the lords and on all the people in it. Those who were slain by him as he died outnumbered those who had been slain by him when he lived. (Judg. 1:29–30)

There is irony, of course, in this reversal of roles, whereby the victim becomes the victor—and a victim as well. We do not suggest, however, that Samson planned for his own mutilation and ridicule only to provide himself with the later opportunity to retaliate massively against the Philistines. Perhaps this was in God's design, as foretold by the angel at Samson's birth, and in God's "seeking a pretext against the Philistines" (Judg. 14:4).

Although Samson's betrayal of the secret of his strength may seem stupid, it was entirely consistent with his previous behavior and apparent preferences. To put it bluntly, Samson was a man of carnal desires: he had lusted after several women before meeting Delilah, and Delilah was not the first to whose charms he fell prey. He would not withhold information if the right woman was around to wheedle it out of him. While Samson could fight the Philistines like a fiend, he could readily be disarmed by women he desired.

FIGURE 16.3 Samson and Delilah, as modeled by Success and Variation. The payoffs of Success are to the left of the slashes and the payoffs of Variation are to the right.

		Samson: Don't tell secret ($\overline{T}$)	Samson: Tell secret (T)
Delilah	Don't nag Samson ($\overline{N}$)	I. Delilah unhappy Samson unforthcoming (2, 4)/(1, 4)	II. Delilah happy Samson forthcoming (4, 2)
	Nag Samson (N)	IV. Delilah frustrated Samson harassed (1, 1)/(2, 1)	III. Delilah persuasive Samson reluctant (3, 3)

Key
(x, y) = (payoff to Delilah, payoff to Samson).
4 = best, 3 = next best, 2 = next worst, 1 = worst.
Nash equilibrium underscored.
Nonmyopic equilibria (NMEs) highlighted in blue.
Arrows indicate progression of states from (2, 4)/(1, 4) to NME of (3, 3).

The payoff matrix of the game we posit that Samson played with Delilah is shown in Figure 16.3. (Ignore for now the payoffs to the right of the slashes in the first column of Figure 16.3.) Samson's desire having been kindled, Delilah could trade on it either by nagging Samson for the secret of his strength (N) or not nagging him ($\overline{N}$) and hoping it would come out anyway. Samson, in turn, could either tell (T) the secret of his strength or not tell it ($\overline{T}$). Consider the consequences of each pair of strategy choices, starting from the upper left-hand state and moving in a clockwise direction:

I. *Delilah unhappy, Samson unforthcoming: (2, 4).* The next-worst state for Delilah, because Samson withholds his secret, though she is not frustrated in an unsuccessful attempt to obtain it; the best state for Samson, because he keeps his secret and is not harassed.

II. *Delilah happy, Samson forthcoming: (4, 2).* The best state for Delilah, because she learns Samson's secret without making a pest of herself; the next-worst state for Samson, because he gives away his secret without being under duress.

III. *Delilah persuasive, Samson reluctant: (3, 3).* The next-best state for both players, because though Delilah would prefer not to nag (if Samson tells) and Samson would prefer not to succumb (if Delilah does not nag), Delilah gets her way when Samson tells; and Samson, under duress, has a respectable reason (i.e., Delilah's nagging) for telling.

IV. *Delilah frustrated, Samson harassed: (1, 1).* The worst state for both players, because Samson does not get peace of mind, and Delilah is frustrated in her effort to learn Samson's secret.

The Figure 16.3 game just described is Success and starts in state I at (2, 4), when Delilah chooses $\overline{N}$ and Samson chooses $\overline{T}$ during their period of getting acquainted. These strategy choices are consistent with Delilah's choosing her dominant strategy, and Samson his best response to this strategy, giving the Nash equilibrium of (2, 4), according to standard game theory.

The standard theory offers no explanation of why the players would ever move to a nonequilibrium outcome. But this is precisely what they do. Delilah switches to *N*, putting the players in state IV at (1, 1), and Samson responds with *T*, leading to state III at (3, 3), neither of which is a Nash equilibrium. By contrast, TOM leads to a different prediction when the initial state is I at (2, 4). From this state, TOM predicts the outcome to be (3, 3). Although (2, 4) and (4, 2) are also NMEs, they do not arise unless play starts elsewhere.

Thus, TOM leads to a unique prediction of (3, 3) if play starts in state I, which is the actual outcome of the story. (Arguably, the outcome is state II, after Delilah has stopped nagging, which is rational because Samson cannot respond by hiding his secret once it is already out; we discuss such infeasible moves later.) Doubtless, Samson did not anticipate having his eyes gouged out and being derided as a fool before the Philistines when he responded to Delilah's nagging by revealing his secret. On the other hand, because he was later able to kill thousands of Philistines at the same time that he ended his own humiliation, the resolution of this story can plausibly be seen as next best for Samson.

Although he surrendered his secret to the treacherous Delilah, Samson apparently never won her love, which seems to be the thing he most wanted. In fact, Delilah's decisive argument in coaxing the truth from Samson was that because he did not confide in her, he did not love her. What better way was there for Samson to counter this contention, and prove his love, than to comply with her request, even if it meant courting not just Delilah but disaster itself?

As for Delilah's preferences, it is hard to quarrel with the assumption that her two best states were associated with Samson's choice of *T*. What is less certain, however, is the order of preferences she held for her two worst states. Contrary to the Success representation in Figure 16.3, Delilah might have preferred to nag Samson than not had he in the end denied her his secret. For even though she would have failed to discover Samson's secret, Delilah would perhaps have felt less badly after having tried than if she had made no effort at all.

