# 8

# Uncertainty and Information

IN CHAPTER 2, we mentioned different ways in which uncertainty can arise in a game (external and strategic) and ways in which players can have limited information about aspects of the game (imperfect and incomplete, symmetric and asymmetric). We have already encountered and analyzed some of these. Most notably, in simultaneous-move games, each player does not know the actions the other is taking; this is strategic uncertainty. In Chapter 6, we saw that strategic uncertainty gives rise to asymmetric and imperfect information, because the different actions taken by one player must be lumped into one information set for the other player. In Chapters 4 and 7, we saw how such strategic uncertainty is handled by having each player formulate beliefs about the other's action (including beliefs about the probabilities with which different actions may be taken when mixed strategies are played) and by applying the concept of Nash equilibrium, in which such beliefs are confirmed. In this chapter we focus on some further ways in which uncertainty and informational limitations arise in games.

We begin by examining various strategies that individuals and societies can use for coping with the imperfect information generated by external uncertainty or risk. Recall that external uncertainty is about matters outside any player's control but affecting the payoffs of the game; weather is a simple example. Here we show the basic ideas behind diversification, or spreading, of risk by an individual player and pooling of risk by multiple players. These strategies can benefit everyone, although the division of total gains among the participants can be unequal; therefore, these situations contain a mixture of common interest and conflict.

We then consider the informational limitations that often exist in situations with strategic interdependence. Information in a game is *complete* only if all of the rules of the game—the strategies available to all players and the payoffs of each player as functions of the strategies of all players—are fully known by all players and, moreover, are common knowledge among them. By this exacting standard, most games in reality have *incomplete information*. Moreover, the incompleteness is usually *asymmetric*: each player knows his own capabilities and payoffs much better than he knows those of other players. As we pointed out in Chapter 2, manipulation of the information becomes an important dimension of strategy in such games. In this chapter, we will discuss when information can or cannot be communicated verbally in a credible manner. We will also examine other strategies designed to convey or conceal one's own information and to elicit another player's information. We spoke briefly of some such strategies—namely, screening and signaling—in Chapters 1 and 2; here, we study those in more detail.

Of course, players in many games would also like to manipulate the actions of others. Managers would like their workers to work hard and well; insurance companies would like their policyholders to exert care to reduce the risk that is being insured. If information were perfect, the actions would be observable. Workers' pay could be made contingent on the quality and quantity of their effort; payouts to insurance policyholders could be made contingent on the care they exercised. But in reality these actions are difficult to observe; that creates a situation of imperfect asymmetric information, commonly called **moral hazard**. Thus, the counterparties in these games have to devise various indirect methods to give incentives to influence others' actions in the right direction.

The study of the topic of information and its manipulation in games has been very active and important in recent decades. It has shed new light on many previously puzzling matters in economics, such as the nature of incentive contracts, the organization of companies, markets for labor and for durable goods, government regulation of business, and myriad others.[1] More recently, political scientists have used the same concepts to explain phenomena such as the relation of tax- and expenditures-policy changes to elections, as well as the delegation of legislation to committees. These ideas have also spread to biology, where evolutionary game theory explains features such as the peacock's large and ornate tail as a signal. Perhaps even more important, you will recognize the important role that signaling and screening play in your daily interaction with family, friends, teachers, coworkers, and so on, and you will be able to improve your strategies in these games.

[1] The pioneers of the theory of asymmetric information in economics—George Akerlof, Michael Spence, and Joseph Stiglitz—received the 2001 Nobel Prize in economics for these contributions.

Although the study of information clearly goes well beyond consideration of external uncertainty and the basic concepts of signaling and screening, we focus only on those few topics in this chapter. We will return to the analysis of information and its manipulation in Chapter 13, however. There we will use the methods developed here to study the design of mechanisms to provide incentives to and elicit information from other players who have some private information.

# 1 IMPERFECT INFORMATION: DEALING WITH RISK

Imagine that you are a farmer subject to the vagaries of weather. If the weather is good for your crops, you will have an income of \$160,000. If it is bad, your income will be only \$40,000. The two possibilities are equally likely (probability 1/2, or 0.5, or 50% each). Therefore, your average or expected income is \$100,000 ($= 1/2 \times 160{,}000 + 1/2 \times 40{,}000$), but there is considerable risk around this average value.

What can you do to reduce the risk that you face? You might try a crop that is less subject to the vagaries of weather, but suppose you have already done all such things that are under your individual control. Then you might be able to reduce your income risk further by getting someone else to accept some of the risk. Of course, you must give the other person something else in exchange. This quid pro quo usually takes one of two forms: cash payment or a mutual exchange or sharing of risks.

## A. Sharing of Risk

We begin with an analysis of the possibility of risk sharing for mutual benefit. Suppose you have a neighbor who faces a similar risk but gets good weather exactly when you get bad weather and vice versa. (Suppose you live on opposite sides of an island, and rain clouds visit one side or the other but not both.) In technical jargon, *correlation* is a measure of alignment between any two uncertain quantities—in this discussion, between one person's risk and another's. Thus, we would say that your neighbor's risk is totally **negatively correlated** with yours. The combined income of you and your neighbor is \$200,000, no matter what the weather: it is totally risk free. You can enter into a contract that gets each of you \$100,000 for sure: you promise to give him \$60,000 in years when you are lucky, and he promises to give you \$60,000 in years when he is lucky. You have eliminated your risks by combining them.

Currency swaps provide a good example of negative correlation of risk in real life. A U.S. firm exporting to Europe gets its revenues in euros, but it is interested in its dollar profits, which depend on the fluctuating euro-dollar exchange rate. Conversely, a European firm exporting to the United States faces similar uncertainty about its profits in euros. When the euro falls relative to the dollar, the U.S. firm's euro revenues convert into fewer dollars, and the European firm's dollar revenues convert into more euros. The opposite happens when the euro rises relative to the dollar. Thus, fluctuations in the exchange rate generate negatively correlated risks for the two firms. Both can reduce these risks by contracting for an appropriate swap of their revenues.

Even without such perfect negative correlation, risk sharing has some benefit. Return to your role as an island farmer and suppose you and your neighbor face risks that are independent from each other, as if the rain clouds could toss a separate coin to decide which one of you to visit. Then there are four possible outcomes, each with a probability of 1/4. The incomes you and your neighbor earn in these four cases are illustrated in panel a of Figure 8.1. However, suppose the two of you were to make a contract to share and share alike; then your incomes would be those shown in panel b of Figure 8.1. Although your average (expected) income in each table is \$100,000, without the sharing contract, you each would have \$160,000, or \$40,000 with probabilities of 1/2 each. With the contract, you each would have \$160,000 with probability 1/4, \$100,000 with probability 1/2, and \$40,000 with probability 1/4. Thus, for each of you, the contract has reduced the probabilities of the two extreme outcomes from 1/2 to 1/4 and increased the probability of the middle outcome from 0 to 1/2. In other words, the contract has reduced the risk for each of you.

In fact, as long as your incomes are not totally **positively correlated**—that is, as long as your luck does not move in perfect tandem—you can both reduce your risks by sharing them. And if there are more than two of you with some degree of independence in your risks, then the law of large numbers makes possible even greater reduction in the risk of each. That is exactly what insurance companies do: by combining the similar but independent risks of many people, an insurance company is able to compensate any one of them when he suffers a large loss. It is also the basis of portfolio diversification: by dividing your

| | | NEIGHBOR | |
|---|---|---|---|
| | | Lucky | Not |
| YOU | Lucky | 160,000, 160,000 | 160,000, 40,000 |
| | Not | 40,000, 160,000 | 40,000, 40,000 |

**(a)** Without sharing

| | | NEIGHBOR | |
|---|---|---|---|
| | | Lucky | Not |
| YOU | Lucky | 160,000, 160,000 | 100,000, 100,000 |
| | Not | 100,000, 100,000 | 40,000, 40,000 |

**(b)** With sharing

**FIGURE 8.1** Sharing Income Risk

wealth among many different assets with different kinds and degrees of risk, you can reduce your total exposure to risk.

However, such arrangements for risk sharing depend on public observability of outcomes and enforcement of contracts. Otherwise, each farmer has the temptation to pretend to have suffered bad luck or simply to renege on the deal and refuse to share when he has good luck. An insurance company may similarly falsely deny claims, but its desire to maintain its reputation in ongoing business may check such reneging.

Here we consider another issue. In the discussion above, we simply assumed that sharing meant equal shares. That seems natural, because you and your farmer-neighbor are in identical situations. But you may have different strategic skills and opportunities, and one may be able to do better than the other in bargaining or contracting.

To understand this, we must recognize the basic reason that farmers want to make such sharing arrangements, namely, that they are averse to risk. As we explain in the appendix to this chapter, attitudes toward risk can be captured by using nonlinear scales to convert money incomes into "utility" numbers. The square root function is a simple example of such a scale that reflects risk aversion, and we apply it here.

When you bear the full risk of getting \$160,000 or \$40,000 with probabilities 1/2 each, your expected (probability-weighted average) utility is

$$1/2 \times \sqrt{160{,}000} + 1/2 \times \sqrt{40{,}000} = 1/2 \times 400 + 1/2 \times 200 = 300.$$

The riskless income that will give you the same utility is the number whose square root is 300, that is, \$90,000. This is less than the average money income you have, namely \$100,000. The difference, \$10,000, is the maximum money sum you would be willing to pay as a price for eliminating the risk in your income entirely. Your neighbor faces a risk of equal magnitude, so if he has the same utility scale, he is also willing to pay the same maximum amount to eliminate all of his risk.

