

# GAMES OF STRATEGY

THIRD EDITION



*Avinash Dixit*

Princeton University

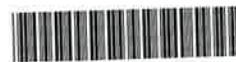
*Susan Skeath*

Wellesley College

*David Reiley*

University of Arizona and Yahoo! Research

DEDALUS - Acervo - FFLCH



20900107647

SBD-FFLCH-USP



302847



W. W. Norton & Company

New York • London

## Strategic Moves

**A** GAME IS specified by the choices or moves available to the players, the order, if any, in which they make those moves, and the payoffs that result from all logically possible combinations of all the players' choices. In Chapter 6, we saw how changing the order of moves from sequential to simultaneous or vice versa can alter the game's outcomes. Adding or removing moves available to a player or changing the payoffs at some terminal nodes or in some cells of the game table also can change outcomes. Unless the rules of a game are fixed by an outside authority, each player has the incentive to manipulate them to produce an outcome that is more to his own advantage. Devices to manipulate a game in this way are called **strategic moves**, which are the subject of this chapter.

A strategic move changes the rules of the original game to create a new two-stage game. In this sense, strategic moves are similar to the direct communications of information that we examined in Chapter 9. With strategic moves, though, the second stage is the original game, often with some alteration of the order of moves and the payoffs; there was no such alteration in our games with direct communication. The first stage in a game with strategic moves specifies how you will act in the second stage. Different first-stage actions correspond to different strategic moves, and we classify them into three types: *commitments*, *threats*, and *promises*. The aim of all three is to alter the outcome of the second-stage game to your own advantage. Which, if any, suits your purpose depends on the context. But most important, any of the three works only if the other player believes that at the second stage you will indeed do what you declared at the first

stage. In other words, the *credibility* of the strategic move is open to question. Only a credible strategic move will have the desired effect and, as was often the case in Chapter 9, mere declarations are not enough. At the first stage, you must take some ancillary actions that lend credibility to your declared second-stage actions. We will study both the kinds of second-stage actions that work to your benefit and the first-stage ancillary moves that make them credible.

You are probably more familiar with the use and credibility of strategic moves than you might think. Parents, for instance, constantly attempt to influence the behavior of their children by using threats ("no dessert unless you finish your vegetables") and promises ("you will get the new racing bike at the end of the term if you maintain at least a B average in school"). And children know very well that many of these threats and promises are not credible; much bad behavior can escape the threatened punishment if the child sweetly promises not to do that again, even though the promise itself may not be credible. Furthermore, when the children get older and become concerned with their own appearance, they find themselves making commitments to themselves to exercise and diet; many of these commitments also turn out to lack credibility. All of these devices—commitments, threats, and promises—are examples of strategic moves. Their purpose is to alter the actions of another player, perhaps even your own future self, at a later stage in a game. But they will not achieve this purpose unless they are credible. In this chapter, we will use game theory to study systematically how to use such strategies and how to make them credible.

Be warned, however, that credibility is a difficult and subtle matter. We can offer you some general principles and an overall understanding of how strategic moves can work—a science of strategy. But actually making them work depends on your specific understanding of the context, and your opponent may get the better of you by having a better understanding of the concepts or the context or both. Therefore the use of strategic moves in practice retains a substantial component of art. It also entails risk, particularly when using the strategy of **brinkmanship**, which can sometimes lead to disasters. You can have success as well as fun trying to put these ideas into practice, but note our disclaimer and warning: Use such strategies at your own risk.

### 1 A CLASSIFICATION OF STRATEGIC MOVES

Because the use of strategic moves depends so critically on the order of moves, to study them we need to know what it means to "move first." Thus far we have taken this concept to be self-evident, but now we need to make it more precise. It has two components. First, your action must be **observable** to the other player; second, it must be **irreversible**.

Consider a strategic interaction between two players, A and B, in which A's move is made first. If A's choice is not observable to B, then B cannot respond to it, and the mere chronology of action is irrelevant. For example, suppose A and B are two companies bidding in an auction. A's committee meets in secret on Monday to determine its bid; B's committee meets on Tuesday; the bids are mailed separately to the auctioneer and opened on Friday. When B makes its decision, it does not know what A has done; therefore the moves are strategically the same as if they were simultaneous.

If A's move is not irreversible, then A might pretend to do one thing, lure B into responding, and then change its own action to its own advantage. B should anticipate this ruse and not be lured; then it will not be responding to A's choice. Once again, in the true strategic sense A does not have the first move.

Considerations of observability and irreversibility affect the nature and types of strategic moves as well as their credibility. We begin with a taxonomy of strategic moves available to players.

### A. Unconditional Strategic Moves

Let us suppose that player A is the one making a strategic observable and irreversible move in the first stage of the game. He can declare: "In the game to follow, I will make a particular move, X." This declaration says that A's future move is unconditional; A will do X irrespective of what B does. Such a statement, if credible, is tantamount to changing the order of the game at stage 2 so that A moves first and B second, and A's first move is X. This strategic move is called a **commitment**.

If the previous rules of the game at the second stage already have A moving first, then such a declaration would be irrelevant. But if the game at the second stage has simultaneous moves or if A is to move second there, then such a declaration, if credible, can change the outcome because it changes B's beliefs about the consequences of his actions. Thus a **commitment** is a simple seizing of the first-mover advantage when it exists.

In the street-garden game of Chapter 3, three women play a sequential-move game in which each must decide whether to contribute toward the creation of a public flower garden on their street; two or more contributors are necessary for the creation of a pleasant garden. The rollback equilibrium entails the first player (Emily) choosing not to contribute while the other players (Nina and Talia) do contribute. By making a credible commitment not to contribute, however, Talia (or Nina) could alter the outcome of the game. Even though she does not get her turn to announce her decision until after Emily and Nina have made theirs public, Talia could let it be known that she has sunk all of her savings (and energy) into a large house-renovation project, and so she will have absolutely nothing left to contribute to the street garden. Then Talia essentially commits herself not to contribute regardless of Emily's and Nina's decisions, before

Emily and Nina make those decisions. In other words, Talia changes the game to one in which she is in effect the first mover. You can easily check that the new rollback equilibrium entails Emily and Nina both contributing to the garden and the equilibrium payoffs are 3 to each of them but 4 to Talia—the equilibrium outcome associated with the game when Talia moves first. Several more detailed examples of commitments are given in the following sections.

### B. Conditional Strategic Moves

Another possibility for A is to declare at the first stage: "In the game to follow, I will respond to your choices in the following way. If you choose  $Y_1$ , I will do  $Z_1$ ; if you do  $Y_2$ , I will do  $Z_2$ , . . ." In other words, A can use a move that is conditional on B's behavior; we call this type of move a **response rule** or **reaction function**. A's statement means that, in the game to be played at the second stage, A will move second, but how he will respond to B's choices at that point is already predetermined by A's declaration at stage 1. For such declarations to be meaningful, A must be physically able to wait to make his move at the second stage until after he has observed what B has irreversibly done. In other words, at the second stage, B should have the true first move in the double sense just explained.

Conditional strategic moves take different forms, depending on what they are trying to achieve and how they set about achieving it. When A wants to stop B from doing something, we say that A is trying to deter B, or to achieve **deterrence**; when A wants to induce B to do something, we say that A is trying to compel B, or to achieve **compellence**. We return to this distinction later. Of more immediate interest is the method used in pursuit of either of these aims. If A declares, "Unless your action (or inaction, as the case may be) conforms to my stated wish, I will respond in a way that will hurt you," that is, a **threat**. If A declares, "If your action (or inaction, as the case may be) conforms to my stated wish, I will respond in a way that will reward you," that is, a **promise**. "Hurt" and "reward" are measured in terms of the payoffs in the game itself. When A hurts B, A does something that lowers B's payoff; when A rewards B, A does something that leads to a higher payoff for B. Threats and promises are the two conditional strategic moves on which we focus our analysis.

To understand the nature of these strategies, consider the dinner game mentioned earlier. In the natural chronological order of moves, first the child decides whether to eat his vegetables, and then the parent decides whether to give the child dessert. Rollback analysis tells us the outcome: the child refuses to eat the vegetables, knowing that the parent, unwilling to see the child hungry and unhappy, will give him the dessert. The parent can foresee this outcome, however, and can try to alter it by making an initial move—namely, by stating a conditional response rule of the form "no dessert unless you finish your vegetables." This declaration constitutes a threat. It is a first move in a pregame, which fixes

how you will make your second move in the actual game to follow. If the child believes the threat, that alters the child's rollback calculation. The child "prunes" that branch of the game tree in which the parent serves dessert even if the child has not finished his vegetables. This may alter the child's behavior; the parent hopes that it will make the child act as the parent wants him to. Similarly, in the "study game," the promise of the bike may induce a child to study harder.