If this is the case, then 2 and 1 for Delilah would be interchanged in Success, giving a different game matrix. But, like Success, it is possible to show that this game, which we call Variation and whose payoffs are to the right of the slashes in the first column of Figure 16.3, yields (3, 3) as the NME when play starts in state I at (1, 4). Hence, this reordering of Delilah's preferences would not affect the rational outcome, according to TOM: thinking nonmyopically, Delilah would still switch to *N*, and Samson in turn would switch to *T*. Thus, even if Samson had only incomplete information about how Delilah ordered her two worst states, his play would not be affected. ◆

To summarize, where players start in a game, including the unique dominant-strategy Nash equilibria, may not be where they end up, according to TOM. However, the restriction that rules 5 and 6 place on the ability of players to cycle in games may not always be descriptive of a situation. Therefore, we next consider alternatives to these rationality rules (namely, rules 5′ and 6′) that permit cycling around a matrix and, in addition, allow players to terminate cycles through the exercise of "moving power."

CYCLIC GAMES AND MOVING POWER

Recall that the rules of play of TOM say nothing about what causes a game to end, only when. Rule 5, which forbids continual cycling, provides one answer. But this ban on cycling may not be realistic, as many protracted conflicts (e.g., the Arab-Israeli conflict), in which the protagonists have revisited the past again and again, make unmistakably clear.

To capture the cyclic aspect of certain conflicts and give players the ability to make choices in which they repeat themselves (why they may want to do so will be considered shortly), we define a subclass of the 78 games in which cycling is possible by defining a subclass of games in which it is *not*. Rule 5′ provides a sufficient condition for cycling not to occur:

5′. If at any state in the move-countermove process a player whose turn it is to move next receives his or her best payoff (i.e., 4), that player will not move from this state.

Rule 5′, in fact, prevents cycling in 42 of the 78 distinct 2 × 2 games. We call the remaining 36 games cyclic.

A **cyclic game** is one in which, when the game cycles either clockwise or counterclockwise, neither player ever receives a best payoff (4) when it is his or her turn to move next.

Consider the circumstances under which players, who know not only their own payoffs but also the payoffs of their opponents, would have an incentive to cycle to try to outlast an opponent. By "outlasting" we mean that one (stronger) player can force the other (weaker) player to stop the move-countermove process at a state where the weaker player has the next move. Forcing stoppage at such a state involves the exercise of moving power.

If one player (*P1*) has **moving power,** he or she can force the other player (*P2*) to stop, in the process of cycling, at one of the two states at which *P2* has the next move.

The state at which *P2* will stop is that which *P2* prefers. Because of the change of rules that now allow for cycling, moving power is very different from order power. For example, it may not always benefit the player who has it compared with the player who does not have it.

Rule 5′ specified what players would *not* do, namely, move from a best (4) state when it was their turn to move. However, this rule did not say anything about *where* cycling would stop, which the exercise of moving power determines by enabling the player who possesses it to break the cycle of moves.

Rule 6′ ensures there will be termination.

6′. At some point in the cycling, *P2* must stop.

This is not to say that *P1* will always exercise his or her moving power. In some games, as we shall see, it is rational for *P1* to terminate play, even though *P1* can always force *P2* to stop first.

EXAMPLE ▶ ***Moving Power in Success and Alternative***

Moving power is **effective** in a cyclic game if the outcome that a player can implement with this power is better for him or her than the outcome that the other player can implement. To illustrate when moving power is effective, consider again the game of Success. The arrows in Figure 16.4a (ignore for now the distinction between the single and double arrows) illustrate the cyclicity of Success in a counterclockwise direction: starting at the upper right state, *C* benefits by moving from $(4, 2)$ to $(2, 4)$; *R* does not benefit by moving from $(2, 4)$ to $(1, 1)$ but departs from a 2, not a 4, state (so does not violate rule 5′); *C* benefits by moving from $(1, 1)$ to $(3, 3)$; and *R* benefits by moving from $(3, 3)$ to $(4, 2)$. Because no player, when it is his or her turn to move, ever departs from his or her best (i.e., 4) state, Success is cyclic.

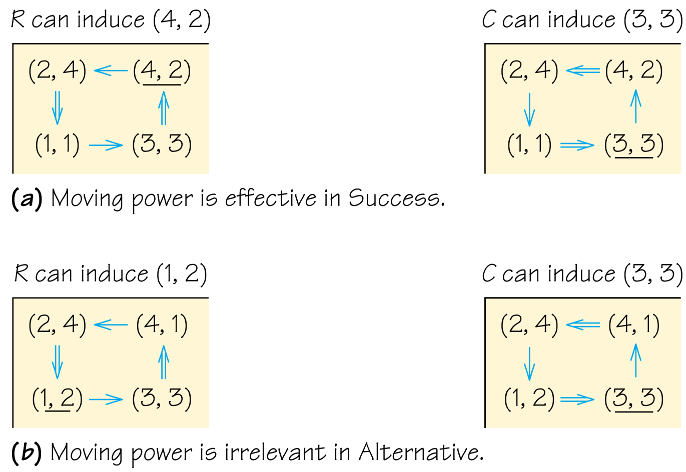


FIGURE 16.4 Moving power in two cyclic games.

Key
(x, y) = (payoff to R, payoff to C).
4 = best, 3 = next best, 2 = next worst, 1 = worst.
Double arrows indicate moves of player with moving power.
Single arrows indicate moves of player without moving power.
Underscored state indicates the outcome player with moving power can induce

To show what outcome *R* can implement if he has moving power, which might be thought of as greater stamina or endurance—in the sense that he can continue moving when the other must eventually stop—let his moves (vertical, as illustrated on the left side of Figure 16.4a) be represented by double arrows. *C*, whose (horizontal) moves are represented by single arrows, must stop in the cycling at either $(1,1)$ or $(4,2)$, from where her single arrows emanate that indicate she has the next move. Since she would prefer to stop at $(4,2)$ rather than $(1,1)$, *R* can implement his best outcome of $(4,2)$ if he has moving power. On the other hand, if *C* has moving power (right side of Figure 16.4a), she can force *R* to stop at either $(2,4)$ or $(3,3)$, whence his single arrows emanate that indicate he has the next move. Since *R* would prefer to stop at $(3,3)$ rather than $(2,4)$, *C* can implement her next-best outcome of $(3,3)$ if she has moving power. Thus, the possession of moving power benefits the player who possesses it—compared with the other player's possession of it—so it is effective in Success.