Consider the situation where your risks are perfectly negatively correlated, so that the sum of your two incomes is \$200,000 no matter what. You make your neighbor the following offer: I will pay you \$90,001 − \$40,000 = \$50,001 when your luck is bad, if you pay me \$160,000 − \$90,001 = \$69,999 when your luck is good. That leaves your neighbor with \$90,001 whether his luck is good or bad (\$160,000 − \$69,999 in the former situation and \$40,000 + \$50,001 in the latter situation). He prefers this situation to facing the risk. When his luck is good, yours is bad; you have \$40,000 of your own but receive \$69,999 from him for a total of \$109,999. When his luck is bad, yours is good; you have \$160,000 of your own but pay him \$50,001, leaving you with \$109,999. You have also eliminated your own risk. Both of you are made better off by this deal, but you have collared almost all the gain.

Of course, your neighbor could have made you the opposite offer. And a whole range of intermediate offers, involving more equitable sharing of the gains from risk sharing, is also conceivable. Which of these will prevail? That depends on the parties' bargaining power, as we will see in more detail in Chapter 17; the full range of mutually beneficial risk-sharing outcomes will correspond to the efficient frontier of negotiation in the bargaining game between the players.

## B. Paying to Reduce Risk

Now we consider the possibility of trading of risks for cash. Suppose you are the farmer facing the same risk as before. But now your neighbor has a sure income of \$100,000. You face a lot of risk, and he faces none. He may be willing to take a little of your risk for a price that is agreeable to both of you. We just saw that \$10,000 is the maximum "insurance premium" you would be willing to pay to get rid of your risk completely. Would your neighbor accept this as payment for eliminating your risk? In effect, he is taking over control of his riskless income plus your risky income, that is, \$100,000 + \$160,000 = \$260,000 if your luck is good and \$100,000 + \$40,000 = \$140,000 if your luck is bad. He gives you \$90,000 in either eventuality, thus leaving him with \$170,000 or \$50,000 with equal probabilities. His expected utility is then

$$1/2 \times \sqrt{170{,}000} + 1/2 \times \sqrt{50{,}000} = 1/2 \times 412.31 + 1/2 \times 223.61 = 317.96.$$

His utility if he did not trade with you would be $\sqrt{100{,}000} = 316.23$, so the trade makes him just slightly better off. The range of mutually beneficial deals in this case is very narrow, so the outcome is almost determinate, but there is not much scope for mutual benefit if you aim to trade all of your risk away.

What about a partial trade? Suppose you pay him $x$ if your luck is good, and he pays you $y$ if your luck is bad. For this to raise expected utilities for both of you, we need both of the following inequalities to hold:

$$1/2 \times \sqrt{160{,}000 - x} + 1/2\sqrt{40{,}000 + y} > 300,$$

$$1/2 \times \sqrt{100{,}000 + x} + 1/2 \times \sqrt{100{,}000 - y} > \sqrt{100{,}000}.$$

As an example, suppose $y = 10{,}000$. Then the second inequality yields $x > 10{,}526.67$, and the first yields $x < 18{,}328.16$. The first value for $x$ is the minimum payment he requires from you to be willing to make the trade, and the second value for $x$ is the maximum you are willing to pay to him to have him assume your risk. Thus, there is a substantial range for mutually beneficial trade and bargaining.

What if your neighbor is risk neutral, that is, concerned solely with expected monetary magnitudes? Then the deal must satisfy

$$1/2 \times (100{,}000 + x) + 1/2 \times (100{,}000 - y) > 100{,}000,$$

or simply $x > y$, to be acceptable to him. Almost-full insurance, where you pay him \$60,001 if your luck is good and he pays you \$59,999 if your luck is bad, is possible. This is the situation where you reap all the gain from the trade in risks.

If your "neighbor" is actually an insurance company, the company can be close to risk neutral because it is combining numerous such risks and is owned by well-diversified investors for each of whom this business is only a small part of their total risk. Then the fiction of a friendly, risk-neutral, good neighbor can become a reality. And if insurance companies compete for your business, the insurance market can offer you almost-complete insurance at a price that leaves almost all of the gain with you.

Common to all such arrangements is the idea that mutually beneficial deals can be struck whereby, for a suitable price, someone facing less risk takes some of the risk off the shoulders of someone else who faces more. In fact, the idea that a price and a market for risk exist is the basis for almost all of the financial arrangements in a modern economy. Stocks and bonds, as well as all of the complex financial instruments, such as derivatives, are just ways of spreading risk to those who are willing to bear it for the lowest asking price. Many people think these markets are purely forms of gambling. In a sense, they are. But those who start out with the least risk take the gambles, perhaps because they have already diversified in the way that we saw earlier. And the risk is sold or shed by those who are initially most exposed to it. This enables the latter to be more adventurous in their enterprises than they would be if they had to bear all of the risk themselves. Thus, financial markets promote entrepreneurship by facilitating risk trading.

Here we have only considered sharing of a given total risk. In practice, people may be able to take actions to reduce that total risk: a farmer can guard crops against frosts, and a car owner can drive more carefully to reduce the risk of an accident. If such actions are not publicly observable, the game will be one of imperfect information, raising the problem of moral hazard that we mentioned in the introduction: people who are well insured will lack the incentive to reduce the risk they face. We will look at such problems, and the design of mechanisms to cope with them, in Chapter 13.

## C. Manipulating Risk in Contests

The farmers above faced risk due to the weather rather than from any actions of their own or of other farmers. If the players in a game can affect the risk they or others face, then they can use such manipulation of risk strategically. A prime example is contests such as research and development races between companies to develop and market new information technology or biotech products; many sports contests have similar features.

The outcome of sports and related contests is determined by a mixture of skill and chance. You win if

$$\text{Your skill} + \text{your luck} > \text{rival's skill} + \text{rival's luck}$$

or

$$\text{Your luck} - \text{rival's luck} > \text{rival's skill} - \text{your skill}.$$

Denote the left-hand side by the symbol *L;* it measures your "luck surplus." *L* is an uncertain magnitude; suppose its probability distribution is a normal, or bell, curve, as illustrated by the black curve in Figure 8.2. At any point on the horizontal axis, the height of the curve represents the probability that *L* takes on that value. Thus, the area under this curve between any two points on the horizontal axis equals the probability that *L* lies between those points. Suppose your rival has more skill, so you are an underdog. Your "skill deficit," which equals the difference between your rival's skill and your skill, is therefore positive, as shown by the point *S*. You win if your luck surplus, *L*, exceeds your skill deficit, *S*. Therefore, the area under the curve to the right of the point *S*, which is shaded in gray in Figure 8.2, represents your probability of winning. If you make the situation chancier, the bell curve will be flatter, like the blue curve in Figure 8.2, because the probability of relatively high and low values of *L* increases while the probability of moderate values decreases. Then the area under the curve to the right of *S* also increases. In Figure 8.2, the area under the original bell curve is shown by gray shading, and the larger area under the flatter bell curve by the blue hatching. As the underdog, you should therefore adopt a strategy that flattens the curve. Conversely, if you are the favorite, you should try to reduce the element of chance in the contest.

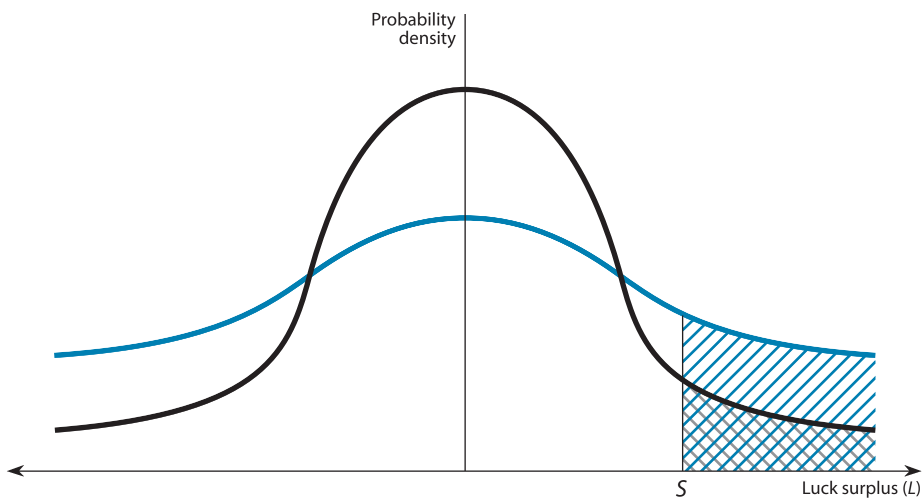


**FIGURE 8.2** The Effect of Greater Risk on the Chances of Winning

Thus, we should see underdogs or those who have fallen behind in a long race try unusual or risky strategies: it is their only chance to get level or ahead. In contrast, favorites or those who have stolen a lead will play it safe. A practical piece of advice based on this principle: if you want to challenge someone who is a better player than you to a game of tennis, choose a windy day.

You may stand to benefit by manipulating not just the amount of risk in your strategy, but also the correlation between the risks. The player who is ahead will try to choose a correlation as high and as positive as possible: then, whether his own luck is good or bad, the luck of his opponent will be the same and his lead protected. Conversely, the player who is behind will try to find a risk as uncorrelated with that of his opponent as possible. It is well known that in a two-sailboat race, the boat that is behind should try to steer differently from the boat ahead, and the boat ahead should try to imitate all the tacks of the one behind.[2]

## 2 ASYMMETRIC INFORMATION: BASIC IDEAS

In many games, one or some of the players may have an advantage of knowing with greater certainty what has happened or what will happen. Such advantages, or asymmetries of information, are common in actual strategic situations. At the most basic level, each player may know his own preferences or payoffs—for example, risk tolerance in a game of brinkmanship, patience in bargaining, or peaceful or warlike intentions in international relations—quite well but those of the other players much more vaguely. The same is true for a player's knowledge of his own innate characteristics (such as the skill of an employee or the riskiness of an applicant for auto or health insurance). And sometimes the actions available to one player—for example, the weaponry and readiness of a country—are not fully known to other players. Finally, some actual outcomes (such as the actual dollar value of loss to an insured homeowner in a flood or an earthquake) may be observed by one player but not by others.