## 2 CREDIBILITY OF STRATEGIC MOVES

We have already seen that payoffs to the other player can be altered by one player's strategic move, but what about the payoffs for the player making that move? Player A gets a higher payoff when B acts in conformity with A's wishes. But A's payoff also may be affected by his own response. In regard to a threat, A's threatened response if B does not act as A would wish may have consequences for A's own payoffs: the parent may be made unhappy by the sight of the unhappy child who has been denied dessert. Similarly, in regard to a promise, rewarding B if he does act as A would wish can affect A's own payoff: the parent who rewards the child for studying hard has to incur the monetary cost of the gift but is happy to see the child's happiness on receiving the gift and even happier about the academic performance of the child.

This effect on A's payoffs has an important implication for the efficacy of A's strategic moves. Consider the threat. If A's payoff is actually increased by carrying out the threatened action, then B reasons that A will carry out this action even if B fulfills A's demands. Therefore B has no incentive to comply with A's wishes, and the threat is ineffective. For example, if the parent is a sadist who enjoys seeing the child go without dessert, then the child thinks, "I am not going to get dessert anyway, so why eat the vegetables?"

Therefore an essential aspect of a threat is that it should be *costly* for the threatener to carry out the threatened action. In the dinner game, the parent must prefer to give the child dessert. Threats in the true strategic sense have the innate property of imposing some cost on the threatener, too; they are threats of *mutual harm*.

In technical terms, a threat fixes your strategy (response rule) in the subsequent game. A strategy must specify what you will do in each eventuality along the game tree. Thus, "no dessert if you don't finish your vegetables" is an incomplete specification of the strategy; it should be supplemented by "and dessert if you do." Threats generally don't specify this latter part. Why not? Because the second part of the strategy is automatically understood; it is implicit. And for the threat to work, this second part of the strategy—the *implied promise* in this case—has to be automatically credible, too.

Thus the threat "no dessert if you don't finish your vegetables" carries with it an implicit promise of "dessert if you do finish your vegetables." This promise also should be credible if the threat is to have the desired effect. In our example, the credibility of the implicit promise is automatic when the parent prefers to see the child get and enjoy his dessert. In other words, the implicit promise is automatically credible precisely when the threatened action is costly for the parent to carry out.

To put it yet another way, a threat carries with it the stipulation that you will do something if your wishes are not met that, if those circumstances actually arise, you will regret having to do. Then why make this stipulation at the first stage? Why tie your own hands in this way when it might seem that leaving one's options open would always be preferable? Because in the realm of game theory, having more options is not always preferable. In regard to a threat, your lack of freedom in the second stage of the game has strategic value. It changes other players' expectations about your future responses, and you can use this change in expectations to your advantage.

A similar effect arises with a promise. If the child knows that the parent enjoys giving him gifts, he may expect to get the racing bike anyway on some occasion in the near future—for example, an upcoming birthday. Then the promise of the bike has little effect on the child's incentive to study hard. To have the intended strategic effect, the promised reward must be so costly to provide that the other player would not expect you to hand over that reward anyway. (This is a useful lesson in strategy that you can point out to your parents: the rewards that they promise must be larger and more costly than what they would give you just for the pleasure of seeing you happy.)

The same is true of unconditional strategic moves (commitments, too). In bargaining, for example, others know that, when you have the freedom to act, you also have the freedom to capitulate; so a "no concessions" commitment can secure you a better deal. If you hold out for 60% of the pie and the other party offers you 55%, you may be tempted to take it. But if you can credibly assert in advance that you will not take less than 60%, then this temptation does not arise and you can do better than you otherwise would.

Thus it is in the very nature of strategic moves that after the fact—that is, when the stage 2 game actually requires it—you do not want to carry out the action that you had stipulated you would take. This is true for all types of strategic moves and it is what makes credibility so problematic. You have to do something at the first stage to create credibility—something that convincingly tells the other player that you will not give in to the temptation to deviate from the stipulated action when the time comes—in order for your strategic move to work. That is why giving up your own freedom to act can be strategically beneficial. Alternatively, credibility can be achieved by changing your own payoffs in the second-stage game in such a way that it becomes truly optimal for you to act as you declare.



Thus there are two general ways of making your strategic moves credible: (1) remove from your own set of future choices the other moves that may tempt you or (2) reduce your own payoffs from those temptation moves so that the stipulated move becomes the actual best one. In the sections that follow, we first elucidate the mechanics of strategic moves, assuming them to be credible. We make some comments about credibility as we go along but postpone our general analysis of credibility until the last section of the chapter.

### 3 COMMITMENTS

We studied the game of chicken in Chapter 4 and found two pure-strategy Nash equilibria. Each player prefers the equilibrium in which he goes straight and the other person swerves.<sup>1</sup> We saw in Chapter 6 that, if the game were to have sequential rather than simultaneous moves, the first mover would choose Straight, leaving the second to make the best of the situation by settling for Swerve rather than causing a crash. Now we can consider the same matter from another perspective. Even if the game itself has simultaneous moves, if one player can make a strategic move—create a first stage in which he makes a credible declaration about his action in the chicken game itself, which is to be played at the second stage—then he can get the same advantage afforded a first mover by making a commitment to act tough (choose Straight).

Although the point is simple, we outline the formal analysis to develop your understanding and skill, which will be useful for later, more complex ex-

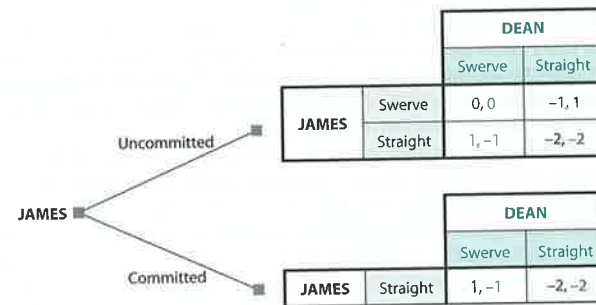


FIGURE 10.1 Chicken: Commitment by Restricting Freedom to Act

<sup>1</sup>We saw in Chapter 7 and will see again in Chapter 13 that the game has a third equilibrium, in mixed strategies, in which both players do quite poorly.

amples. Remember our two players, James and Dean. Suppose James is the one who has the opportunity to make a strategic move. Figure 10.1 shows the tree for the two-stage game. At the first stage, James has to decide whether to make a commitment. Along the upper branch emerging from the first node, he does not make the commitment. Then at the second stage the simultaneous-move game is played, and its payoff table is the familiar one shown in Figures 4.14 and 6.6. This second-stage game has multiple equilibria, and James gets his best payoff in only one of them. Along the lower branch, James makes the commitment. Here, we interpret this commitment to mean giving up his freedom to act in such a way that Straight is the only action available to James at this stage. Therefore the second-stage game table has only one row for James, corresponding to his declared choice of Straight. In this table, Dean's best action is Swerve; so the equilibrium outcome gives James his best payoff. Therefore, at the first stage, James finds it optimal to make the commitment; this strategic move ensures his best payoff, while not committing leaves the matter uncertain.

How can James make this commitment credibly? Like any first move, the commitment move must be (1) irreversible and (2) visible to the other player. People have suggested some extreme and amusing ideas. James can disconnect the steering wheel of the car and throw it out of the window so that Dean can see that James can no longer Swerve. (James could just tie the wheel so that it could no longer be turned, but it would be more difficult to demonstrate to Dean that the wheel was truly tied and that the knot was not a trick one that could be undone quickly.) These devices simply remove the Swerve option from the set of choices available to James in the stage 2 game, leaving Straight as the only thing he can do.

More plausibly, if such games are played every weekend, James can acquire a general reputation for toughness that acts as a guarantee of his action on any one day. In other words, James can alter his own payoff from swerving by subtracting an amount that represents the loss of reputation. If this amount is large enough—say, 3—then the second-stage game when James has made the commitment has a different payoff table. The complete tree for this version of the game is shown in Figure 10.2.

Now, in the second stage with commitment, Straight has become truly optimal for James; in fact, it is his dominant strategy in that stage. Dean's optimal strategy is then Swerve. Looking ahead to this outcome at stage 1, James sees that he gets 1 by making the commitment (changing his own stage 2 payoffs), while without the commitment he cannot be sure of 1 and may do much worse. Thus a rollback analysis shows that James should make the commitment.

Both (or all) can play the game of commitment, so success may depend both on the speed with which you can seize the first move and on the credibility with which you can make that move. If there are lags in observation, the two may even make incompatible simultaneous commitments: each disconnects his steering wheel and tosses it out of the window just as he sees the other's wheel come flying out, and then the crash is unavoidable.

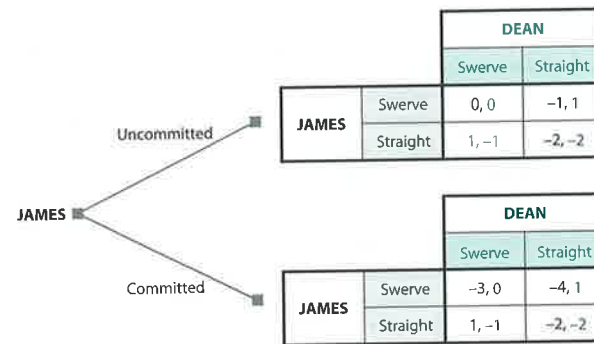


FIGURE 10.2 Chicken: Commitment by Changing Payoffs

Even if one of the players has the advantage in making a commitment, the other player can defeat the first player's attempt to do so. The second player could demonstrably remove his ability to "see" the other's commitment, for example, by cutting off communication.