This is not the case in a game in which 1 and 2 are interchanged for *C* in Success, which we will refer to as the game Alternative. Applying the same reasoning to Alternative, shown in Figure 16.4b, we see that *R* can implement only $(1,2)$—C prefers this to $(4,1)$, the other state where she moves from—but C can implement $(3,3)$—R prefers this to $(2,4)$, the other state he moves from. Since R also prefers $(3,3)$ to $(1,2)$, moving power is not effective: R cannot implement a better outcome when he has it than when C has it. Instead, moving power is irrelevant in Alternative, because it would be in R's interest to stop at $(3,3)$, even if he has moving power, rather than to force C to stop at $(1,2)$.

More generally, moving power is **irrelevant** when the outcome one player can implement is better for both. Unlike Variation (see Figure 16.3), in which a switch of 1 and 2 by one player (R, or Delilah) does not change the outcome of Success when play commences at the upper left-hand state—(3,3) remains the NME from this state—a switch of 1 and 2 by C, yielding Alternative, does change the effects of moving power in Success.

Moving power obviously constrains the freedom of choice of the player who does not possess it, because the player who possesses it can force the other player to stop. As we have seen in Alternative, R has no reason to exercise this kind of power, preferring to stop himself at $(3,3)$ in the move-countermove process, which renders his possession of moving power in this game irrelevant. On the other hand, the possession of moving power in Success by either R or C helps him or her obtain a better outcome than if the other player possessed it, rendering moving power effective in this game. ◆

In many real-world conflicts, there may be no clear recognition of which, if either, player has moving power. In fact, there may be a good deal of misinformation. For example, if both players believe they can hold out longer, cycling is likely to persist until one player succeeds in demonstrating greater strength or both players are exhausted by the repeated cycling. The latter may well have occurred in the Egyptian-Israeli conflict between 1948 and 1979, as well as similar recent conflicts (e.g., in South Africa and Northern Ireland). Although Egypt and Israel fought five wars in this period (1948, 1956, 1967, 1969–70, and 1973), at great cost to both sides, it still required considerable pressure from the United States to achieve the 1978 Camp David accords that paved the way for the signing of a peace treaty between Egypt and Israel in 1979. Likewise, similar outside pressure was exerted in the South African and Northern Ireland conflicts to induce settlements, as it was in the former Yugoslavia to induce the warring sides to sign a peace treaty in November 1995 after four years of bitter conflict that cost some 250,000 lives.

We offer a note of caution in interpreting cycling and the exercise of moving power. While Egypt and Israel cycled in and out of war for a generation, cycling in games like that of Samson and Delilah may, though theoretically possible, not be feasible. In particular, although Delilah was readily able to switch from being a seductress to being a nag, and Samson from being mum about the secret of his strength to revealing it, once Samson's secret was out, he

Israel's late Prime Minister Yitzhak Rabin (*left)* with PLO Chairman Yasir Arafat. Egyptian President Hosni Mubarak (*center*) hosted the historic PLO-Israeli peace accord in May 1994.

could not retract it, especially in light of the fact that Delilah tested every explanation, true or false, that he gave for his strength.

Thus, cycling in the Samson and Delilah game is not **feasible:** it contradicts what a reasonable interpretation of the strategies in this game permit. Whereas it is permissible for Samson to move from $\overline{T}$ to T (see Figure 16.3), a reverse switch is hard to entertain in the context of the story, though not necessarily in other interpretations of Success. For example, consider the original interpretation of this game given in Figure 16.1, in which $(3, 3)$ is "Compromise." Starting from this state, one can readily imagine situations in which C switches from t_2 to t_1, plunging the players into "Disaster," hoping that this move will drive R to switch from s_2 to s_1, yielding "C succeeds," with payoffs of $(2, 4)$ to the players.

In using TOM to model a strategic situation, the analyst must be sensitive to what strategy changes are feasible and infeasible. Also, even if moves are rational according to the rules, the state to which a game moves may, if feasible, be one from which a player cannot emerge to move on. As a case in point, assume C, starting at $(3, 3)$ in the Figure 16.1 game, switches from t_2 to t_1, changing the state to $(1, 1)$. Given that this move is feasible (which we assumed was not the case in the Samson and Delilah story), can R subsequently move on to $(2, 4)$?

We suggest that the answer to this question depends on what interpretation one gives to "Disaster" at $(1, 1)$. If this state means, say, nuclear war, nobody may survive in order to be able to "move on," making $(1, 1)$, figuratively speaking, a black hole. On the other hand, if *C*'s move to (1, 1) creates a severe crisis, like the Cuban missile crisis of 1962, *R* may be able to respond, as the Soviet Union did, by withdrawing its nuclear weapons from Cuba, which abated this crisis.

To conclude, we have shown that, depending on the rules, the starting point, or who possesses order or moving power, can matter. Nevertheless, not all rational moves may be feasible, either because changing strategies or moving from certain states violates the interpretation of a game.