By manipulating what the other players know about your abilities and preferences, you can affect the equilibrium outcome of a game. Therefore, such manipulation of asymmetric information itself becomes a game of strategy. You may think that each player will always want to conceal his own information and elicit information from the others, but that is not so. Here is a list of various possibilities, with examples. The better-informed player may want to do one of the following:

[2] Avinash Dixit and Barry Nalebuff, *Thinking Strategically* (New York: W. W. Norton & Company, 1991), give a famous example of the use of this strategy in sailboat racing. For a more general theoretical discussion, see Luis Cabral, "R&D Competition When the Firms Choose Variance," *Journal of Economics and Management Strategy*, vol. 12, no. 1 (Spring 2003), pp. 139–50.

1. *Conceal information* or *reveal misleading information*. When mixing moves in a zero-sum game, you don't want the other player to see what you have done; you bluff in poker to mislead others about your cards.
2. *Reveal selected information truthfully.* When you make a strategic move, you want others to see what you have done so that they will respond in the way you desire. For example, if you are in a tense situation but your intentions are not hostile, you want others to know this credibly, so that there will be no unnecessary fight.

Similarly, the less-informed player may want to do one of the following:

1. *Elicit information* or *filter truth from falsehood*. An employer wants to find out the skill of a prospective employee and the effort of a current employee. An insurance company wants to know an applicant's risk class, the amount of a claimant's loss, and any contributory negligence by the claimant that would reduce its liability.
2. *Remain ignorant.* Being unable to know your opponent's strategic move can immunize you against his commitments and threats. Top-level politicians or managers often benefit from having such "credible deniability."

In most cases, we will find that words alone do not suffice to convey credible information; rather, actions speak louder than words. Even actions may not convey information credibly if they are too easily performed by any random player. In general, however, the less-informed players should pay attention to what a better-informed player does, not to what he says. And knowing that the others will interpret actions in this way, the better-informed player should in turn try to manipulate his actions for their information content.

When you are playing a strategic game, you may find that you have information that other players do not. You may have information that is "good" (for yourself) in the sense that, if the other players knew this information, they would alter their actions in a way that would increase your payoff. You know that you are a nonsmoker, for example, and should qualify for lower life-insurance premiums. Or you may have "bad" information whose disclosure would cause others to act in a way that would hurt you. You cheated your way through college, for example, and don't deserve to be admitted to a prestigious law school. You know that others will infer your information from your actions. Therefore, you try to think of, and take, actions that will induce them to believe your information is good. Such actions are called **signals**, and the strategy of using them is called **signaling**. Conversely, if others are likely to conclude that your information is bad, you may be able to stop them from making this inference by confusing them. This strategy, called **signal jamming**, is typically a mixed strategy, because the randomness of mixed strategies makes inferences imprecise.

If other players know more than you do or take actions that you cannot directly observe, you can use strategies that reduce your informational disadvantage. The strategy of making another player act so as to reveal his information is called **screening**, and specific methods used for this purpose are called **screening devices**.[3]

Because a player's private information often consists of knowledge of his own abilities or preferences, it is useful to think of players who come to a game possessing different private information as different **types**. When credible signaling works, in the equilibrium of the game the less-informed players will be able to infer the information of the more-informed ones correctly from the actions; the law school, for example, will admit only the truly qualified applicants. Another way to describe the outcome is to say that in equilibrium, the different types are correctly revealed or separated. Therefore, we call this a **separating equilibrium**. In some cases, however, one or more types may successfully mimic the actions of other types, so that the uninformed players cannot infer types from actions and cannot identify the different types; insurance companies, for example, may offer only one kind of life insurance policy. Then, in equilibrium we say the types are pooled together, and we call this a **pooling equilibrium**. When studying games of incomplete information, we will see that identifying the kind of equilibrium that occurs is of primary importance.

## 3 DIRECT COMMUNICATION, OR "CHEAP TALK"

The simplest way to convey information to others would seem to be to tell them; likewise, the simplest way to elicit information would seem to be to ask. But in a game of strategy, players should be aware that others may not tell the truth and, likewise, that their own assertions may not be believed by others. That is, the *credibility* of mere words may be questionable. It is a common saying that talk is cheap; indeed, direct communication has zero or negligible *direct* cost. However, it can *indirectly* affect the outcome and payoffs of a game by changing one players's beliefs about another player's actions or by influencing the selection of one equilibrium out of multiple equilibria. Direct communication that has no direct cost has come to be called *cheap talk* by game theorists, and the equilibrium achieved by using direct communication is termed a **cheap talk equilibrium**.

[3] A word of warning: Don't confuse screening with signal jamming. In ordinary language, the word *screening* can have different meanings. The one used in game theory is that of testing or scrutinizing. Thus, a less-informed player uses screening to find out what a better-informed player knows. For the alternative sense of screening—that is, concealing—the game-theoretic term is signal jamming. Thus, a better-informed player uses a signal-jamming action to prevent the less-informed player from correctly inferring the truth from the action (that is, from screening the better-informed player).

## A. Perfectly Aligned Interests

Direct communication of information works well if the players' interests are well aligned. The assurance game first introduced in Chapter 4 provides the most extreme example of this. We reproduce its payoff table (Figure 4.11) as Figure 8.3.

The interests of Harry and Sally are perfectly aligned in this game; they both want to meet, and both prefer meeting in Local Latte. The problem is that the game is played noncooperatively; they are making their choices independently, without knowledge of what the other is choosing. But suppose that Harry is given an opportunity to send a message to Sally (or Sally is given an opportunity to ask a question and Harry replies) before their choices are made. If Harry's message (or reply; we will not keep repeating this) is "I am going to Local Latte," Sally has no reason to think he is lying.[4] If she believes him, she should choose Local Latte, and if he believes she will believe him, it is equally optimal for him to choose Local Latte, making his message truthful. Thus, direct communication very easily achieves the mutually preferable outcome. This is indeed the reason that, when we considered this game in Chapter 4, we had to construct an elaborate scenario in which such communication was infeasible; recall that the two were in separate classes until the last minute before their meeting and did not have their cell phones.

Let us examine the outcome of allowing direct communication in the assurance game more precisely in game-theoretic terms. We have created a two-stage game. In the first stage, only Harry acts, and his action is his message to Sally. In the second stage, the original simultaneous-move game is played. In the full two-stage game, we have a rollback equilibrium where the strategies (complete plans of action) are as follows. The second-stage action plans for both players are: "If Harry's first-stage message was 'I am going to Starbucks,' then choose Starbucks; if Harry's first-stage message was 'I am going to Local Latte,' then

| | | SALLY | |
|---|---|---|---|
| | | Starbucks | Local Latte |
| HARRY | Starbucks | 1, 1 | 0, 0 |
| | Local Latte | 0, 0 | 2, 2 |

**FIGURE 8.3** Assurance

[4] This reasoning assumes that Harry's payoffs are as stated, and that this fact is common knowledge between the two. If Sally suspects that Harry wants her to go to Local Latte so he can go to Starbucks to meet another girlfriend, her strategy will be different! Analysis of games of asymmetric information thus depends on how many different possible "types" of players are actually conceivable.

choose Local Latte." (Remember that players in sequential games must specify *complete* plans of action.) The first-stage action for Harry is to send the message "I am going to Local Latte." Verification that this is indeed a rollback equilibrium of the two-stage game is easy, and we leave it to you.

However, this equilibrium where cheap talk "works" is not the only rollback equilibrium of this game. Consider the following strategies: the second-stage action plan for each player is to go to Starbucks regardless of Harry's first-stage message; and Harry's first-stage message can be anything. We can verify that this also is indeed a rollback equilibrium. Regardless of Harry's first-stage message, if one player is going to Starbucks, then it is optimal for the other player to go there also. Thus, in each of the second-stage subgames that could arise—one after each of the two messages that Harry could send—both choosing Starbucks is a Nash equilibrium of the subgame. Then, in the first stage, Harry, knowing his message is going to be disregarded, is indifferent about which message he sends.

The cheap talk equilibrium—where Harry's message is not disregarded—yields higher payoffs, and we might normally think that it would be the one selected as a focal point. However, there may be reasons of history or culture that favor the other equilibrium. For example, for some reasons quite extraneous to this particular game, Harry may have a reputation for being totally unreliable. He might be a compulsive practical joker or just absent minded. Then people might generally disregard his statements and, knowing this to be the usual state of affairs, Sally might not believe this particular one.

Such problems exist in all communication games. They always have alternative equilibria where the communication is disregarded and therefore irrelevant. Game theorists call these **babbling equilibria**. Having noted that they exist, however, we will focus on the cheap talk equilibria, where communication does have some effect.

## B. Totally Conflicting Interests

The credibility of direct communication depends on the degree of alignment of players' interests. As a dramatic contrast with the assurance game example, consider a game where the players' interests are totally in conflict—namely, a zero-sum game. A good example is the tennis point in Figure 4.14 from Chapter 4; we reproduce its payoff matrix as Figure 8.4. Remember that the payoffs are Evert's success percentages. Remember also that this game has only a mixed-strategy Nash equilibrium (derived in Chapter 7); Evert's expected payoff in this equilibrium is 62.

Now suppose that we construct a two-stage game. In the first stage, Evert is given an opportunity to send a message to Navratilova. In the second stage, the simultaneous-move game of Figure 8.4 is played. What will be the rollback equilibrium?

| | | NAVRATILOVA | |
|---|---|---|---|
| | | DL | CC |
| EVERT | DL | 50, 50 | 80, 20 |
| | CC | 90, 10 | 20, 80 |

**FIGURE 8.4** Tennis Point

It should be clear that Navratilova will not believe any message she receives from Evert. For example, if Evert's message is "I am going to play DL," and Navratilova believes her, then Navratilova should choose to cover DL. But if Evert thinks that Navratilova will cover DL, then Evert's best choice is CC. At the next level of thinking, Navratilova should see through this and not believe the assertion of DL.