Games of chicken may be a 1950s anachronism, but our second example is perennial and familiar. In a class, the teacher's deadline enforcement policy can be Weak or Tough, and the students' work can be Punctual or Late. Figure 10.3 shows this game in the strategic form. The teacher does not like being tough; for him the best outcome (a payoff of 4) is when students are punctual even when he is weak; the worst (1) is when he is tough but students are still late. Of the two intermediate strategies, he recognizes the importance of punctuality and rates (Tough, Punctual) better than (Weak, Late). The students most prefer the outcome (Weak, Late), where they can party all weekend without suffering any penalty for the late assignment. (Tough, Late) is the worst for them, just as it is for the teacher. Between the intermediate ones, they prefer (Weak, Punctual) to (Tough, Punctual) because they have higher self-esteem if they can think that

		STUDENT	
		Punctual	Late
TEACHER	Weak	4, 3	2, 4
	Tough	3, 2	1, 1

FIGURE 10.3 Payoff Table for Class Deadline Game

they acted punctually of their own volition rather than because of the threat of a penalty.<sup>2</sup>

If this game is played as a simultaneous-move game or if the teacher moves second, Weak is dominant for the teacher, and then the student chooses Late. The equilibrium outcome is (Weak, Late), and the payoffs are (2, 4). But the teacher can achieve a better outcome by committing at the outset to the policy of Tough. We do not draw a tree as we did in Figures 10.1 and 10.2. The tree would be very similar to that for the preceding chicken case, and so we leave it for you to draw. Without the commitment, the second-stage game is as before, and the teacher gets a 2. When the teacher is committed to Tough, the students find it better to respond with Punctual at the second stage, and the teacher gets a 3.

The teacher commits to a move different from what he would do in simultaneous play or indeed, his best second move if the students moved first. This is where strategic thinking enters. The teacher has nothing to gain by declaring that he will have a Weak enforcement regime; the students expect that anyway in the absence of any declaration. To gain advantage by making a strategic move, he must commit not to follow what would be his equilibrium strategy of the simultaneous-move game. This strategic move changes the students' expectations and therefore their action. Once they believe the teacher is really committed to tough discipline, they will choose to turn in their assignments punctually. If they tested this out by being late, the teacher would like to forgive them, maybe with an excuse to himself, such as "just this once." The existence of this temptation to shift away from your commitment is what makes its credibility problematic.

Even more dramatic, in this instance the teacher benefits by making a strategic move that commits him to a dominated strategy. He commits to choosing Tough, which is dominated by Weak. The choice of Tough gets the teacher a 3 if the student chooses Punctual and a 1 if the student chooses Late, whereas if the teacher had chosen Weak, his corresponding payoffs would have been 4 and 2. If you think it paradoxical that one can gain by choosing a dominated strategy, you are extending the concept of dominance beyond the proper scope of its validity. Dominance entails either of two calculations: (1) After the other player does something, how do I respond, and is some choice best (or worst), given all possibilities? (2) If the other player is simultaneously doing action X, what is best (or worst) for me, and is this the same for all the X actions that the other could be choosing? Neither is relevant when you are moving first. Instead, you must look

<sup>2</sup>You may not regard these specific rankings of outcomes as applicable either to you or to your own teachers. We ask you to accept them for this example, whose main purpose is to convey some general ideas about commitment in a simple way. The same disclaimer applies to all the examples that follow.

ahead to how the other will respond. Therefore the teacher does not compare his payoffs in vertically adjacent cells of the table (taking the possible actions of the students one at a time). Instead, he calculates how the students will react to each of his moves. If he is committed to Tough, they will be Punctual, but if he is committed to Weak (or uncommitted), they will be Late, so the only pertinent comparison is that of the top-right cell with the bottom left, of which the teacher prefers the latter.

To be credible, the teacher's commitment must be everything a first move has to be. First, it must be made before the other side makes its move. The teacher must establish the ground rules of deadline enforcement before the assignment is due. Next, it must be observable—the students must know the rules by which they must abide. Finally, and perhaps most important, it must be irreversible—the students must know that the teacher cannot, or at any rate will not, change his mind and forgive them. A teacher who leaves loopholes and provisions for incompletely specified emergencies is merely inviting imaginative excuses accompanied by fulsome apologies and assertions that "it won't happen again."

The teacher might achieve credibility by hiding behind general university regulations; this simply removes the Weak option from his set of available choices at stage 2. Or, as is true in the chicken game, he might establish a reputation for toughness, changing his own payoffs from Weak by creating a sufficiently high cost of loss of reputation.

## 4 THREATS AND PROMISES

We emphasize that threats and promises are *response rules*: your actual future action is conditioned on what the other players do in the meantime, but your freedom of future action is constrained to following the stated rule. Once again, the aim is to alter the other players' expectations and therefore their actions in a way favorable to you. Tying yourself to a rule that you would not want to follow if you were completely free to act at the later time is an essential part of this process. Thus the initial declaration of intention must be credible. Once again, we will elucidate some principles for achieving credibility of these moves, but we remind you that their actual implementation remains largely an art.

Remember the taxonomy given in Section 1. A *threat* is a response rule that leads to a bad outcome for the other players if they act contrary to your interests. A *promise* is a response rule by which you offer to create a good outcome for the other players if they act in a way that promotes your own interests. Each of these responses may aim either to stop the other players from doing something that they would otherwise do (*deterrence*) or to induce them to do

something that they would otherwise not do (*compellence*). We consider these features in turn.

### A. Example of a Threat: U.S.–Japan Trade Relations

Our example comes from a hardy perennial of U.S. international economic policy—namely, trade friction with Japan. Each country has the choice of keeping its own markets open or closed to the other's goods. They have somewhat different preferences regarding the outcomes.

Figure 10.4 shows the payoff table for the trade game. For the United States, the best outcome (a payoff of 4) comes when both markets are open; this is partly because of its overall commitment to the market system and free trade and partly because of the benefit of trade with Japan itself—U.S. consumers get high-quality cars and consumer electronics products, and U.S. producers can export their agricultural and high-tech products. Similarly, its worst outcome (payoff 1) occurs when both markets are closed. Of the two outcomes when only one market is open, the United States would prefer its own market to be open, because the Japanese market is smaller, and loss of access to it is less important than the loss of access to Hondas and video games.

As for Japan, for the purpose of this example we accept the protectionist, producer-oriented picture of Japan, Inc. Its best outcome is when the U.S. market is open and its own is closed; its worst is when matters are the other way around. Of the other two outcomes, it prefers that both markets be open, because its producers then have access to the much larger U.S. market.<sup>3</sup>

Both sides have dominant strategies. No matter how the game is played—simultaneously or sequentially with either move order—the equilibrium outcome is (Open, Closed), and the payoffs are (3, 4). This outcome also fits well the common American impression of how the actual trade policies of the two countries work.

		JAPAN	
		Open	Closed
UNITED STATES	Open	4, 3	3, 4
	Closed	2, 1	1, 2

FIGURE 10.4 Payoff Table for U.S.–Japan Trade Game

<sup>3</sup>Again, we ask you to accept this payoff structure as a vehicle for conveying the ideas. You can experiment with the payoff tables to see what difference that would make to the role and effectiveness of the strategic moves.



Japan is already getting its best payoff in this equilibrium and so has no need to try any strategic moves. The United States, however, can try to get a 4 instead of a 3. But in this case an ordinary unconditional commitment will not work. Japan's best response, no matter what commitment the United States makes, is to keep its market closed. Then the United States does better for itself by committing to keep its own market open, which is the equilibrium without any strategic moves anyway.

But suppose the United States can choose the following conditional response rule: "We will close our market if you close yours." The situation then becomes the two-stage game shown in Figure 10.5. If the United States does not use the threat, the second stage is as before and leads to the equilibrium in which the U.S. market is open and it gets a 3, whereas the Japanese market is closed and it gets a 4. If the United States does use the threat, then at the second stage only Japan has freedom of choice; given what Japan does, the United States then merely does what its response rule dictates. Therefore, along this branch of the tree, we show only Japan as an active player and write down the payoffs to the two parties: If Japan keeps its market closed, the United States closes its own, and the United States gets a 1 and Japan gets a 2. If Japan keeps its market open, then the United States threat has worked, it is happy to keep its own market open, and it gets a 4, while Japan gets a 3. Of these two possibilities, the second is better for Japan.

Now we can use the familiar rollback reasoning. Knowing how the second stage will work in all eventualities, it is better for the United States to deploy its threat at the first stage. This threat will result in an open market in Japan, and the United States will get its best outcome.