RETURN TO PRISONERS' DILEMMA AND CHICKEN

In Figure 16.5 we show Prisoners' Dilemma and Chicken, in which $C =$ cooperation and $\overline{C} =$ noncooperation. As we did in the game of Success in Figure 16.2, we also show in brackets the anticipation games for each of these games, giving the NMEs into which each state goes, according to rules 1–6. Note that Prisoners' Dilemma has two NMEs: starting at (2, 2), the players would remain stuck at Conflict; but from (4, 1) or (1, 4) they would go into Compromise, or stay there if they start out at (3, 3). Hence, the dilemma arises only if the status quo state is Conflict. This is a somewhat more auspicious view of this game, at least for nonmyopic players, than the standard game theory solution (i.e., Conflict) suggested, unless there is repeated play that enables the players to use a strategy like tit-for-tat to induce cooperation (Chapter 15).

TOM also indicates that Compromise is an NME in Chicken if the players start out there. Order power comes into play if the players commence at either (4, 2) or (2, 4), and one player is advantaged: if the advantaged player has order power at these states, he or she can ensure (3, 3) as the NME. This consequence of TOM is interesting, because it suggests that the advantaged player in each of these states should move to his or her *less* preferred state of (3, 3), lest the other player induce his or her best outcome by moving the process through (1, 1) to the NME of (2, 4) or (4, 2), respectively, from (4, 2) and (2, 4). Finally, if the initial state is Disaster, then the possession of order power enables the player who possesses it to force the other player to move away first (i.e., "chicken out"), leading to the nonmover's best state (i.e., 4).

In real-life games that can be modeled by Prisoners' Dilemma or Chicken, Compromise has sometimes been achieved, it seems, because the players were nonmyopic. For example, even before the collapse of the Soviet Union in the late 1980s, the superpowers agreed to certain limitations in their arms race, which was frequently modeled as a Prisoners' Dilemma. In the Cuban missile

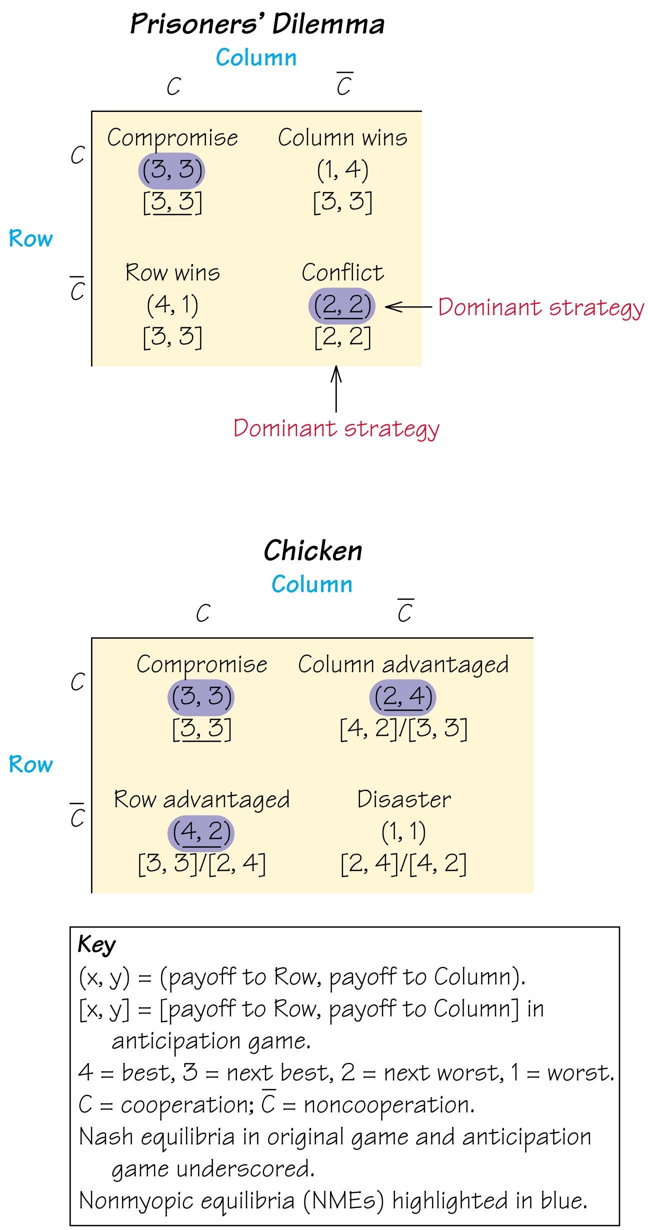


FIGURE 16.5 Prisoners' Dilemma and Chicken.

crisis of 1962, which was often modeled as a game of Chicken (notice that the payoffs to the players in the four states are the same as those in Success, but the configuration is different), a compromise was reached when the Soviets withdrew their missiles from Cuba and the United States promised not to invade the island in the future (an invasion of Cuba, with U.S. support, had been unsuccessfully attempted in 1961 at the infamous Bay of Pigs).

Assume rules 5′ and 6′ are operative. Then moving power is not defined in either Prisoners' Dilemma or Chicken because these games are not cyclic: whether the players cycle clockwise or counterclockwise, a state is reached in which a player whose turn it is to move next receives his or her best payoff and so would not move, which precludes cycling (rule 5′). In Prisoners' Dilemma, for example, if the move-countermove process is clockwise, the column player at some point would have to move from (4, 1) to (3, 3), whereas if it is counterclockwise, the row player would have to move from (1, 4) to (3, 3). Hence, while order power is applicable in Chicken, moving power is defined in neither this game nor Prisoners' Dilemma.

LARGER GAMES

So far we have applied TOM only to 2 × 2 games. But in this concluding section we turn to a game with more than two players and in which each player has more than two strategies.