But there is more. Navratilova should not believe that Evert would do exactly the opposite of what she says either. Suppose Evert's message is "I am going to play DL," and Navratilova thinks "She is just trying to trick me, and so I will take it that she will play CC." This will lead Navratilova to choose to cover CC. But if Evert thinks that Navratilova will disbelieve her in this simple way, then Evert should choose DL after all. And Navratilova should see through this, too.

Thus, Navratilova's disbelief should mean that she should just totally disregard Evert's message. Then the full two-stage game has only the babbling equilibrium. The two players' actions in the second stage will be simply those of the original equilibrium, and Evert's first-stage message can be anything. This is true of all zero-sum games.

## C. Partially Aligned Interests

But what about more general games in which there is a mixture of conflict and common interest? Whether direct communication is credible in such games depends on how the two aspects of conflict and cooperation mix when players' interests are only partially aligned. Thus, we should expect to see both cheap talk and babbling equilibria in games of this type. More generally, the greater the alignment of interests, the more information should be communicable. We illustrate this intuition with an example.

Consider a situation that you may have already experienced or, if not, soon will when you start to earn and invest. When your financial adviser recommends an investment, he may be doing so as part of developing a long-run relationship with you for the steady commissions that your business will bring him or he may be a fly-by-night operator who touts a loser, collects the up-front fee, and disappears. The credibility of his recommendation depends on what type of relationship you establish with him.

Suppose you want to invest \$100,000 in the asset recommended by your adviser and that you anticipate three possible outcomes. The asset could be a bad investment (B), leading to a 50% loss, or a payoff of $-50$ measured in thousands of dollars. The asset could be a mediocre investment (M), yielding a 1% return, or a payoff of 1. Finally, it could be a good investment (G), yielding a 55% return, or a payoff of 55. If you choose to invest, you pay the adviser a 2% fee up front regardless of the performance of the asset; this fee gives your adviser a payoff of 2 and simultaneously lowers your payoff by 2. Your adviser will also earn 20% of any gain you make, leaving you with a payoff of 80% of the gain, but he will not have to share in any loss.

With no specialized knowledge related to the particular asset that has been recommended to you, you cannot judge which of the three outcomes might be more likely. Therefore, you simply assume that all three possibilities—B, M, and G—are equally likely: there is a one-third chance of each outcome occurring. In this situation, in the absence of any further information, you calculate your expected payoff from investing in the recommended asset as $[(1/3 \times -50) + (1/3 \times 0.8 \times 1) + (1/3 \times 0.8 \times 55)] - 2 = [1/3 \times (-50 + 0.8 + 44)] - 2 = [1/3 \times (-5.2)] - 2 = -1.73 - 2 = -3.73$. This calculation indicates an expected loss of \$3,730. Therefore, you would not make the investment, and your adviser would not get any fee. Similar calculations show that you would also choose not to invest, due to a negative expected payoff, if you believed the asset was definitely the B type, definitely the M type, or definitely any probability-weighted combination of the B and M types alone.

Your adviser is in a different situation. He has researched the investment and knows which of the three possibilities—B, M, or G—is the truth. We want to determine what he will do with his information, specifically whether he will truthfully reveal to you what he knows about the asset. We consider the various possibilities below, assuming that you update your belief about the asset's type based on the information you receive from your adviser. For this example, we assume that you simply believe what you are told: you assign probability 1 to the asset being the type stated by your adviser.[5]

**I. SHORT-TERM RELATIONSHIP** If your adviser tells you that the recommended asset is type B, you will choose not to invest. Why? Because your expected payoff from that asset is $-50$ and investing would cost you an additional 2 (in fees to the adviser) for a final payoff of $-52$. Similarly, if he tells you the asset is M, you will also not invest. In that case, your expected payoff is 80% of the return of 1 minus

[5] In the language of probability theory, the probability you assign to a particular event after having observed, or heard, information or evidence about that event is known as the *posterior probability* of the event. You thus assign posterior probability 1 to the stated quality of the asset. *Bayes' theorem*, which we explain in detail in the appendix to this chapter, provides a formal quantification of the relationship between prior and posterior probabilities.

the 2 in fees for a total of −1.2. Only if the adviser tells you that the asset is G will you choose to invest. In this situation, your expected payoff is 80% of the 55 return less the 2 in fees, or 42.

What will your adviser do with his knowledge then? If the truth is G, your adviser will want to tell you the truth in order to induce you to invest. But if he anticipates no long-term relationship with you, he will be tempted to tell you that the truth is G, even when he knows the asset is either M or B. If you decide to invest based on his statement, he simply pockets his 2% fee and flees; he has no further need to stay in touch. Knowing that there is a possibility of getting bad advice, or false information, from an adviser with whom you will interact only once, you should ignore the adviser's recommendation altogether. Therefore, in this asymmetric information, short-term relationship game, credible communication is not possible. The only equilibrium is the babbling one in which you ignore your adviser; there is no cheap talk equilibrium in this case.

**II. LONG-TERM RELATIONSHIP: FULL REVELATION** Now suppose your adviser works for a firm that you have invested with for years: losing your future business may cost him his job. If you invest in the asset he recommends, you can compare its actual performance to your adviser's forecast. That forecast could prove to have been wrong in a small way (the forecast was M and the truth is B, or the forecast was G and the truth is M) or in a large way (the forecast was G and the truth is B). If you discover such misrepresentations, your adviser and his firm lose your future business. They may also lose business from others if you bad-mouth them to friends and acquaintances. If the adviser attaches a cost to his loss of reputation, he is implicitly concerned about your possible losses, and therefore his interests are *partially aligned* with yours. Suppose the payoff cost to his reputation of a small misrepresentation is 2 (the monetary equivalent of a $2,000 loss) and that of a large misrepresentation is 4 (a $4,000 loss). We can now determine whether the partial alignment of your interests with those of your adviser is sufficient to induce him to be truthful.

As we discussed earlier, your adviser will tell you the truth if the asset is G to induce you to invest. We need to consider his incentives when the truth is *not* G, when the asset is actually B or M. Suppose first that the asset is B. If your adviser truthfully reveals the asset's type, you will not invest, he will not collect any fee, but he will also suffer no reputational cost: his payoff from reporting B when the truth is B is 0. If he tells you the asset is M (even though it is B), you still will not buy because your expected payoff is −1.2 as we calculated earlier. Then the adviser will still get 0, so he has no incentive to lie and tell you that a B-type asset is really M.[6] But what if he reports G? If you believe him and invest, he will get the

[6] We are assuming that if you do not invest in the recommended asset, you do not find out its actual return, so the adviser can suffer no reputation cost in that case. This assumption fits nicely with the general interpretation of "cheap talk." Any message has no direct payoff consequences to the sender; those arise only if the receiver acts upon the information received in the message.

up-front fee of 2, but he will also suffer the reputational cost of the large error, 4.[7] His payoff from reporting G (when the truth is B) is negative: your adviser would do better to reveal B truthfully. Thus, in situations when the truth about the asset is G or B, the adviser's incentives are to reveal the type truthfully.

But what if the truth is M? Truthful revelation does not induce you to invest: the adviser's payoff is 0 from reporting M. If he reports G and you believe him, you invest. The adviser gets his fee of 2, 20% of the 1 that is your return under M, and he also suffers the reputation cost of the small misrepresentation, 2. His payoff is $2 + (0.2 \times 1) - 2 = 0.2 > 0$. Thus, your adviser does stand to benefit by falsely reporting G when the truth is M. Knowing this, you would not believe any report of G.

Because your adviser has an incentive to lie when the asset he is recommending is M, full information cannot be credibly revealed in this situation. The babbling equilibrium, where any report from the adviser is ignored, is still a possible equilibrium. But is it the *only* equilibrium here or is some partial communication possible? The failure to achieve full revelation occurs because the adviser will misreport M as G, so suppose we lump those two possibilities together into one event and label it "not-B." Thus, the adviser asks himself what he should report: "B or not-B?"[8] Now we can consider whether your adviser will choose to report truthfully in this case of partial communication.

**III. LONG-TERM RELATIONSHIP: PARTIAL REVELATION** To determine your adviser's incentives in the "B or not-B" situation, we need to figure out what inference you will draw (that is, what posterior probability you will calculate) from the report "not-B," assuming you believe it. Your prior (original) belief was that B, M, and G were equally likely, with probabilities 1/3 each. If you are told "not-B," you are left with the two possibilities of M and G. You regarded the two as equally likely originally, and there is no reason to change that assumption, so you now give each a probability of 1/2. These are your new, posterior, probabilities, conditioned on the information you receive from the adviser's report. With these probabilities, your expected payoff if you invest when the report is "not-B" is: $[1/2 \times (0.8 \times 1)] + [1/2 \times (0.8 \times 55)] - 2 = 0.4 + 22 - 2 = 20.4 > 0$. This positive expected payoff is sufficient to induce you to invest when given a report of "not-B."

Knowing that you will invest if you are told "not-B," we can determine whether your adviser will have any incentive to lie. Will he want to tell you "not-B" even if the truth is B? When the asset is actually type B and the adviser tells the truth (reports B), his payoff is 0 as we calculated earlier. If he reports "not-B" instead, and you believe him, he gets 2 in fees.[9] He also incurs the

[7] The adviser's payoff calculation does not include a 20% share of your return here. The adviser knows the truth to be B and so knows you will make a loss, in which he will not share.

[8] Our apologies to William Shakespeare.