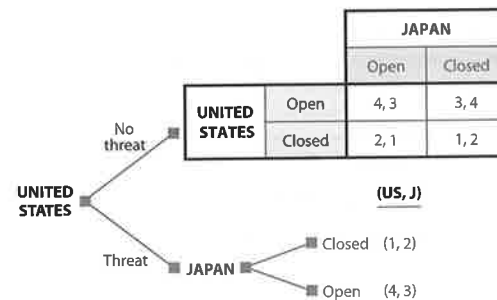


FIGURE 10.5 Tree for the U.S.-Japan Trade Game with Threat

Having described the mechanics of the threat, we now point out some of its important features.

1. When the United States deploys its threat credibly, Japan doesn't follow its dominant strategy Closed. Again, the idea of dominance is relevant only in the context of simultaneous moves or when Japan moves second. Here, Japan knows that the United States will take actions that depart from its dominant strategy. In the payoff table, Japan is looking at a choice between just two cells, the top left and the bottom right, and, of those two, it prefers the latter.

2. Credibility of the threat is problematic because, if Japan puts it to the test by keeping its market closed, the United States faces the temptation to refrain from carrying out the threat. In fact, if the threatened action were the best U.S. response after the fact, then there would be no need to make the threat in advance (but the United States might issue a *warning* just to make sure that the Japanese understand the situation). The strategic move has a special role exactly because it locks a player into doing something other than what it would have wanted to do after the fact. As explained earlier, a threat in the true strategic sense is necessarily costly for the threatener to carry out; the threatened action would inflict *mutual* harm.

3. The conditional rule "We will close our market if you close yours" does not completely specify the U.S. strategy. To be complete, it needs an additional clause indicating what the United States will do in response to an open Japanese market: "and we will keep our market open if you keep yours open." This additional clause, the implicit promise, is really part of the threat, but it does not need to be stated explicitly, because it is automatically credible. Given the payoffs of the second-stage game, it is in the best interests of the United States to keep its market open if Japan keeps its market open. If that were not the case, if the United States would respond by keeping its market closed even when Japan kept its own market open, then the implicit promise would have to be made explicit and somehow made credible. Otherwise, the U.S. threat would become tantamount to the unconditional commitment "We will keep our market closed," and that would not elicit the desired response from Japan.

4. The threat, when credibly deployed, results in a change in Japan's action. We can regard this as deterrence or compellence, depending on the status quo. If the Japanese market is initially open, and the Japanese are considering a switch to protectionism, then the threat deters them from that action. But if the Japanese market is initially closed, then the threat compels them to open it. Thus whether a strategic move is deterrent or compellent depends on the status quo. The distinction may seem to be a matter of semantics, but in practice the credibility of a move and the way that it works are importantly affected by this distinction. We return to this matter later in the chapter.



5. Here are a few ways in which the United States can make its threat credible. First, it can enact a law that mandates the threatened action under the right circumstances. This removes the temptation action from the set of available choices at stage 2. Some reciprocity provisions in the World Trade Organization agreements have this effect, but the procedures are very slow and uncertain. Second, it can delegate fulfillment to an agency such as the U.S. Commerce Department that is captured by U.S. producers who would like to keep our markets closed and so reduce the competitive pressure on themselves. This changes the U.S. payoffs at stage 2—replacing the true U.S. payoffs by those of the Commerce Department—with the result that the threatened action becomes truly optimal. (The danger is that the Commerce Department will then retain a protectionist stance even if Japan opens its market; gaining credibility for the threat may lose credibility for the implied promise.)

6. If a threat works, it doesn't have to be carried out. So its cost to you is immaterial. In practice, the danger that you may have miscalculated or the risk that the threatened action will take place by error even if the other player complies is a strong reason to refrain from using threats more severe than necessary. To make the point starkly, the United States could threaten to pull out of defensive alliances with Japan if it didn't buy our rice and semiconductors, but that threat is "too big" and too risky for the United States ever to carry out; therefore it is not credible.

But sometimes a range of threats is not available from which a player can choose one that is, on the one hand, sufficiently big that the other player fears it and alters his action in the way that the first player desires and, on the other hand, not so big as to be too risky for the first player ever to carry out and therefore lacking credibility. If the only available threat is too big, then a player can reduce its size by making its fulfillment a matter of chance. Instead of saying, "If you don't open your markets, we will refuse to defend you in the future," the United States can say to Japan, "If you don't open your markets, the relations between our countries will deteriorate to the point where Congress may refuse to allow us to come to your assistance if you are ever attacked, even though we do have an alliance." In fact, the United States can deliberately foster sentiments that raise the probability that Congress will do just that, so the Japanese will feel the danger more vividly. A threat of this kind, which creates a risk but not a certainty of the bad outcome, is called brinkmanship. It is an extremely delicate and even dangerous variant of the strategic move. We will study brinkmanship in greater detail in Chapter 15.

7. Japan gets a worse outcome when the United States deploys its threat than it would without this threat, so it would like to take strategic actions that defeat or disable U.S. attempts to use the threat. For example, suppose its market is currently closed, and the United States is attempting compellence. The Japanese can accede in principle but stall in practice, pleading unavoidable delays for assembling the necessary political consensus to legislate the market

opening, then delays for writing the necessary administrative regulations to implement the legislation, and so on. Because the United States does not want to go ahead with its threatened action, at each point it has the temptation to accept the delay. Or Japan can claim that its domestic politics makes it difficult to open all markets fully; will the United States accept the outcome if Japan keeps just a few of its industries protected? It gradually expands this list, and at any point the extra small step is not enough cause for the United States to unleash a trade war. This device of defeating a compellent threat by small steps, or "slice by slice," is called *salami tactics*.

### B. Example of a Promise: The Restaurant Pricing Game

We now illustrate a promise by using the restaurant pricing game of Chapter 5. We saw in Chapter 5 that the game is a prisoners' dilemma, and we simplify it here by supposing that only two choices of price are available: the jointly best price of \$26 or the Nash equilibrium price of \$20. The profits for each restaurant in this version of the game can be calculated by using the functions in Section 1 of Chapter 5; the results are shown in Figure 10.6. Without any strategic moves, the game has the usual equilibrium in dominant strategies in which both stores charge the low price of 20, and both get lower profits than they would if they both charged the high price of 26.

If either side can make the credible promise "I will charge a high price if you do," the cooperative outcome is achieved. For example, if Xavier's makes the promise, then Yvonne's knows that its choice of 26 will be reciprocated, leading to the payoff shown in the lower-right cell of the table, and that its choice of 20 will bring forth Xavier's usual action—namely, 20—leading to the upper-left cell. Between the two, Yvonne's prefers the first and therefore chooses the high price.

The analysis can be done more properly by drawing a tree for the two-stage game in which Xavier's has the choice of making or not making the promise at the first stage. We omit the tree, partly so that you can improve your understanding of the process by constructing it yourself and partly to show how such detailed analysis becomes unnecessary as one becomes familiar with the ideas.

		YVONNE'S BISTRO	
		20 (low)	26 (high)
XAVIER'S TAPAS	20 (low)	288, 288	360, 216
	26 (high)	216, 360	324, 324

FIGURE 10.6 Payoff Table for Restaurant Prisoners' Dilemma (\$100s per month)

The credibility of Xavier's promise is open to doubt. To respond to what Yvonne's does, Xavier's must arrange to move second in the second stage of the game; correspondingly, Yvonne's must move first in stage 2. Remember that a first move is an irreversible and observable action. Therefore, if Yvonne's moves first and prices high, it leaves itself vulnerable to Xavier's cheating, and Xavier's is very tempted to renege on its promise to price high when it sees Yvonne's in this vulnerable position. Xavier's must somehow convince Yvonne's that it will not give in to the temptation to charge a low price when Yvonne's charges a high price.

How can it do so? Perhaps Xavier's owner can leave the pricing decision in the hands of a local manager, with clear written instructions to reciprocate with the high price if Yvonne's charges the high price. Xavier's owner can invite Yvonne's to inspect these instructions, after which he leaves on a solo round-the-world sailing trip so that he cannot rescind them. (Even then, Yvonne's management may be doubtful—Xavier might secretly carry a telephone or a laptop computer on board.) This scenario is tantamount to removing the cheating action from the choices available to Xavier's at stage 2.

Or Xavier's restaurant can develop a reputation for keeping its promises, in business and in the community more generally. In a repeated relationship, the promise may work because reneging on the promise once may cause future cooperation to collapse. In essence, an ongoing relationship means splitting the game into smaller segments, in each of which the benefit from reneging is too small to justify the costs. In each such game, then, the payoff from cheating is altered by the cost of collapse of future cooperation.<sup>4</sup>

We saw earlier that every threat has an implicit attached promise. Similarly, every promise has an implicit attached threat. In this case, the threat is "I will charge the low price if you do." It does not have to be stated explicitly, because it is automatically credible—it describes Xavier's best response to Yvonne's low price.

There is also an important difference between a threat and a promise. If a threat is successful, it doesn't have to be carried out and is then costless to the threatener. Therefore a threat can be bigger than what is needed to make it effective (although making it too big may be too risky, even to the point of losing its credibility as suggested earlier). If a promise is successful in altering the other's action in the desired direction, then the promisor has to deliver what he had promised, and so it is costly. In the preceding example, the cost is simply giving up the opportunity to cheat and get the highest payoff; in other instances where the promisor offers an actual gift or an inducement to the other, the cost may

<sup>4</sup>In Chapter 11, we will investigate in great detail the importance of repeated or ongoing relationships in attempts to reach the cooperative outcome in a prisoners' dilemma.

be more tangible. In either case, the player making the promise has a natural incentive to keep its size small—just big enough to be effective.