EXAMPLE ▶ ***A Truel***

A **truel** is like a duel, except that there are three players. Each player can either fire, or not fire, his or her gun at either of the other two players. We assume the goal of each player is, first, to survive and, second, to survive with as few other players as possible. Each player has one bullet and is a perfect shot, and no communication (e.g., to pick out a common target) that results in a binding agreement with other players is allowed, making the game noncooperative. We will discuss the answers that standard game theory, on the one hand, and TOM, on the other, give to what it is optimal for the players to do in the truel.

According to standard game theory, *at the start of play, each player fires at one of the other two players, killing that player.*

Why will the players all fire at each other? Because their own survival does not depend an iota on what they do. Since they cannot affect what happens to themselves but can only affect how many others survive (the fewer the better, according to the postulated secondary goal), they should all blaze away at each other. (Even if the rules of the play permitted shooting oneself, the primary goal of survival would preclude committing suicide.) In fact, the players all have dominant strategies to shoot at each other, because whether or not a player survives—we will discuss shortly the probabilities of doing so—he or she does at least as well shooting an opponent.

The game, and optimal strategies in it, would change if (i) the players were allowed more options, such as to fire in the air and thereby disarm themselves, or (ii) they did not have to choose simultaneously, and a particular order of play were specified. Thus, if the order of play were *A*, followed by *B* and *C* choosing simultaneously, followed by any player with a bullet remaining choosing, then *A* would fire in the air and *B* and *C* would subsequently shoot each other. (*A* is no threat to *B* or *C*, so neither of the latter will fire at *A*; on the other hand, if *B* or *C* did not fire immediately at the other, each might not survive to get in the last shot, so they both fire.) Thus, *A* will be the sole survivor. In 1992, a modified version of this scenario was played out in late-night television programming among the three major television networks, with ABC's effectively going first with "Nightline," its well-established news program, and CBS and NBC dueling on which host, David Letterman or Jay Leno, to choose for their entertainment shows. Regardless of their ultimate choices, ABC "won" when CBS and NBC were forced to divide the entertainment audience.

To return to the original game, the players' strategies of all firing have two possible consequences: either one player survives (even if two players fire at the same person, the third must fire at one of them, leaving only one survivor), or no player survives (if each player fires at a different person). In either event, there is no guarantee of survival. In fact, if each player has an equal probability of firing at one of the two other players, the probability that any particular player will survive is only .25.

The reason is that if the three players are *A*, *B*, and *C*, *A* will be killed when either *B* fires at him or her, *C* does, or both do. The only circumstance in which *A* will survive is if *B* and *C* fire at each other, which gives *A* one chance in four. Although this calculation implies that one of *A*, *B*, or *C* will survive with probability .75 if all outcomes are equally likely, more meaningful for each player is the low .25 individual probability of survival.

According to TOM, *no player will fire at any other, so all will survive.*

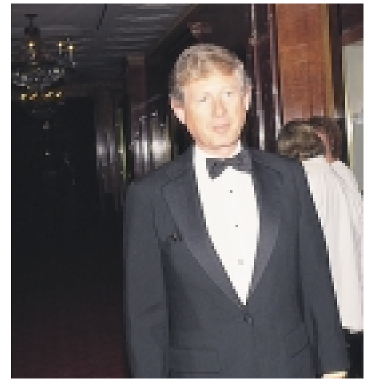
At the start of the truel, all the players are alive, which satisfies their primary goal of survival, though not their secondary goal of surviving with as few others as possible. Now assume that *A* contemplates shooting *B*, thereby reducing the number of survivors. But looking ahead, *A* knows that by firing first and killing *B*, he or she will be defenseless and be immediately shot by *C*, who will then be the sole survivor.

It is in *A*'s interest, therefore, not to shoot anybody at the start, and the same logic applies to each of the other players. Hence, everybody will survive, which is a happier outcome than that given by game theory's answer, in which everyone's primary goal is not satisfied—or, quantitatively speaking, satisfied only 25% of the time. ◆

The purpose of TOM, however, is not to produce "happier" outcomes but to provide a plausible model of a strategic situation that mimics what people might actually think and do in such a situation. We believe that the players in the truel, artificial as this kind of shoot-out may seem, would be motivated

Jay Leno

David Letterman



Ted Koppel

to think ahead, given the dire consequences of their actions. Following the reasoning of TOM, therefore, they would hold their fire, knowing that if one fired first, he or she would be the next target.

In Figure 16.6 we show this logic somewhat more formally with a **game tree,** in which A has three strategies, as indicated by the three branches that sprout from A: not shoot ($\overline{S}$), shoot B ($S \rightarrow B$), or shoot C ($S \rightarrow C$). The latter two branches, in turn, give survivors C and B, respectively, two strategies: not shoot ($\overline{S}$) or shoot A ($S \rightarrow A$).

We assume that the players rank the outcomes as follows, which is consistent with their primary and secondary goals: 4 = best (lone survivor), 3 = next best (survivor with one other), 2 = next worst (survivor with two others), and 1 = worst (nonsurvivor). These payoffs are given for ordered triples (A, B, C); thus (3, 3, 1) indicates the next-best payoffs for A and B and the worst payoff for C.

Note that play necessarily terminates when there is only one survivor, as is the case at (1, 1, 4) and (1, 4, 1). To keep the tree simple, we assume that play also terminates when either A initially or B or C subsequently chooses $\overline{S}$, giving outcomes of (2, 2, 2), (3, 3, 1), and (3, 1, 3), respectively. Of course, we could allow the two or three surviving players in the latter cases to make subsequent choices in an extended game tree, but this example is meant only to illustrate the analysis of a game tree, not be the definitive statement on truel possibilities (more will be explored in the exercises).