[9] Again, the adviser's calculation includes no portion of your gain because you will make a loss: the truth is B and the adviser knows the truth.

reputation cost associated with misrepresentation. Because you assume that M or G is equally likely in the "not-B" report, the expected value of the reputation cost in this case will be 1/2 times the cost of 2 for small misrepresentation plus 1/2 times the cost of 4 for large misrepresentation: the expected reputation cost is then $(1/2 \times 2) + (1/2 \times 4) = 3$. Your adviser's net payoff from saying "not-B" when the truth is B is $2 - 3 = -1$. Therefore, he does not gain by making a false report to you. Because telling the truth is your adviser's best strategy here, a cheap talk equilibrium with credible *partial revelation* of information is possible.

The concept of the partial-revelation cheap talk equilibrium can be made more precise using the concept of a partition. Recall that you anticipate three possible cases or events—B, M, and G. This set of events can be divided, or partitioned, into distinct subsets, and your adviser then reports to you the subset containing the truth. (Of course, the verity of his report remains to be examined as part of the analysis.) Here we have a situation with a partition into two subsets, one consisting of the singleton B, and the other consisting of the pair of events {M, G}. In the partial-revelation equilibrium, these two subsets can be distinguished based on the adviser's report, but the finer distinction between M and G, leading to the finest possible partition into three subsets each consisting only of a singleton, cannot be made. That finer distinction would be possible only in a case in which a full-revelation equilibrium exists.

We advisedly said earlier that a cheap talk equilibrium with credible partial revelation of information is *possible*. This game is one with multiple equilibria because the babbling equilibrium also remains possible. The configuration of strategies and beliefs where you ignore the adviser's report, and the adviser sends the same report (or even a random report) regardless of the truth, is still an equilibrium. Given each player's strategies, the other has no reason to change his actions or beliefs. In the terminology of partitions, we can think of this babbling equilibrium as having the coarsest possible, and trivial, partition with just one (sub)set {B, M, G} containing all three possibilities. In general, whenever you find a non-babbling equilibrium in a cheap talk game, there will also be at least one other equilibrium with a coarser or cruder partition of outcomes.

**IV. MULTIPLE EQUILIBRIA** As an example of a situation in which coarser partitions are associated with additional equilibria, consider the case in which your adviser's cost of reputation is higher than assumed above. Let the reputation cost be 4 (instead of 2) for a small misrepresentation of the truth and 8 (instead of 4) for a large misrepresentation. Our analysis above showed that your adviser will report G if the truth is G, and that he will report B if the truth is B. These results continue to hold. Your adviser wants you to invest when the truth is G, and he still gets the same payoff from reporting B when the truth is B as he does from reporting M in that situation. The higher reputation cost gives him even less incentive to falsely report G when the truth is B. So if the asset is either B or G, the adviser can be expected to report truthfully.

The problem for full revelation in our earlier example arose because of the adviser's incentive to lie when the asset is M. With our earlier numbers, his payoff from reporting G when the truth is M was higher than that from reporting truthfully. Will that still be true with the higher reputation costs?

Suppose the truth is M and the adviser reports G. If you believe him and invest in the asset, his expected payoff is 2 (his fee) $+ 0.2 \times 1$ (his share in the actual return from an M-type asset) $-$ 4 (his reputation cost) $= -1.8 < 0$. The truth would get him 0. He no longer has the temptation to exaggerate the quality of the stock. The outcome where he always reports the truth, and you believe him and act upon his report, is now a cheap talk equilibrium with full revelation. This has the finest possible partition consisting of three singleton subsets, {B}, {M}, and {G}.

There are also *three* other equilibria in this case, each with a coarser partition than the full-revelation equilibrium. Both two-subset situations—one with {B, M} and {G} and the other with {B} and {M, G}—and the babbling situation with {B, M, G} are all alternative possible equilibria. We leave it to you to verify this. Which one prevails can depend on all the considerations addressed in Chapter 4 in our discussion of games with multiple equilibria.

The biggest practical difficulty associated with attaining a non-babbling equilibrium with credible information communication lies in the players' knowledge about the extent to which their interests are aligned. The extent of alignment of interest between the two players must be common knowledge between them. In the investment example, it is critical that you know from past interactions or other credible sources (for example, a contract) that the adviser has a large reputational concern in your investment outcome. If you did not know to what extent his interests were aligned with yours, you would be justified in suspecting that he was exaggerating to induce you to invest for the sake of the fee he would earn immediately.

What happens when even richer messages are possible? For example, suppose that your adviser could report a number $g$, representing his estimate of the rate of growth of the stock price, and that $g$ could range over a continuum of values. In this situation, as long as the adviser gets some extra benefit if you buy a bad stock that he recommends, he has some incentive to exaggerate $g$. Therefore, fully accurate truthful communication is no longer possible. But a partial-revelation cheap talk equilibrium may be possible. The continuous range of growth rates may split into intervals—say, from 0% to 1%, from 1% to 2%, and so on—such that the adviser finds it optimal to tell you truthfully into which of these intervals the actual growth rate falls, and you find it optimal to accept this advice and take your optimal action on its basis. The higher the adviser's valuation of his reputation, the finer the possible partition will be—for example, half-percentage points instead of whole or quarter-percentage points instead of half. However, we must leave further explanation of this idea to more advanced treatments of the subject.[10]

## D. Formal Analysis of Cheap Talk Games

Our analysis of cheap talk games so far has been heuristic and verbal. This approach often suffices for understanding and predicting behavior, but the formal techniques for setting up and solving games—trees and matrices—are available and can be deployed if needed. To show how this is done and to connect the games in this chapter with the theory of previous chapters, we now consider the game between you and your financial adviser in this framework. For this analysis, we assume that the "language" of communication from your adviser distinguishes all three possibilities B, M, and G—that is, we consider the finest possible partition of information. After reading this section, you should be able to complete a similar analysis for the case where the adviser's report has to be the coarser choice between B or "not-B."

We start by constructing the tree for this game, illustrated in Figure 8.5. The fictitious player Nature, introduced in Chapter 3, makes the first move, producing one of three scenarios for the return on your investment, namely B, M, or G, with equal probabilities of 1/3 each. Your adviser observes Nature's move and chooses his action, namely the report to you, which can again be B, M, or G. We simplify the tree a little right away by noting that the adviser never has any incentive to understate the return on the investment; he will never report B when the truth is M or G, nor will he report M when the truth is G. (You could leave those possible actions in the tree, but they make it unnecessarily complex. Our application of one step of rollback shows that none of them is ever optimal for the adviser, so none could ever be part of an equilibrium.)

Finally, you are the third mover and you must choose whether to invest (I) or not invest (N). You do not observe Nature's move directly, however—you know only the adviser's report. Therefore, for you, both nodes where the adviser reports M are gathered in one information set while all three nodes where the adviser reports G are gathered in another information set: both information sets are indicated by dotted ovals around the relevant nodes in Figure 8.5. The presence of the information sets indicates that your actions are constrained. In the information set where the adviser has reported M, you must make the same investment choice at both nodes in the set. You must choose either I at both nodes or N at both nodes: you cannot distinguish between the two nodes inside the

[10] The seminal paper by Vincent Crawford and Joel Sobel, "Strategic Information Transmission," *Econometrica*, vol. 50, no. 6 (November 1982), pp. 1431–52, developed this theory of partial communication. An elementary exposition and survey of further work is in Joseph Farrell and Matthew Rabin, "Cheap Talk," *Journal of Economic Perspectives*, vol. 10, no. 3 (Summer 1996), pp. 103–118.

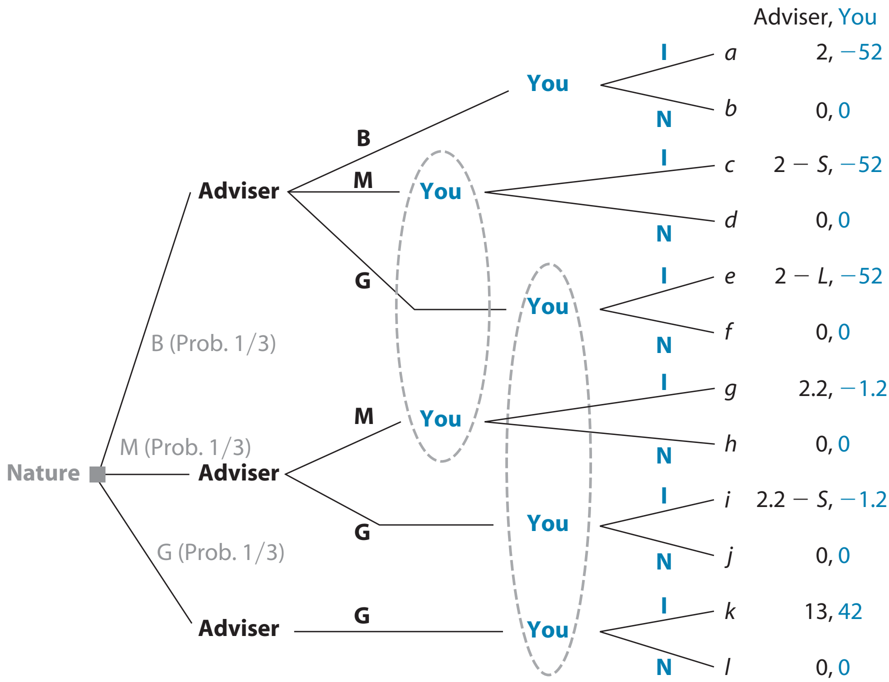


**FIGURE 8.5** Cheap Talk Game Tree: Financial Adviser and Investor

information set to choose I at one and N at the other. Likewise, you must choose either I at all three nodes or N at all three nodes of the "report G" information set.