### C. Example Combining Threat and Promise: Joint U.S.–China Political Action

When we considered threats and promises one at a time, the explicit statement of a threat included an implicit clause of a promise that was automatically credible, and vice versa. There can, however, be situations in which the credibility of both aspects is open to question; then the strategic move has to make both aspects explicit and make them both credible.

Our example of an explicit-threat-and-promise combination comes from a context in which multiple nations must work together toward some common goal in dealing with a dangerous situation in a neighboring country. Specifically, we consider an example of the United States and China contemplating whether to take action to compel North Korea to give up its nuclear weapons programs. We show in Figure 10.7 the payoff table for the United States and China when each must choose between action and inaction.

Each country would like the other to take on the whole burden of taking action against the North Koreans; so the top-right cell has the best payoff for China (4), and the bottom-left cell is best for the United States. The worst situation for the United States is where no action is taken, because it finds the increased threat of nuclear war in that case to be unacceptable. For China, however, the worst outcome arises when it takes on the whole burden of action, because the costs of action are so high. Both regard a joint involvement as the second best (a payoff of 3). The United States assigns a payoff of 2 to the situation in which it is the only one to act. And for China, a payoff of 2 is assigned to the case in which no action is taken.

Without any strategic moves, the intervention game is dominance solvable. Inaction is the dominant strategy for China, and then Action is the best choice for the United States. The equilibrium outcome is the top-right cell, with payoffs of 2 for the United States and 4 for China. Because China gets its best outcome, it has no reason to try any strategic moves. But the United States can try to do better than a 2.

		CHINA	
		Action	Inaction
UNITED STATES	Action	3, 3	2, 4
	Inaction	4, 1	1, 2

FIGURE 10.7 Payoff Table for U.S.–China Political Action Game



What strategic move will work to improve the equilibrium payoff for the United States? An unconditional move (commitment) will not work, because China will respond with "Inaction" to either first move by the United States. A threat alone ("We won't take action unless you do") does not work, because the implied promise ("We will if you do") is not credible—if China does act, the United States would prefer to back off and leave everything to China, getting a payoff of 4 instead of the 3 that would come from fulfilling the promise. A promise alone won't work: because China knows that the United States will intervene if China does not, an American promise of "We will intervene if you do" becomes tantamount to a simple commitment to intervene; then China can stay out and get its best payoff of 4.

In this game, an explicit promise from the United States must carry the implied threat "We won't take action if you don't," but that threat is not automatically credible. Similarly, America's explicit threat must carry the implied promise "We will act if you do," but that is not automatically credible, either. Therefore the United States has to make both the threat and the promise explicit. It must issue the combined threat-cum-promise "We will act if, and only if, you do." It needs to make both clauses credible. Usually such credibility has to be achieved by means of a treaty that covers the whole relationship, not just with agreements negotiated separately when each incident arises.

## S SOME ADDITIONAL TOPICS

### A. When Do Strategic Moves Help?

We have seen several examples in which a strategic move brings a better outcome to one player or another, compared with the original game without such moves. What can be said in general about the desirability of such moves?

An unconditional move—a commitment—need not always be advantageous to the player making it. In fact, if the original game gives the advantage to the second mover, then it is a mistake to commit oneself to move in advance, thereby effectively becoming the first mover.

The availability of a conditional move—threat or promise—can never be an actual disadvantage. At the very worst, one can commit to a response rule that would have been optimal after the fact. However, if such moves bring one an actual gain, it must be because one is choosing a response rule that in some eventualities specifies an action different from what one would find optimal at that later time. Thus whenever threats and promises bring a positive gain, they do so precisely when (one might say precisely because) their credibility is inherently questionable and must be achieved by some specific credibility "device."

We have mentioned some such devices in connection with each earlier example and will later discuss the topic of achieving credibility in greater generality.

What about the desirability of being on the receiving end of a strategic move? It is never desirable to let the other player threaten you. If a threat seems likely, you can gain by looking for a different kind of advance action—one that makes the threat less effective or less credible. We will consider some such actions shortly. However, it is often desirable to let the other player make promises to you. In fact, both players may benefit when one can make a credible promise, as in the prisoners' dilemma example of restaurant pricing earlier in this chapter, in which a promise achieved the cooperative outcome. Thus it may be in the players' mutual interest to facilitate the making of promises by one or both of them.

### B. Deterrence Versus Compellence

In principle, either a threat or a promise can achieve either deterrence or compellence. For example, a parent who wants a child to study hard (compellence) can promise a reward (a new racing bike) for good performance in school or can threaten a punishment (a strict curfew the following term) if the performance is not sufficiently good. Similarly, a parent who wants the child to keep away from bad company (deterrence) can try either a reward (promise) or a punishment (threat). In practice, the two types of strategic moves work somewhat differently, and that will affect the ultimate decision regarding which to use. Generally, deterrence is better achieved by a threat and compellence by a promise. The reason is an underlying difference of timing and initiative.

A deterrent threat can be passive—you don't need to do anything so long as the other player doesn't do what you are trying to deter. And it can be static—you don't have to impose any time limit. Thus you can set a trip wire and then leave things up to the other player. So the parent who wants the child to keep away from bad company can say, "If I ever catch you with X again, I will impose a 7 P.M. curfew on you for a whole year." Then the parent can sit back to wait and watch; only if the child acts contrary to the parent's wishes does the parent have to act on her threat. Trying to achieve the same deterrence by a promise would require more complex monitoring and continual action: "At the end of each month in which I know that you did not associate with X, I will give you \$25."

Compellence must have a deadline or it is pointless—the other side can defeat your purpose by procrastinating or by eroding your threat in small steps (salami tactics). This makes a compellent threat harder to implement than a compellent promise. The parent who wants the child to study hard can simply say, "Each term that you get an average of B or better, I will give you CDs or games worth \$500." The child will then take the initiative in showing the parent each time he has fulfilled the conditions. Trying to achieve the same thing by a threat—"Each term that your average falls below B, I will take away one of your

computer games"—will require the parent to be much more vigilant and active. The child will postpone bringing the grade report or will try to hide the games.

The concepts of reward and punishment are relative to those of some status quo. If the child has a perpetual right to the games, then taking one away is a punishment; if the games are temporarily assigned to the child on a term-by-term basis, then renewing the assignment for another term is a reward. Therefore you can change a threat into a promise or vice versa by changing the status quo. You can use this change to your own advantage when making a strategic move. If you want to achieve compellence, try to choose a status quo such that what you do when the other player acts to comply with your demand becomes a reward, and so you are using a compellent promise. To give a rather dramatic example, a mugger can convert the threat "If you don't give me your wallet, I will take out my knife and cut your throat" into the promise "Here is a knife at your throat; as soon as you give me your wallet I will take it away." But if you want to achieve deterrence, try to choose a status quo such that, if the other player acts contrary to your wishes, what you do is a punishment, and so you are using a deterrent threat.

## 6 ACQUIRING CREDIBILITY

We have emphasized the importance of credibility of strategic moves throughout, and we accompanied each example with some brief remarks about how credibility could be achieved in that particular context. Devices for achieving credibility are indeed often context specific, and there is a lot of art to discovering or developing such devices. Some general principles can help you organize your search.

We pointed out two broad approaches to credibility: (1) reducing your own future freedom of action in such a way that you have no choice but to carry out the action stipulated by your strategic move and (2) changing your own future payoffs in such a way that it becomes optimal for you to do what you stipulate in your strategic move. We now elaborate some practical methods for implementing each of these approaches.

### A. Reducing Your Freedom of Action

**I. AUTOMATIC FULFILLMENT** Suppose at stage 1 you relinquish your choice at stage 2 and hand it over to a mechanical device or similar procedure or mechanism that is programmed to carry out your committed, threatened, or promised action under the appropriate circumstances. You demonstrate to the other player that you have done so. Then he will be convinced that you have no freedom to change your mind, and your strategic move will be credible. The **doomsday device**, a

nuclear explosive device that would detonate and contaminate the whole world's atmosphere if the enemy launched a nuclear attack, is the best-known example, popularized by the early 1960s movies *Fail Safe* and *Dr. Strangelove*. Luckily, it remained in the realm of fiction. But automatic procedures that retaliate with import tariffs if another country tries to subsidize its exports to your country (countervailing duties) are quite common in the arena of trade policy.

**II. DELEGATION** A fulfillment device does not even have to be mechanical. You could delegate the power to act to another person or to an organization that is required to follow certain preset rules or procedures. In fact, that is how the countervailing duties work. They are set by two agencies of the U.S. government—the Commerce Department and the International Trade Commission—whose operating procedures are laid down in the general trade laws of the country.