As in 2 × 2 strategic-form games using TOM, we work backwards in extensive-form games, starting the analysis at the bottom of the tree. (By "bottom" we mean where play terminates; because this is where the tree branches out, the tree looks upside down in Figure 16.6.) Thus, because C prefers (1, 1, 4) to (3, 1, 3), we indicate that C would not choose $\overline{S}$ by "cutting" this branch with a scissors; similarly, B would not choose $\overline{S}$. Cutting a branch here is analogous to the blockage of a move in a 2 × 2 game.

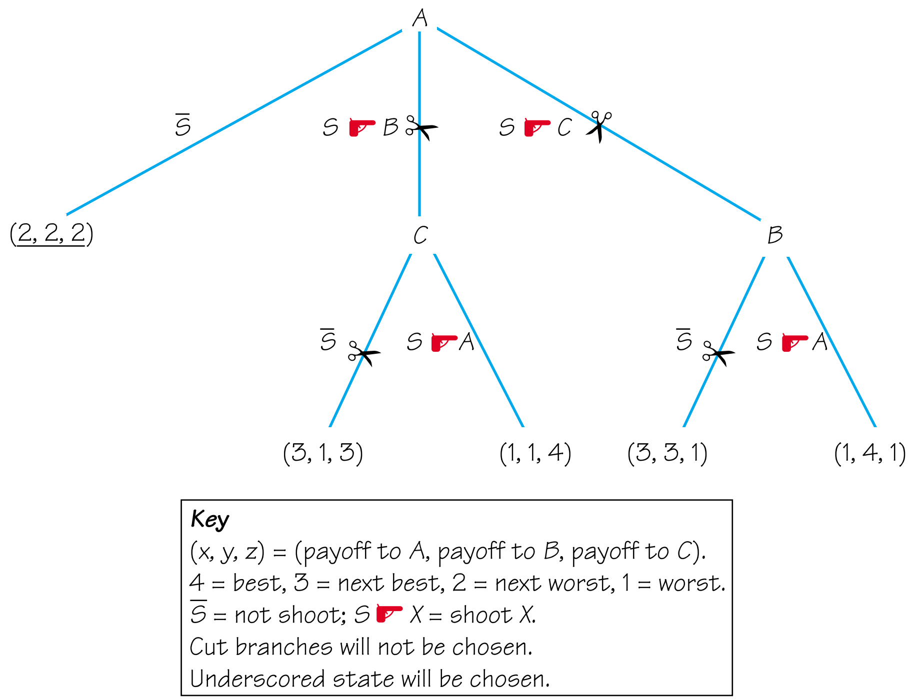


FIGURE 16.6 A game tree of a truel.

Moving up to the next level, A would know that if he or she chose $S \rightarrow B$, $(1, 1, 4)$ would be the outcome; if he or she chose $S \rightarrow C$, $(1, 4, 1)$ would be the outcome, making each, in our earlier terminology, the survivor from the bottom level. Choosing between these two outcomes and $(2, 2, 2)$, A would prefer the latter, so we cut the two branches, $S \rightarrow B$ and $S \rightarrow C$. Hence, A would choose $\overline{S}$, terminating play with nobody's shooting anybody else.

This, of course, is the conclusion we reached earlier, based on the reasoning that if A shot either B or C, he or she would end up dead, too. Because we could allow each player, like A, to choose among his or her three initial strategies in a $3 \times 3 \times 3$ game, and subsequently make moves and countermoves from the initial state (if feasible), the foregoing analysis applies to all players.

Underlying the completely different answers of game theory and TOM is a change in the rules of play that TOM introduces, namely, that players do not have to fire simultaneously at the start of play in the $3 \times 3 \times 3$ game. In standard game theory, however, if the players do not fire simultaneously, an order of play is posited, in which case the player to move first—and then the later players—would not fire, given that play continues until all bullets are expended or nobody chooses to fire.

To be sure, this is the same answer as that given by TOM. The point we stress here is that the standard theory does not raise the question of which order of play the players—if, thinking ahead, they could make this choice—would adopt, given their goals.

TOM, by contrast, leaves open the order of play by asking of each player: given your present situation (all alive), and the situation you anticipate will ensue if you fire first, should you do so? Because each player prefers living to the state he or she would bring about by being the first to shoot (certain death), none shoots. This analysis suggests that truels might be more effective than duels in preventing the outbreak of conflict.

We will not try to develop this argument into a more general model. The main point is that TOM introduces into a payoff matrix a look-ahead approach to game-theoretic analysis, which requires radical changes in the usual rules of play. In particular, these changes require the comparison of the past or present (initial state) with the future (final state)—perhaps several steps ahead—to which the moves and countermoves may transport the players. TOM also allows for the exercise of power (order or moving) when there is an asymmetry in the abilities of the players, which, as we saw, can also affect the outcome.

Unfortunately, there is no surefire way to determine which set of rules, or what kinds of power, are most applicable in a given situation. Our aim in this chapter has been to show that there are alternatives to standard game theory, particularly in looking at the *processes* by which outcomes are chosen, thereby making the analysis more dynamic.

REVIEW VOCABULARY

Anticipation game A game, described by a payoff matrix, whose entries, which are given in brackets, are the nonmyopic equilibria (NMEs) into which each state of the original game goes.

Asymmetric game A game in which the row player's strategies do not lead to the same payoffs as the column player's strategies.

Backward induction A reasoning process in which players, working backward from the last possible move in a game, anticipate each other's rational choices.

Blockage Occurs when it is not rational, based on backward induction, for a player to move from a state.

Complete information Each player knows the rules of the game, the preferences of every player for all possible states, and which, if either, player has order or moving power.

Cyclic game A 2×2 ordinal game in which, when the game cycles either clockwise or counterclockwise, neither player ever receives a best payoff when it is his or her turn to move next.

Effective moving power Moving power is effective when possessing it induces a better outcome for a player than when the other player possesses it.