At each terminal node, the adviser's payoff is shown first, and your payoff is shown second. The payoff numbers, measured in thousands of dollars, reflect the same numeric values used in our heuristic analysis earlier. You pay your adviser a 2% fee on your \$100,000 investment, and your return is −50 if you invest in B, 1 if you invest in M, and 55 if you invest in G. Your adviser retains 20% of any gain you earn from his recommendation. We make one change to our former model by not specifying the exact value of the adviser's reputation cost of misrepresentation. Instead, we use $S$ to denote the reputation cost of a small misrepresentation and $L$ for that of a large misrepresentation; to be consistent with our analysis above, we assume that both are positive and that $S < L$. This approach allows us to consider both levels of reputational consideration discussed earlier.

As a sample of how each pair of payoffs is calculated, consider the node at which Nature has produced an asset of type M, the adviser has reported G, and you have chosen I; this node is labeled $i$ in Figure 8.5. With these choices, your payoff includes the up-front fee of 2 paid to your adviser along with 80% of the investment's return of 1 for a total of $0.8 - 2 = -1.2$. The adviser gets his fee of

2 and his 20% share of the asset's return (0.2) but suffers the reputation cost of $S$, so his total payoff is $2.2 - S$. We leave it to you to confirm that all of the other payoffs have been computed correctly.

With the help of the tree in Figure 8.5, we can now construct a payoff matrix for this game. Technically, that matrix should include all of the strategies available to both you and your adviser. But, as in our construction of the tree, we can eliminate some possible strategies from consideration before even putting them into the table: any obviously poor strategies, for example, can be removed. This process allows us to build a much smaller table, and therefore one that is much more manageable, than would be produced if we were to include all possible strategies.

What strategies can we leave out of consideration as equilibrium strategies? The answer is twofold. First, we can ignore strategies that are obviously not going to be deployed. We already eliminated some such choices for your adviser (for example, "report B if truth is G") in building the tree. We can now see that you also have some choices that can be removed. For example, the strategy "choose I if report is B" at terminal node *a* is dominated by "choose N if report is B," so we can ignore it. Similarly, inside the "report M" information set, your action "choose I if report is M" is dominated by "choose N if report is M"; it is the worst choice at *both* terminal nodes (*c* and *g*) and can therefore also be ignored. Second, we can remove strategies that make no difference to the search for cheap talk equilibria. For the adviser, for example, "report B" and "report M" both lead to your choosing N, so we remove them as well. In addition to the terminal nodes we have already eliminated in Figure 8.5 (*a*, *c*, and *g*), we can now eliminate *b*, *d*, and *h* as well.

This simplification process leaves us only six terminal nodes to consider as possible equilibrium outcomes of the games (*e*, *f*, *i*, *j*, *k*, and *l*). Those nodes arise from strategies that include the adviser reporting that the asset is G and your choice in response to a report of G. Specifically, we are left with three interesting strategies for the adviser ["report G always (regardless of whether the truth is B, M, or G)," "report G only when the truth is M or G," and "report G if and only if the truth is G"] and two for you ("choose I if report is G" and "choose N even if report is G"). These five strategies yield the three-by-two payoff matrix illustrated in Figure 8.6.

The payoffs for each strategy combination in Figure 8.6 are *expected payoffs* calculated using the values shown at the terminal nodes of the tree that can be reached under that strategy combination, weighted by the appropriate probabilities. As an example, consider the top-left cell of the table, where the adviser reports G regardless of the true type of the asset, and you invest because the report is G. This strategy combination leads to terminal nodes *e*, *i*, and *k*, each with probability 1/3. Thus, the adviser's expected payoff in that cell is $\{[1/3 \times (2 - L)] + [1/3 \times (2.2 - S)] + (1/3 \times 13)\} = 1/3 \times (17.2 - L - S)$. Similarly, your

| | | YOU | |
|---|---|---|---|
| | | I if G | N if G |
| ADVISER | Always G | $(17.2 - L - S)/3$, $-11.2/3$ | 0, 0 |
| | G only if M or G | $(15.2 - S)/3$, $40.8/3$ | 0, 0 |
| | G if and only if G | $13/3$, $42/3$ | 0, 0 |

**FIGURE 8.6** Payoff Matrix for Cheap Talk Game

expected payoff in the same cell is $[(1/3 \times -52) + (1/3 \times -1.2) + (1/3 \times 42)] = 1/3 \times (-11.2)$. Again we leave it to you to confirm that the remaining expected payoffs have been computed correctly.

Now that we have a complete payoff matrix, we can use the techniques developed in Chapter 4 to identify equilibria, with the caveat that the values of $L$ and $S$ will play a role in our analysis. Simple best-response analysis shows that your best response to "Always G" is "N if G," but your best response to the adviser's other two strategies is "I if G." Similarly, the adviser's best response to your "N if G" can be any of his three choices. Thus, we have our first result: the top-right cell is always a Nash equilibrium. If the adviser reports G regardless of the truth (or for that matter sends any report that is the same in all three scenarios), then you do better by choosing N, and given that you are choosing N, the adviser has no incentive to deviate from his choice. This equilibrium is the babbling equilibrium with no information communication that we saw earlier.

Next consider the adviser's best response to your choice of "I if G." The only possible equilibria occur when he chooses "G only if M or G" or "G if and only if G." But whether he will pick one or the other of these, or indeed neither, depends on the specific values of $L$ and $S$. For the strategy pair {"G only if M or G," "I if G"} to be a Nash equilibrium, it must be true that $15.2 - S > 17.2 - L - S$ and that $15.2 - S > 13$. The first expression holds if $L > 2$; the second if $S < 2.2$. So if the values of $L$ and $S$ meet these requirements, the middle-left cell will be a cheap talk (Nash) equilibrium. In this equilibrium, the report G does not allow you to infer whether the true scenario is M or G, but you know that the truth is definitely not B. Knowing this much, you can be sure that your expected payoff will be positive, and you choose to invest. In this situation, G really means "not-B," and the equilibrium outcome is formally equivalent to the partial-revelation equilibrium we discussed earlier.[11]

[11] Incidentally, this highlights a certain arbitrariness in language. It does not matter whether the report is G or "not-B," as long as its significance is clearly understood by the parties. One can even have upside-down conventions where "bad" means "good" and vice versa, if the translation from the terms to meaning is common knowledge to all parties involved in the communication.

We can also check for the conditions under which the strategy pair {"G if and only if G," "I if G"} is a Nash equilibrium. That outcome requires both $13 > 17.2 - L - S$ and $13 > 15.2 - S$. These are less easily handled than the pair of expressions above. Note however that the latter expression requires $S > 2.2$ and that we have assumed that $L > S$; so $L > 2.2$ must hold when $S > 2.2$ holds. You can now use these requirements to check whether the first expression will hold. Use the minimum value of $L$ and $S$, 2.2, and plug these into $13 > 17.2 - L - S$ to find $13 > 12.8$, which is always true. These calculations indicate that the bottom-left cell is a cheap talk equilibrium when $S > 2.2$, as long as $L > S$. This equilibrium is the one with full revelation that we identified at the end of our earlier analysis.

In each case described here, the babbling equilibrium exists along with either the {"G if and only if G," "I if G"} or the {"G only if M or G," "I if G"} equilibrium. Note that we get *only* the babbling equilibrium when the reputational cost to your adviser is small ($L < 2$, and $S < L$), which is consistent with the intuition we presented earlier. Finally, if we restrict the language of messages to the coarser partition between B and "not-B," then an extension of the analysis here shows that the strategy set {" 'not-B' if M or G," "I if 'not-B' "} is also a Nash equilibrium of that game.

In each instance, our formal analysis confirms the verbal arguments we made in Section 3.C. Some of you may find the verbal approach sufficient for most, if not all, of your needs. Others may prefer the more formal model presented in this section. Be aware, however, that game trees and matrices can only go so far: once your model becomes sufficiently complex, with a continuum of report choices, for example, you will need to rely almost entirely on mathematics to identify equilibria. Being able to solve models of asymmetric information in a variety of forms—verbally, with trees and tables, or with algebra or calculus—is an important skill. Later in this chapter, we present additional examples of such games: we solve one using a combination of intuition and algebra and the other with a game tree and payoff table. In each case, the one solution method does not preclude the other, so you may attempt alternative solutions on your own.

# 4 ADVERSE SELECTION, SIGNALING, AND SCREENING

## A. Adverse Selection and Market Failure

In many games, one of the players knows something pertinent to the outcomes that the other players don't know. An employer knows much less about the skills of a potential employee than does the employee himself; vaguer but important matters such as work attitude and collegiality are even harder to observe. An insurance company knows much less about the health or driving skills of someone applying for medical or auto insurance than does the applicant. The seller of a used car knows a lot about the car from long experience; a potential buyer can at best get a little information by inspection.

In such situations, direct communication will not credibly signal information. Unskilled workers will claim to have skills to get higher-paid jobs; people who are bad risks will claim good health or driving habits to get lower insurance premiums; owners of bad cars will assert that their cars run fine and have given them no trouble in all the years they have owned them. The other parties to the transactions will be aware of the incentives to lie and will not trust the information conveyed by the words. There is no possibility of a cheap talk equilibrium of the type described in Section 3.

What if the less-informed parties in these transactions have no way of obtaining the pertinent information at all? In other words, to use the terminology introduced in Section 2 above, suppose that no credible screening devices nor signals are available. If an insurance company offers a policy that costs 5 cents for each dollar of coverage, then the policy will be especially attractive to people who know that their own risk (of illness or a car crash) exceeds 5%. Of course, some people who know their risk to be lower than 5% will still buy the insurance because they are risk averse. But the pool of applicants for this insurance policy will have a larger proportion of the poorer risks than the proportion of these risks in the population as a whole. The insurance company will selectively attract an unfavorable, or adverse, group of customers. This phenomenon is very common in transactions involving asymmetric information and is known as **adverse selection**. (This term in fact originated within the insurance industry.)