An agent should not have his own objectives that defeat the purpose of his strategic move. For example, if one player delegates to an agent the task of inflicting threatened punishment and the agent is a sadist who enjoys inflicting punishment, then he may act even when there is no reason to act—that is, even when the second player has complied. If the second player suspects this, then the threat loses its effectiveness, because the punishment becomes a case of "damned if you do and damned if you don't."

Delegation devices are not complete guarantees of credibility. Even the doomsday device may fail to be credible if the other side suspects that you control an override button to prevent the risk of a catastrophe. And delegation and mandates can always be altered; in fact, the U.S. government has often set aside the stipulated countervailing duties and reached other forms of agreements with other countries so as to prevent costly trade wars.

**III. BURNING BRIDGES** Many invaders, from Xenophon in ancient Greece to William the Conqueror in England to Cortés in Mexico, are supposed to have deliberately cut off their own army's avenue of retreat to ensure that it will fight hard. Some of them literally burned bridges behind them, while others burned ships, but the device has become a cliché. Its most recent users in military contexts may have been the Japanese *kamikaze* pilots in World War II, who took only enough fuel to reach the U.S. naval ships into which they were to ram their airplanes. The principle even appears in the earliest known treatise on war, in a commentary attributed to Prince Fu Ch'ai: "Wild beasts, when they are at bay, fight desperately. How much more is this true of men! If they know there is no alternative they will fight to the death."<sup>5</sup>

Related devices are used in other high-stakes games. Although the European Monetary Union could have retained separate currencies and merely fixed

<sup>5</sup>Sun Tzu, *The Art of War*, trans. Samuel B. Griffith (Oxford: Oxford University Press, 1963), p. 110.



the exchange rates among them, a common currency was adopted precisely to make the process irreversible and thereby give the member countries a much greater incentive to make the union a success. (In fact, it is the extent of the necessary commitment that has kept some nations, Great Britain in particular, from agreeing to be part of the European Monetary Union.) It is not totally impossible to abandon a common currency and go back to separate national ones; it is just inordinately costly. If things get really bad inside the Union, one or more countries may yet choose to get out. As with automatic devices, the credibility of burning bridges is not an all-or-nothing matter, but one of degree.

**IV. CUTTING OFF COMMUNICATION** If you send the other player a message demonstrating your commitment and at the same time cut off any means for him to communicate with you, then he cannot argue or bargain with you to reverse your action. The danger in cutting off communication is that, if both players do so simultaneously, then they may make mutually incompatible commitments that can cause great mutual harm. Additionally, cutting off communication is harder to do with a threat, because you have to remain open to the one message that tells you whether the other player has complied and therefore whether you need to carry out your threat. In this age, it is also quite difficult for a person to cut himself off from all contact.

But players who are large teams or organizations can try variants of this device. Consider a labor union that makes its decisions at mass meetings of members. To convene such a meeting takes a lot of planning—reserving a hall, communicating with members, and so forth—and several weeks of time. A meeting is convened to decide on a wage demand. If management does not meet the demand in full, the union leadership is authorized to call a strike and then it must call a new mass meeting to consider any counteroffer. This process puts management under a lot of time pressure in the bargaining; it knows that the union will not be open to communication for several weeks at a time. Here, we see that cutting off communication for extended periods can establish some degree of credibility, but not absolute credibility. The union's device does not make communication totally impossible; it only creates several weeks of delay.

## B. Changing Your Payoffs

**I. REPUTATION** You can acquire a **reputation** for carrying out threats and delivering on promises. Such a reputation is most useful in a repeated game against the same player. It is also useful when playing different games against different players, if each of them can observe your actions in the games that you play with others. The circumstances favorable to the emergence of such a reputation are the same as those for achieving cooperation in the prisoners' dilemma,

and for the same reasons. The greater the likelihood that the interaction will continue and the greater the concern for the future relative to the present, the more likely the players will be to sacrifice current temptations for the sake of future gains. The players will therefore be more willing to acquire and maintain reputations.

In technical terms, this device links different games, and the payoffs of actions in one game are altered by the prospects of repercussions in other games. If you fail to carry out your threat or promise in one game, your reputation suffers and you get a lower payoff in other games. Therefore when you consider any one of these games, you should adjust your payoffs in it to take into consideration such repercussions on your payoffs in the linked games.

The benefit of reputation in ongoing relationships explains why your regular car mechanic is less likely to cheat you by doing an unnecessary or excessively costly or shoddy repair than is a random garage that you go to in an emergency. But what does your regular mechanic actually stand to gain from acquiring this reputation if competition forces him to charge a price so low that he makes no profit on any deal? His integrity in repairing your car must come at a price—you have to be willing to let him charge you a little bit more than the rates that the cheapest garage in the area might advertise.

The same reasoning also explains why, when you are away from home, you might settle for the known quality of a restaurant chain instead of taking the risk of going to an unknown local restaurant. And a department store that expands into a new line of merchandise can use the reputation that it has acquired in its existing lines to promise its customers the same high quality in the new line.

In games where credible promises by one or both parties can bring mutual benefit, the players can agree and even cooperate in fostering the development of reputation mechanisms. But if the interaction ends at a known finite time, there is always the problem of the endgame.

In the Middle East peace process that started in 1993 with the Oslo Accord, the early steps, in which Israel transferred some control over Gaza and small isolated areas of the West Bank to the Palestinian Authority and in which the latter accepted the existence of Israel and reduced its anti-Israel rhetoric violence, continued well for a while. But as the final stages of the process approached, mutual credibility of the next steps became problematic, and by 1998 the process stalled. Sufficiently attractive rewards could have come from the outside; for example, the United States or Europe could have given both parties contingent offers of economic aid or prospects of expanded commerce to keep the process going. The United States offered Egypt and Israel large amounts of aid in this way to achieve the Camp David Accords in 1978. But such rewards were not offered in the more recent situation and, at the date of this writing, prospects for progress do not look bright.

**II. DIVIDING THE GAME INTO SMALL STEPS** Sometimes a single game can be divided into a sequence of smaller games, thereby allowing the reputation mechanism to come into effect. In home-construction projects, it is customary to pay by installments as the work progresses. In the Middle East peace process, Israel would never have agreed to a complete transfer of the West Bank to the Palestinian Authority in one fell swoop in return for a single promise to recognize Israel and cease the terrorism. Proceeding in steps has enabled the process to go at least part of the way. But this again illustrates the difficulty of sustaining momentum as the endgame approaches.

**III. TEAMWORK** Teamwork is yet another way to embed one game into a larger game to enhance the credibility of strategic moves. It requires a group of players to monitor each another. If one fails to carry out a threat or a promise, others are required to inflict punishment on him; failure to do so makes them in turn vulnerable to similar punishment by others, and so on. Thus a player's payoffs in the larger game are altered in a way that makes adhering to the team's creed credible.

Many universities have academic honor codes that act as credibility devices for students. Examinations are not proctored by the faculty; instead, students are required to report to a student committee if they see any cheating. Then the committee holds a hearing and hands out punishment, as severe as suspension for a year or outright expulsion, if it finds the accused student guilty of cheating. Students are very reluctant to place their fellow students in such jeopardy. To stiffen their resolve, such codes include the added twist that failure to report an observed infraction is itself an offense against the code. Even then, the general belief is that the system works only imperfectly. A poll conducted at Princeton University last year found that only a third of students said that they would report an observed infraction, especially if they knew the guilty person.

**IV. IRRATIONALITY** Your threat may lack credibility because the other player knows that you are rational and that it is too costly for you to follow through with your threatened action. Therefore others believe you will not carry out the threatened action if you are put to the test. You can counter this problem by claiming to be irrational so that others will believe that your payoffs are different from what they originally perceived. Apparent irrationality can then turn into strategic rationality when the credibility of a threat is in question. Similarly, apparently irrational motives such as honor or saving face may make it credible that you will deliver on a promise even when tempted to renege.

The other player may see through such **rational irrationality**. Therefore if you attempt to make your threat credible by claiming irrationality, he will not readily believe you. You will have to acquire a reputation for irrationality, for example, by acting irrationally in some related game. You could also use one of the strategies discussed in Chapter 9 and do something that is a credible signal

of irrationality to achieve an equilibrium in which you can separate from the falsely irrational.

**V. CONTRACTS** You can make it costly to yourself to fail to carry out a threat or to deliver on a promise by signing a **contract** under which you have to pay a sufficiently large sum in that eventuality. If such a contract is written with sufficient clarity that it can be enforced by a court or some outside authority, the change in payoffs makes it optimal to carry out the stipulated action, and the threat or the promise becomes credible.

In regard to a promise, the other player can be the other party to the contract. It is in his interest that you deliver on the promise, so he will hold you to the contract if you fail to fulfill the promise. A contract to enforce a threat is more problematic. The other player does not want you to carry out the threatened action and will not enforce the contract unless he gets some longer-term benefit in associated games from being subject to a credible threat in this one. Therefore in regard to a threat, the contract has to be with a third party. But when you bring in a third party and a contract merely to ensure that you will carry out your threat if put to the test, the third party does not actually benefit from your failure to act as stipulated. The contract thus becomes vulnerable to any renegotiation that would provide the third-party enforcer with some positive benefits. If the other player puts you to the test, you can say to the third party, "Look, I don't want to carry out the threat. But I am being forced to do so by the prospect of the penalty in the contract, and you are not getting anything out of all this. Here is a real dollar in exchange for releasing me from the contract." Thus the contract itself is not credible; therefore neither is the threat. The third party must have its own longer-term reasons for holding you to the contract, such as wanting to maintain its reputation, if the contract is to be renegotiation proof and therefore credible.