Feasibility A move is feasible if it can plausibly be interpreted as possible in the situation being modeled.

Final state A final state is the state induced after all rational moves and countermoves (if any) from the initial state have been made, making it the outcome of the game.

Game A game is the totality of the rules that describe it.

Game tree A symbolic tree, based on the rules of play of a game, in which the vertices, or nodes, of the tree represent choice points, and the branches represent alternative courses of action that the players can select.

Indeterminate state A state in which the outcome induced depends on which player moves first (in which case order power is effective).

Initial state The state from which play commences.

Irrelevant moving power Moving power is irrelevant when the outcome induced by one player is better for both players than the outcome that the other player can induce.

Move A player's switch from one strategy to another in the payoff matrix of a strategic-form game.

Moving power In a cyclic game, the ability of one player to continue moving when the other player must eventually stop; a player exercises moving power in order to try to implement a preferred outcome.

Nonmyopic equilibrium (NME) A state to which rational players would move (or not move), anticipating all possible rational moves and countermoves from some initial state.

Order power The ability of a player to determine the order of moves in which the players depart from an indeterminate initial state in order to ensure a preferred outcome.

Outcome The final state of a game, from which no player would choose to move and at which the players receive their payoffs.

Precedence Occurs when the outcome induced by the player who moves overrides the outcome induced by the player who stays.

Rational choice A choice that leads to a preferred outcome, based on the rules of the game.

Rules of play Describe the possible choices of the players at each stage of play.

State An entry in a payoff matrix from which the players can move. Play of a game starts at an initial state and terminates at a final state, or outcome.

Stoppage Occurs when blockage occurs for the first time from some initial state.

Survivor The state that is selected at any stage as the result of backward induction.

Termination rule Prescribes that play will terminate after one complete cycle.

Theory of moves (TOM) A dynamic theory that describes optimal strategic choices in games in which the players, thinking ahead, can make moves and countermoves from an initial state.

Truel The analogue of a duel, in which each of three players can fire or not fire his or her gun at either of the other two players.

Two-sidedness rule Describes how a player determines whether or not to move from a state, based on the other player's rational choices as well as his or her own.

SUGGESTED READINGS

BRAMS, STEVEN J. *Superior Beings: If They Exist, How Would We Know? Game-Theoretic Implications of Omniscience, Omnipotence, Immortality, and Incomprehensibility,* Springer-Verlag, New York, 1983. An application of game theory, and an early version of the theory of moves, to questions in the philosophy of religion, such as ascertaining the existence of a superior being and explaining the problem of evil in the world, based on an analysis of 2×2 ordinal games.

BRAMS, STEVEN J. *Theory of Moves,* Cambridge University Press, Cambridge, 1994. The main source of the theory in this chapter; it also contains diverse applications, ranging from literature to theology.

RAPOPORT, ANATOL, MELVIN GUYER, AND DAVID GORDON. *The 2×2 Game,* University of Michigan Press, Ann Arbor, 1976. A comprehensive review and classification of the 78 distinct 2×2 strict ordinal games, including the means by which the number 78 was determined.

TAYLOR, ALAN D. *Mathematics and Politics: Strategy, Voting, Power, and Proof,* Springer-Verlag, New York, 1995. A mathematics textbook in which game theory, theory of moves, and social choice theory are used to model power, voting, and conflict and escalation processes.

EXERCISES

▲ *Optional.* ■ *Advanced.* ◆ *Discussion.*

Game Theory Revisited

◆ 1. Define an ordinal game to be one of total conflict in which the best state (4) for one player is the worst state (1) for the other player, and the next-best state (3) for one player is the next-worst state (2) for the other. What relationship do these games have to zero-sum games?

2. The three different total-conflict games with (4, 1), (1, 4), (3, 2), and (2, 3) as states are as follows:

(2, 3)	(4, 1)
(1, 4)	(3, 2)

(3, 2)	(4, 1)
(2, 3)	(1, 4)

(2, 3)	(4, 1)
(3, 2)	(1, 4)

Find the dominant strategies and Nash equilibria in each (if they exist).

Rules of TOM

3. Show that two of the total-conflict games in Exercise 2 have one NME and the third has two NMEs. What, if any, relationship is there between the Nash equilibria and NMEs in the total-conflict games?

4. Both Success and Chicken have as states (3, 3), (2, 4), (4, 2), and (1, 1), but these two games have very different properties. Find one other *symmetric* 2 × 2 game (like Chicken) with these four states, and one other *asymmetric* game (like Success). Are there any other games that have different configurations of these four states?

5. Show that the other symmetric game in Exersise 4 has one NME, and the other asymmetric game has two NMEs. Which of the NMEs in these games coincide with Nash equilibria?

Interpreting TOM

◆ 6. An alternative game called Variation was suggested for Delilah's preferences in Figure 16.3. Does Variation seem more reasonable to you than Success as a model of the Samson and Delilah story? Why?

7. The following "mugging game" has been proposed to model the conflict between a mugger and a victim:

		Mugger	
		Use force (F)	Don't use force ($\bar{F}$)
Victim	Resist (R)	I. Fight (2, 2) [3, 4]	II. Mugger fails (4, 1) [2, 2]
	Don't resist ($\bar{R}$)	II. Involuntary submission (1, 3) [3, 4]	III. Voluntary submission (3, 4) [3, 4]

◆ Do the ordinal payoffs in this game seem plausible to you? If not, propose a more plausible alternative game.

8. Verify that the NMEs in the mugging game from each state are those shown in its anticipation game. If play of the game starts at state II, does it seem reasonable that, because the victim would *not* move from (4, 1) according to TOM, the mugger would move to (2, 2)?