Potential consequences of adverse selection for market transactions were dramatically illustrated by George Akerlof in a paper that became the starting point of economic analysis of asymmetric information situations and won him a Nobel Prize in 2001.[12] We use his example to introduce you to the effects that adverse selection may have.

## B. The Market for "Lemons"

Think of the market in 2014 for a specific kind of used car, say a 2011 Citrus. Suppose that in use these cars have proved to be either largely trouble free and reliable or have had many things go wrong. The usual slang name for the latter type is "lemon," so for contrast let us call the former type "orange."

[12] George Akerlof, "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, vol. 84, no. 3 (August 1970), pp. 488–500.

Suppose that each owner of an orange Citrus values it at \$12,500; he is willing to part with it for a price higher than this but not for a lower price. Similarly, each owner of a lemon Citrus values it at \$3,000. Suppose that potential buyers are willing to pay more than these values for each type. If a buyer could be confident that the car he was buying was an orange, he would be willing to pay \$16,000 for it; if the car was a known lemon, he would be willing to pay \$6,000. Since the buyers value each type of car more than do the original owners, it benefits everyone if all the cars are traded. The price for an orange can be anywhere between \$12,500 and \$16,000; that for a lemon anywhere between \$3,000 and \$6,000. For definiteness, we will suppose that there is a limited stock of such cars and a larger number of potential buyers. Then the buyers, competing with each other, will drive the price up to their full willingness to pay. The prices will be \$16,000 for an orange and \$6,000 for a lemon—if each type could be identified with certainty.

But information about the quality of any specific car is not symmetric between the two parties to the transaction. The owner of a Citrus knows perfectly well whether it is an orange or a lemon. Potential buyers don't, and the owner of a lemon has no incentive to disclose the truth. For now, we confine our analysis to the private used-car market in which laws requiring truthful disclosure are either nonexistent or hard to enforce. We also assume away any possibility that the potential buyer can observe something that tells him whether the car is an orange or a lemon; similarly, the car owner has no way to indicate the type of car he owns. Thus, for this example, we consider the effects of the information asymmetry alone without allowing either side of the transaction to signal or screen.

When buyers cannot distinguish between oranges and lemons, there cannot be distinct prices for the two types in the market. There can be just one price, $p$, for a Citrus; the two types—oranges and lemons—must be pooled. Whether efficient trade is possible under such circumstances will depend on the proportions of oranges and lemons in the population. We suppose that oranges are a fraction $f$ of used Citruses and lemons the remaining fraction $(1 - f)$.

Even though buyers cannot verify the quality of an individual car, they can know the proportion of good cars in the population as a whole, for example, from newspaper reports, and we assume this to be the case. If all cars are being traded, a potential buyer will expect to get a random selection, with probabilities $f$ and $(1 - f)$ of getting an orange and a lemon, respectively. The expected value of the car purchased is $16{,}000 \times f + 6{,}000 \times (1 - f) = 6{,}000 + 10{,}000 \times f$. He will buy such a car if its expected value exceeds the price he is asked to pay, that is, if $6{,}000 + 10{,}000 \times f > p$.

Now consider the point of view of the seller. The owners know whether their cars are oranges or lemons. The owner of a lemon is willing to sell it as long as the price exceeds its value to him, that is, if $p > 3{,}000$. But the owner of an orange requires $p > 12{,}500$. If this condition for an orange owner to sell is satisfied, so is the sell condition for a lemon owner.

To meet the requirements for all buyers and sellers to want to make the trade, therefore, we need $6{,}000 + 10{,}000 \times f > p > 12{,}500$. If the fraction of oranges in the population satisfies $6{,}000 + 10{,}000 \times f > 12{,}500$, or $f > 0.65$, a price can be found that does the job; otherwise there cannot be efficient trade. If $6{,}000 + 10{,}000 \times f < 12{,}500$ (leaving out the exceptional and unlikely case where the two are just equal), owners of oranges are unwilling to sell at the maximum price the potential buyers are willing to pay. We then have adverse selection in the set of used cars put up for sale; no oranges will appear in the market at all. The potential buyers will recognize this, will expect to get a lemon for sure, and will pay at most \$6,000. The owners of lemons will be happy with this outcome, so lemons will trade. But the market for oranges will collapse completely due to the asymmetric information. The outcome will be a kind of Gresham's law, where bad cars drive out the good.

Because the lack of information makes it impossible to get a reasonable price for an orange, the owners of oranges will want a way to convince the buyers that their cars are the good type. They will want to *signal* their type. The trouble is that the owners of lemons would also like to pretend that their cars are oranges, and to this end they can imitate most of the signals that owners of oranges might attempt to use. Michael Spence, who developed the concept of signaling and shared the 2001 Nobel Prize for information economics with Akerlof and Stiglitz, summarizes the problems facing our orange owners in his pathbreaking book on signaling: "Verbal declarations are costless and therefore useless. Anyone can lie about why he is selling the car. One can offer to let the buyer have the car checked. The lemon owner can make the same offer. It's a bluff. If called, nothing is lost. Besides, such checks are costly. Reliability reports from the owner's mechanic are untrustworthy. The clever nonlemon owner might pay for the checkup but let the purchaser choose the inspector. The problem for the owner, then, is to keep the inspection cost down. Guarantees do not work. The seller may move to Cleveland, leaving no forwarding address."[13]

In reality, the situation is not so hopeless as Spence implies. People and firms that regularly sell used cars as a business can establish a reputation for honesty and profit from this reputation by charging a markup. (Of course, some used car dealers are unscrupulous.) Some buyers are knowledgeable about cars; some buy from personal acquaintances and can therefore verify the history of the car they are buying. Or dealers may offer warranties, a topic we discuss in

[13] A. Michael Spence, *Market Signaling: Information Transfer in Hiring and Related Screening Processes* (Cambridge, Mass.: Harvard University Press, 1974), pp. 93–94. The present authors apologize on behalf of Spence to any residents of Cleveland who may be offended by any unwarranted suggestion that that's where shady sellers of used cars go!

more detail later. And in other markets it is harder for bad types to mimic the actions of good types, so credible signaling will be viable. For a specific example of such a situation, consider the possibility that education can signal skill. Then it may be hard for the unskilled to acquire enough education to be mistaken for highly skilled people. The key requirement for education to separate the types is that education should be sufficiently more costly for the truly unskilled to acquire than for the truly skilled. To show how and when signaling can successfully separate types, therefore, we turn to the labor market.

## C. Signaling and Screening: Sample Situations

The basic idea of signaling or screening to convey or elicit information is very simple: players of different "types" (that is, possessing different information about their own characteristics or about the game and its payoffs more generally) should find it optimal to take different actions so that their actions truthfully reveal their types. Situations of such information asymmetry, and signaling and screening strategies to cope with them, are ubiquitous. Here are some additional situations to which the methods of analysis developed throughout this chapter can be applied.

**I. INSURANCE** The prospective buyers of an insurance policy vary in their risk categories, or their levels of riskiness to the insurer. For example, among the numerous applicants for an automobile collision insurance policy will be some drivers who are naturally cautious and others who are simply less careful. Each potential customer has a better knowledge of his or her own risk class than does the insurance company. Given the terms of any particular policy, the company will make less profit (or a greater loss) on the more risky customers. However, the more risky customers will be the ones who find the specified policy more attractive. Thus, the company attracts the less favorable group of customers, and we have a situation of adverse selection.[14] Clearly, the insurance company would like to distinguish between the risk classes. They can do so using a screening device.

Suppose as an example that there are just two risk classes. The company can then offer two policies from which any individual customer chooses one. The first has a lower premium (in units of so many cents per dollar of coverage), but covers a lower percentage of any loss incurred by the customer; the second has a higher premium, but covers a higher percentage, perhaps even 100%, of the loss. (In the case of collision insurance, this loss represents the cost of having

[14] Here we are not talking about the possibility that a well-insured driver will deliberately exercise less care. That is moral hazard, and it can be mitigated using co-insurance schemes similar to those discussed here. But for now our concern is purely adverse selection, where some drivers are just by their nature careful, and others are equally uncontrollably spaced out and careless when they drive.

an auto body shop complete the needed repairs to one's car.) A higher-risk customer is more likely to suffer the uncovered loss and therefore is more willing to pay the higher premium to get more coverage. The company can then adjust the premiums and coverage ratios so that customers of the higher-risk type choose the high-premium, high-coverage policy and customers of the less-risky type choose the lower-premium, lower-coverage policy. If there are more risk types, there have to be correspondingly more policies in the menu offered to prospective customers: with a continuous spectrum of risks, there may be a corresponding continuum of policies.

Of course, this insurance company has to compete with other insurance companies for the business of each customer. That competition affects the packages of premiums and levels of coverage it can offer. Sometimes the competition may even preclude the attainment of an equilibrium as each offering can be defeated by another.[15] But the general idea behind differential premium policies for differential risk-class customers is valid and important.

**II. WARRANTIES** Many types of durable goods—cars, computers, washing machines—vary in their quality. Any company that has produced such a good will have a pretty good idea of its quality. But a prospective buyer will be much less informed. Can a company that knows its product to be of high quality signal this fact credibly to its potential customers?

The most obvious, and most commonly used, signal is a good warranty. The cost of providing a warranty is lower for a genuinely high-quality product; the high-quality producer is less likely to be called on to provide repairs or replacement than the company with a shoddier product. Therefore, warranties can serve as signals of quality, and consumers are intuitively quite aware of this fact when they make their purchase decisions.