Written contracts are usually more binding than verbal ones, but even verbal ones may constitute commitments. When George H. W. Bush said, "Read my lips; no new taxes," in the presidential campaign of 1988, the American public took this promise to be a binding contract; when Bush reneged on it in 1990, the public held that against him in the election of 1992.

**VI. BRINKMANSHIP** In the U.S.-Japan trade-policy game, we found that a threat might be too "large" to be credible. If a smaller but effective threat cannot be found in a natural way, the size of the large threat can be reduced to a credible level by making its fulfillment a matter of chance. The United States cannot credibly say to Japan, "If you don't keep your markets open to U.S. goods, we will not defend you if the Russians or the Chinese attack you." But it can credibly say, "If you don't keep your markets open to U.S. goods, the relations between our countries will deteriorate, which will create the risk that, if you are faced with

an invasion, Congress at that time will not sanction U.S. military involvement in your aid." As mentioned earlier, such deliberate creation of risk is called brinkmanship. This is a subtle idea, difficult to put into practice. Brinkmanship is best understood by seeing it in operation, and the detailed case study of the Cuban missile crisis in Chapter 15 serves just that purpose.

We have described several devices for making one's strategic moves credible and examined how well they work. In conclusion, we want to emphasize a feature common to the entire discussion. Credibility in practice is not an all-or-nothing matter but one of degree. Even though the theory is stark—rollback analysis shows either that a threat works or that it does not—practical application must recognize that between these polar extremes lies a whole spectrum of possibility and probability.

## 7 COUNTERING YOUR OPPONENT'S STRATEGIC MOVES

If your opponent can make a commitment or a threat that works to your disadvantage, then, before he actually does so, you may be able to make a strategic countermove of your own. You can do so by making his future strategic move less effective, for example, by removing its irreversibility or undermining its credibility. In this section, we examine some devices that can help achieve this purpose. Some are similar to devices that the other side can use for its own needs.

### A. Irrationality

Irrationality can work for the would-be receiver of a commitment or a threat just as well as it does for the other player. If you are known to be so irrational that you will not give in to any threat and will suffer the damage that befalls you when your opponent carries out that threat, then he may as well not make the threat in the first place, because having to carry it out will only end up hurting him, too. Everything that we said earlier about the difficulties of credibly convincing the other side of your irrationality holds true here as well.

### B. Cutting Off Communication

If you make it impossible for the other side to convey to you the message that it has made a certain commitment or a threat, then your opponent will see no point in doing so. Thomas Schelling illustrates this possibility with the story of a child who is crying too loudly to hear his parent's threats.<sup>6</sup> Thus it is pointless

<sup>6</sup>Thomas C. Schelling, *The Strategy of Conflict* (Oxford: Oxford University Press, 1960), p. 146.

for the parent to make any strategic moves; communication has effectively been cut off.

### C. Leaving Escape Routes Open

If the other side can benefit by burning bridges to prevent its retreat, you can benefit by dousing those fires or perhaps even by constructing new bridges or roads by which your opponent can retreat. This device was also known to the ancients. Sun Tzu said, "To a surrounded enemy, you must leave a way of escape." The intent is not actually to allow the enemy to escape. Rather, "show him there is a road to safety, and so create in his mind the idea that there is an alternative to death. Then strike."<sup>7</sup>

### D. Undermining Your Opponent's Motive to Uphold His Reputation

If the person threatening you says, "Look, I don't want to carry out this threat, but I must because I want to maintain my reputation with others," you can respond, "It is not in my interest to publicize the fact that you did not punish me. I am only interested in doing well in this game. I will keep quiet; both of us will avoid the mutually damaging outcome; and your reputation with others will stay intact." Similarly, if you are a buyer bargaining with a seller and he refuses to lower his price on the grounds that "if I do this for you, I would have to do it for everyone else," you can point out that you are not going to tell anyone else. This may not work; the other player may suspect that you would tell a few friends who would tell a few others, and so on.

### E. Salami Tactics

Salami tactics are devices used to whittle down the other player's threat in the way that a salami is cut—one slice at a time. You fail to comply with the other's wishes (whether for deterrence or compellence) to a very small degree so that it is not worth the other's while to carry out the comparatively more drastic and mutually harmful threatened action just to counter that small transgression. If that works, you transgress a little more, and a little more again, and so on.

You know this perfectly well from your own childhood. Schelling<sup>8</sup> gives a wonderful description of the process:

Salami tactics, we can be sure, were invented by a child. . . . Tell a child not to go in the water and he'll sit on the bank and submerge his bare feet; he is not yet "in" the water. Acquiesce, and he'll stand up; no more of him is in the

<sup>7</sup>Sun Tzu, *The Art of War*, pp. 109–110.

<sup>8</sup>Thomas C. Schelling, *Arms and Influence* (New Haven: Yale University Press, 1966), pp. 66–67.



water than before. Think it over, and he'll start wading, not going any deeper. Take a moment to decide whether this is different and he'll go a little deeper, arguing that since he goes back and forth it all averages out. Pretty soon we are calling to him not to swim out of sight, wondering whatever happened to all our discipline.

Salami tactics work particularly well against compellence, because they can take advantage of the *time* dimension. When your mother tells you to clean up your room "or else," you can put off the task for an extra hour by claiming that you have to finish your homework, then for a half day because you have to go to football practice, then for an evening because you can't possibly miss *The Simpsons* on TV, and so on.

To counter the countermove of salami tactics you must make a correspondingly graduated threat. There should be a scale of punishments that fits the scale of noncompliance or procrastination. This can also be achieved by gradually raising the risk of disaster, another application of brinkmanship.

### SUMMARY

Actions taken by players to fix the rules of later play are known as *strategic moves*. These first moves must be *observable* and *irreversible* to be true first moves, and they must be credible if they are to have their desired effect of altering the equilibrium outcome of the game. *Commitment* is an unconditional first move used to seize a first-mover advantage when one exists. Such a move usually entails committing to a strategy that would not have been one's equilibrium strategy in the original version of the game.

Conditional first moves such as *threats* and *promises* are *response rules* designed either to *deter* rivals' actions and preserve the status quo or to *compel* rivals' actions and alter the status quo. Threats carry the possibility of mutual harm but cost nothing if they work; threats that create only the risk of a bad outcome fall under the classification of *brinkmanship*. Promises are costly only to the maker and only if they are successful. Threats can be arbitrarily large, although excessive size compromises credibility, but promises are usually kept just large enough to be effective. If the implicit promise (or threat) that accompanies a threat (or promise) is not credible, players must make a move that combines both a promise and a threat and see to it that both components are credible.

Credibility must be established for any strategic move. There are a number of general principles to consider in making moves credible and a number of specific devices that can be used to acquire credibility. They generally work either by reducing your own future freedom to choose or by altering your own

payoffs from future actions. Specific devices of this kind include establishing a *reputation*, using teamwork, demonstrating apparent irrationality, burning bridges, and making *contracts*, although the acquisition of credibility is often context specific. Similar devices exist for countering strategic moves made by rival players.

### KEY TERMS

brinkmanship (363)  
commitment (364)  
compellence (365)  
contract (387)  
deterrence (365)  
doomsday device (382)  
irreversible (363)  
observable (363)

promise (365)  
rational irrationality (386)  
reputation (384)  
response rule (365)  
salami tactics (377)  
strategic moves (389)  
threat (365)

### SOLVED EXERCISES

- S1. "One could argue that the size of a promise is naturally bounded, while in principle a threat can be arbitrarily severe so long as it is credible (and error free)." First, briefly explain why the statement is true. Despite the truth of the statement, players might find that an arbitrarily severe threat might not be to their advantage. Explain why the latter statement is also true.
- S2. For each of the following three games, answer these questions:
- What is the equilibrium if neither player can use any strategic moves?
  - Can one player improve his payoff by using a strategic move (commitment, threat, or promise) or a combination of such moves? If so, which player makes what strategic move(s)?

(i)

		COLUMN	
		Left	Right
ROW	Up	0, 0	2, 1
	Down	1, 2	0, 0



(ii)

		COLUMN	
		Left	Right
ROW	Up	4, 3	3, 4
	Down	2, 1	1, 2

(iii)

		COLUMN	
		Left	Right
ROW	Up	4, 1	2, 2
	Down	3, 3	1, 4

S3. In the classic film *Mary Poppins*, the Banks children are players in a strategic game with a number of different nannies. In their view of the world, nannies are inherently harsh, and playing tricks on nannies is great fun. That is, they view themselves as playing a game in which the nanny moves first, showing herself to be either Harsh or Nice, and the children move second, choosing to be either Good or Mischievous. The nanny prefers to have Good children to take care of but is also inherently harsh, and so she gets her highest payoff of 4 from (Harsh, Good) and her lowest payoff of 1 from (Nice, Mischievous), with (Nice, Good) yielding 3 and (Harsh, Mischievous) yielding 2. The children similarly most prefer to have a Nice nanny and then to be Mischievous; they get their highest two payoffs when the nanny is Nice (4 if Mischievous, 3 if Good) and their lowest two payoffs when the nanny is Harsh (2 if Mischievous, 1 if Good).