◆ 9. Assume that the victim anticipates the (2, 2) outcome in the mugging game, starting at state II. Argue that it would be reasonable for him or her to take the initiative and move to (3, 4), making Voluntary submission rather than Fight the outcome from (4, 1). *Note:* In fact, TOM postulates a "two-sidedness convention" to cover this and six other 2×2 games. If one player (the victim in the mugging game) can induce a better state for *both* players [at (3, 4)] by moving rather than staying—instead of forcing the other player (the mugger) to move [to (2, 2)]—then that player (the victim) will move, even if he or she would otherwise prefer to stay [at (4, 1)], to induce a better outcome [(3, 4) rather than (2, 2)]. When amended by the two-sided convention, TOM says that (3, 4) rather than (2, 2) is the NME from (4, 1), making the mugging game a one-NME rather than a two-NME game.

Cyclic Games and Moving Power

10. Show that the mugging game is cyclic in a counterclockwise direction. Is moving power effective or irrelevant in this game?

◆ 11. It was shown that moving power is irrelevant in Alternative (Figure 16-4b), leading to (3, 3), whichever player possessed it. But since (2, 4) as well as (3, 3) is an NME in Alternative, it would appear *not* to be in *C*'s interest to exercise her moving power in order to try to get her preferred NME of (2, 4). Does this seem paradoxical?

■ 12. If a 2×2 strict ordinal game is cyclic, show that it can cycle in only one direction (clockwise or counterclockwise, but not both).

Return to Prisoners' Dilemma and Chicken

13. If (2, 4) or (4, 2) is the initial state in Chicken, show by backward induction that the player receiving 4 would not move, whereas the player receiving 2 would move the process from (2, 4) to (4, 2) or from (4, 2) to (2, 4), making (4, 2) and (2, 4) the NMEs from (2, 4) and (4, 2), respectively. Knowing this, would the 4-player at (2, 4) or (4, 2) have an incentive to "beat the 2-player to the punch" in order to move the process immediately to (3, 3)—instead of not moving and suffering a 2-outcome if the 2-player moved first? Does this reasoning agree with backward induction?

Larger Games

14. Extend the game tree of the truel in Figure 16.6 to allow the additional possibility that if *A* does not shoot initially, then *B* has the choice of shooting or not shooting *C*. Will *A*, in fact, not shoot initially, and will *B* then shoot *C*?

15. Extend the game tree in Exercise 14 to still another level to allow the possibility that if *A* does not shoot initially, and *B* shoots *C*, then *A* has the choice of shooting or not shooting *B*. What will happen in this case?

16. Change Exercise 15 to allow for the possibility that if *A* does not shoot initially, and *B* shoots *or does not shoot C*, then *A* has the choice of shooting or not shooting *B*. What will happen in this case?

◆ 17. What general conclusions would you draw in light of your answers to Exercises 14, 15, and 16?

Extensions to Threats

18. Say that a player has a *threat* if he or she can choose a strategy that leads to the two worst states (1 and 2) of the other player. Show that both players have threats in Prisoners' Dilemma and Chicken. Which, if either, player has a threat in Success, Variation, Alternative, and the mugging game?

Additional Exercises

19. Show that there are four different "almost" total-conflict games with states $(4, 1)$, $(1, 4)$, $(3, 3)$, $(2, 2)$. (If you should find more than four games, check whether interchanging players, their strategies, or both can transform one game into another, so that they are not strategically different.) Find the dominant strategies and Nash equilibria in each (if they exist).

▲ 20. Find the NMEs in the "almost" total-conflict games in Exercise 19.

◆ 21. In Shakespeare's *Macbeth,* Lady Macbeth badgers and cajoles Macbeth into murdering King Duncan. Argue that Success can be used to model their conflict—or show that another game better mirrors the preferences of the two players, and analyze the latter game.

22. Show that if the players commenced play from the upper left state of either Success or Variation, it would make no difference which game was the "true" game—the players in both games would move to the lower right state of $(3, 3)$. Does TOM make the same prediction if the initial state is $(1, 1)$ in Success and $(2, 1)$ in the Variation?

23. Show that Variation is cyclic and that moving power is effective in this game. Is it plausible to assume that moving power would be exercised in Variation if this game is used to model Samson and Delilah?

■ 24. Show that no symmetric 2×2 strict ordinal game is cyclic (but *not* by checking all possible symmetric games—there are a total of 12).

25. Indicate the game that results if the row player's preferences are those in Prisoners' Dilemma, and the column player's preferences are those in Chicken. Show that this game has only one NME, which favors the player with Prisoners' Dilemma preferences. Does it seem strange to you that this game has only one NME, whereas Prisoners' Dilemma has two and Chicken has three?

■ 26. If a fourth player is added to the original truel, show that each player will have an incentive to shoot another (as in a duel).

◆ 27. If the players in Prisoners' Dilemma and Chicken say that they will carry out their threats if and only if the other player does not choose his or her cooperative strategy initially, and these threats are believed by both players, will (3, 3) be the outcome in both games? Might threats like this help in ameliorating, rather than aggravating, conflict in such games?

WRITING PROJECTS

1 ▸ Quentin Tarantino's films *Reservoir Dogs* (1992) and *Pulp Fiction* (1994) both have truels, but the choices that the characters make in each are completely different. Does TOM offer any insight into why?

2 ▸ Model a conflict in a Bible story, a work of fiction, or a real-life situation as a 2×2 strict ordinal game, comparing the results predicted by game theory and by TOM. Do you think these theories take proper account of the feelings or emotions of the characters, especially when their actions appear "irrational"?

3 ▸ Strikes and threatened strikes in different professional sports have become common, even as the salaries of professional athletes have skyrocketed. Explain why this has happened, using either standard game theory or TOM to model the conflict between the athletes and the team owners. Does either theory suggest how strikes might be avoided and stable settlements reached?