- Draw the game tree for this game and find the subgame-perfect equilibrium in the absence of any strategic moves.
- In the film, before the arrival of Mary Poppins, the children write their own ad for a new nanny in which they state: "If you won't scold and dominate us, we will never give you cause to hate us; we won't hide your spectacles so you can't see, put toads in your bed, or pepper in your tea." Use the tree from part (a) to argue that this statement constitutes a promise. What would the outcome of the game be if the children keep their promise?
- What is the implied threat that goes with the promise in part (b)? Is that implied threat automatically credible? Explain your answer.
- How could the children make the promise in part (b) credible?

- Is the promise in part (b) compellent or deterrent? Explain your answer by referring to the status quo in the game—namely, what would happen in the absence of the strategic move.

S4. The following is an interpretation of the rivalry between the United States and the Soviet Union for geopolitical influence during the 1970s and 1980s.<sup>9</sup> Each side has the choice of two strategies: Aggressive and Restrained. The Soviet Union wants to achieve world domination, so being Aggressive is its dominant strategy. The United States wants to prevent the Soviet Union from achieving world domination; it will match Soviet aggressiveness with aggressiveness, and restraint with restraint. Specifically, the payoff table is:

		SOVIET UNION	
		Restrained	Aggressive
UNITED STATES	Restrained	4, 3	1, 4
	Aggressive	3, 1	2, 2

For each player, 4 is best and 1 is worst.

- Consider this game when the two countries move simultaneously. Find the Nash equilibrium.
  - Next consider three different and alternative ways in which the game could be played with sequential moves: (i) The United States moves first, and the Soviet Union moves second. (ii) The Soviet Union moves first, and the United States moves second. (iii) The Soviet Union moves first, and the United States moves second, but the Soviet Union has a further move in which it can change its first move. For each case, draw the game tree and find the subgame-perfect equilibrium.
  - What are the key strategic matters (commitment, credibility, and so on) for the two countries?
- S5. Consider the following games. In each case, (i) identify which player can benefit from making a strategic move, (ii) identify the nature of the strategic move appropriate for this purpose, (iii) discuss the conceptual and practical difficulties that will arise in the process of making this move credible, and (iv) discuss whether and how the difficulties can be overcome.

<sup>9</sup>We thank political science professor Thomas Schwartz at UCLA for the idea for this exercise.

- (a) The other countries of the European Monetary Union (France, Germany, and so on) would like Britain to join the common currency and the common central bank.
- (b) The United States would like North Korea to stop exporting missiles and missile technology to countries such as Iran and would like China to join the United States in working toward this aim.
- (c) The United Auto Workers would like U.S. auto manufacturers not to build plants in Mexico and would like the U.S. government to restrict imports of autos made abroad.

### UNSOLVED EXERCISES

- U1. In a scene from the movie *Manhattan Murder Mystery*, Woody Allen and Diane Keaton are at a hockey game in Madison Square Garden. She is obviously not enjoying herself, but he tells her: "Remember our deal. You stay here with me for the entire hockey game, and next week I will come to the opera with you and stay until the end." Later, we see them coming out of the Met into the deserted Lincoln Center Plaza while inside the music is still playing. Keaton is visibly upset: "What about our deal? I stayed to the end of the hockey game, and so you were supposed to stay till the end of the opera." Allen answers: "You know I can't listen to too much Wagner. At the end of the first act, I already felt the urge to invade Poland." Comment on the strategic choices made here by using your knowledge of the theory of strategic moves and credibility.
- U2. Consider a game between a parent and a child. The child can choose to be good (G) or bad (B); the parent can punish the child (P) or not (N). The child gets enjoyment worth a 1 from bad behavior, but hurt worth  $-2$  from punishment. Thus a child who behaves well and is not punished gets a 0; one who behaves badly and is punished gets  $1 - 2 = -1$ ; and so on. The parent gets  $-2$  from the child's bad behavior and  $-1$  from inflicting punishment.
- (a) Set up this game as a simultaneous-move game, and find the equilibrium.
  - (b) Next, suppose that the child chooses G or B first and that the parent chooses its P or N after having observed the child's action. Draw the game tree and find the subgame-perfect equilibrium.
  - (c) Now suppose that before the child acts, the parent can commit to a strategy. For example, the threat "P if B" ("If you behave badly, I will punish you"). How many such strategies does the parent have? Write the table for this game. Find all pure-strategy Nash equilibria.

- (d) How do your answers to parts (b) and (c) differ? Explain the reason for the difference.

U3. The general strategic game in Thucydides' history of the Peloponnesian War has been expressed in game-theoretic terms by Professor William Charron of St. Louis University.<sup>10</sup> Athens had acquired a large empire of coastal cities around the Aegean as part of its leadership role in defending the Greek world from Persian invasions. Sparta, fearing Athenian power, was contemplating war against Athens. If Sparta decided against war, Athens would have to decide whether to retain or relinquish its empire. But Athens in turn feared that if it gave independence to the cities, they could choose to join Sparta in a greatly strengthened alliance against Athens and receive very favorable terms from Sparta for doing so. Thus there are three players, Sparta, Athens, and Small cities, who move in this order. There are four outcomes, and the payoffs are as follows (4 being best):

Outcome	Sparta	Athens	Small cities
War	2	2	2
Athens retains empire	1	4	1
Small cities join Sparta	4	1	4
Small cities stay independent	3	3	3

- (a) Draw the game tree and find the rollback equilibrium. Is there another outcome that is better for all players?
  - (b) What strategic move or moves could attain the better outcome? Discuss the credibility of such moves.
- U4. It is possible to reconfigure the payoffs in the game in Exercise S3 so that the children's statement in their ad is a threat, rather than a promise.
- (a) Redraw the tree from part (a) of Exercise S3 and fill in payoffs for both players so that the children's statement becomes a *threat* in the full technical sense.
  - (b) Define the status quo in your game, and determine whether the threat is deterrent or compellent.
  - (c) Explain why the threatened action is not automatically credible, given your payoff structure.
  - (d) Explain why the implied promise is automatically credible.

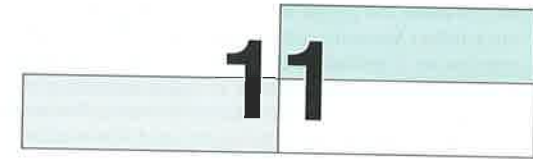
<sup>10</sup>William C. Charron, "Greeks and Games: Forerunners of Modern Game Theory," *Forum for Social Economics*, vol. 29, no. 2 (Spring 2000), pp. 1–32.

- (e) Explain why the children would want to make a threat in the first place, and suggest a way in which they might make their threatened action credible.

U5. Answer the questions in Exercise S5 for the following situations:

- (a) The students at your university or college want to prevent the administration from raising tuition.  
 (b) Most participants, as well as outsiders, want to achieve a durable peace in Afghanistan, Iraq, Israel, and Palestine.  
 (c) Nearly all nations of the world want Iran to shut down its nuclear program.

U6. Write a brief description of a game in which you have participated, entailing strategic moves such as a commitment, threat, or promise and paying special attention to the essential aspect of credibility. Provide an illustration of the game if possible, and explain why the game that you describe ended as it did. Did the players use sound strategic thinking in making their choices?



## The Prisoners' Dilemma and Repeated Games

IN THIS CHAPTER, we continue our study of broad classes of games with an analysis of the prisoners' dilemma game. It is probably the classic example of the theory of strategy and its implications for predicting the behavior of game players, and most people who learn only a little bit of game theory learn about it. Even people who know *no* game theory may know the basic story behind this game or they may have at least heard that it exists. The prisoners' dilemma is a game in which each player has a dominant strategy, but the equilibrium that arises when all players use their dominant strategies provides a worse outcome for every player than would arise if they all used their dominated strategies instead. The paradoxical nature of this equilibrium outcome leads to several more complex questions about the nature of the interactions that only a more thorough analysis can hope to answer. The purpose of this chapter is to provide that additional thoroughness.

We already considered the prisoners' dilemma in Section 3 of Chapter 4. There we took note of the curious nature of the equilibrium that is actually a "bad" outcome for the players. The "prisoners" can find another outcome that both prefer to the equilibrium outcome, but they find it difficult to bring about. The focus of this chapter is the potential for achieving that better outcome. That is, we consider whether and how the players in a prisoners' dilemma can attain and sustain their mutually beneficial cooperative outcome, overcoming their separate incentives to defect for individual gain. We first review the standard prisoners' dilemma game and then develop four categories of solutions. The first and most important method of solution consists of repetition of the