Typically in such situations, the signal has to be carried to excess in order to make it sufficiently costly to mimic. Thus, the producer of a high-quality car has to offer a sufficiently long or strong warranty to signal the quality credibly. This requirement is especially relevant for any company that is a relative newcomer or one that does not have a previous reputation for offering high-quality products. Hyundai, for example, began selling cars in the United States in 1986 and for its first decade had a low-quality reputation. In the mid-1990s, it invested heavily in better technology, design, and manufacturing. To revamp its image, it offered the then-revolutionary 10-year, 100,000-mile warranty. Now it ranks with consumer groups as one of the better-quality automobile manufacturers.

[15] See Michael Rothschild and Joseph Stiglitz, "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, vol. 90, no. 4 (November 1976), pp. 629–49.

**III. PRICE DISCRIMINATION** The buyers of most products are heterogeneous in terms of their willingness to pay, their willingness to devote time to searching for a better price, and so on. Companies would like to identify those potential customers with a higher willingness to pay and charge them one, presumably fairly high, price while offering selective good deals to those who are not willing to pay so much (as long as that willingness to pay still exceeds the cost of supplying the product). The companies can successfully charge different prices to different groups of customers by using screening devices to separate the types. We will discuss such strategies, known as price discrimination in the economics literature, in more detail in Chapter 13. Here we provide just a brief overview.

The example of discriminatory prices best known to most people comes from the airline industry. Business travelers are willing to pay more for their airline tickets than are tourists, often because at least some of the cost of the ticket is borne by the business traveler's employer. It would be illegal for airlines blatantly to identify each traveler's type and then to charge them different prices. But the airlines take advantage of the fact that tourists are also more willing to commit to an itinerary well in advance, while business travelers need to retain flexibility in their plans. Therefore, airlines charge different prices for nonrefundable versus refundable fares and leave it to the travelers to choose the fare type. This pricing strategy is an example of *screening by self-selection*.[16] Other devices—advance purchase or Saturday night stay requirements, different classes of onboard service (first versus business versus coach)—serve the same screening purpose.

Price discrimination is not specific to high-priced products like airline tickets. Other discriminatory pricing schemes can be observed in many markets where product prices are considerably lower than those for air travel. Coffee and sandwich shops, for example, commonly offer "frequent buyer" discount cards. These cards effectively lower the price of coffee or a sandwich to the shop's regular customers. The idea is that regular customers are more willing to search for the best deal in the neighborhood, while visitors or occasional users would go to the first coffee or sandwich shop they see without spending the time necessary to determine whether any lower prices might be available. The higher regular price and "free 11th item" discount represent the menu of options from which the two types of customer select, thereby separating them by type.

Books are another example. They are typically first published in a higher-price hardcover version; a cheaper paperback comes out several months to a year or more later. The difference in the costs of producing the two versions is negligible. But the versions serve to separate the buyers who want to read the

[16] We investigate the idea of self-selection more formally in Section 5 of this chapter.

book immediately and are willing to pay more for the ability to do so from those who wish to pay less and are willing to wait longer.

**IV. PRODUCT DESIGN AND ADVERTISING** Can an attractive, well-designed product exterior serve the purpose of signaling high quality? The key requirement is that the cost of the signal be sufficiently higher for a company trying to pretend high quality than for one that has a truly high-quality product. Typically, the cost of the product exterior is the same regardless of the innate quality that resides within. Therefore, the mimic would face no cost differential, and the signal would not be credible.

But such signals may have some partial validity. Exterior design is a fixed cost that is spread over the whole product run. Buyers do learn about quality from their own experience, from friends, and from reviews and comments in the media. These considerations indicate that a high-quality good can expect to have a longer market life and higher total sales. Therefore, the cost of an expensive exterior is spread over a larger volume and adds less to the cost of each unit of the product, if that product is of higher innate quality. The firm is in effect making a statement: “We have a good product that will sell a lot. That is why we can afford to spend so much on its design. A fly-by-night firm would find this prohibitive for the few units it expects to sell before people find out its poor quality and don’t buy any more from it.” Even expensive, seemingly useless and uninformative product launch and advertising campaigns can have a similar signaling effect.[17]

Similarly, when you walk into a bank and see solid, expensive marble counters and plush furnishings, you may be reassured about its stability. However, for this particular signal to work, it is important that the building, furnishings, and décor be specific to the bank. If everything could easily be sold to other types of establishments and the space converted into a restaurant, say, then a fly-by-night operator could mimic a truly solid bank at no higher cost. In that situation, the signal would not be credible.

**V. TAXIS** The examples above are drawn primarily from economics, but here is one from the field of sociology about taxi service. The overwhelming majority of people who hail a taxi simply want to go to their destination, pay the fare, and depart. But a few are out to rob the driver or hijack the cab, perhaps with some physical violence involved. How can taxi drivers screen their prospective customers and accept only the good ones? Sociologists Diego Gambetta and Heather Hamill researched this question using extensive interviews with taxi

[17] Kyle Bagwell and Gary Ramey, “Coordination Economies, Advertising, and Search Behavior in Retail Markets,” *American Economic Review*, vol. 84, no. 3 (June 1994), pp. 498–517.

drivers in New York (where robbery is the main problem) and Northern Ireland (where sectarian violence was a serious problem at the time of their study).[18]

The drivers need an appropriate screening device, knowing that the bad types of potential customers are trying to mimic the actions of the good type. The usual differential cost condition applies. A New York customer wearing a suit is not guaranteed to be harmless, because a robber can buy and wear a suit for the same cost as a good customer; race and gender cannot be used to screen customers either. In Northern Ireland, the warring factions were also not easily distinguishable by external characteristics.

Gambetta and Hamill found that some screens were more useful to the taxi drivers than others. For example, ordering a cab by phone was a better signal of a customer's trustworthiness than hailing on the street: when you revealed a pickup location, the taxi company literally "knew where you lived."[19] More important, some signaling devices worked better for customers (and were therefore better screens for the drivers) when used in combination rather than individually. Wearing a suit was no good as a credible screen all by itself, but a customer coming out of an office building wearing a suit was deemed safer than a random suit-wearing customer standing on a street corner. Most office buildings have some security in the lobby these days, and such a customer could be deemed to have already passed one level of security testing.

Perhaps most important were the involuntary signals that people give off—microexpressions, gestures, and so forth—that experienced drivers learn to read and interpret. Exactly because these are involuntary, they act as signals with an infinite cost of mimicry and are therefore the most effective in screening to separate types.[20]

**VI. POLITICAL BUSINESS CYCLES** And now we provide two examples from the field of political economy. Incumbent governments often increase spending to get the economy to boom just before an election, thereby hoping to attract more votes and win the election. But shouldn't rational voters see through this stratagem and recognize that, as soon as the election is over, the government will be forced to retrench, perhaps leading to a recession? For pre-election spending to be an effective signal of type, there has to be some uncertainty in the voters' mind about the "competence-type" of the government. The future recession will create a political cost for the government. This cost will be smaller if the government is more competent in its handling of the economy. If the cost differential

[18] Diego Gambetta and Heather Hamill, *Streetwise: How Taxi Drivers Establish Their Customers' Trustworthiness* (New York: Russell Sage Foundation, 2005).

[19] Even if the location was a restaurant or office, not a home, you leave more evidence about yourself when you call for pickup than when you hail a cab on the street.

[20] Paul Ekman, *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage* (New York: W. W. Norton & Company, 2009), reports on how such inadvertent signals can be read and interpreted.

between competent and incompetent government types is large enough, a sufficiently high expenditure spike can credibly signal competence.[21]

Another similar example relates to inflation controls. Many countries at many times have suffered high inflation, and governments have piously declared their intentions to reduce this level. Can a government that truly cares about price stability credibly signal its type? Yes. Governments can issue bonds protected against inflation: the interest rate on such bonds is automatically ratcheted up by the rate of inflation or the capital value of the bond rises in proportion to the increase in the price level. Issuing government debt in this form is more costly to a government that likes policies that lead to higher inflation, because it has to make good on the contract of paying more interest or increasing the value of its debt. Therefore, a government with genuinely anti-inflation preferences can issue inflation-protected bonds as a credible signal, separating itself from the inflation-loving type of government.

**VII. EVOLUTIONARY BIOLOGY** Finally, an example from the natural sciences. In many species of birds, the males have very elaborate and heavy plumage that females find attractive. One should expect the females to seek genetically superior males so that their offspring will be better equipped to survive to adulthood and to attract mates in their turn. But why does elaborate plumage indicate such desirable genetic qualities? One would think that such plumage might be a handicap, making the male bird more visible to predators (including human hunters) and less mobile, therefore less able to evade these predators. Why do females choose these seemingly handicapped males? The answer comes from the conditions for credible signaling. Although heavy plumage is indeed a handicap, it is less of a handicap to a male who is sufficiently genetically superior in qualities such as strength and speed. The weaker the male, the harder it will be for him to produce and maintain plumage of a given quality. Thus, it is precisely the heaviness of the plumage that makes it a credible signal of the male's quality.[22]

## D. Experimental Evidence

The characterization of and solution for equilibria in games of signaling and screening entail some quite subtle concepts and computations. Thus, in each case above, formal models must be carefully described in order to formulate reasonable and accurate predictions for player choices. In all such games, players must revise or update their probabilities about other players' type(s) based

[21] These ideas and the supporting evidence are reviewed by Alan Drazen in "The Political Business Cycle after 25 Years," in *NBER Macroeconomics Annual 2000*, ed. Ben S. Bernanke and Kenneth S. Rogoff (Cambridge, Mass.: MIT Press, 2001), pp. 75–117.

[22] Matt Ridley, *The Red Queen: Sex and the Evolution of Human Behavior* (New York: Penguin, 1995), p. 148.

on observation of those other players' actions. This updating requires an application of Bayes' theorem, which is explained in the appendix to this chapter. We also carefully analyze an example of a game with this kind of updating in Section 6 below.

You can imagine, without going into any of the details of the appendixes, that these probability-updating calculations are quite complex. Should we expect players to perform them correctly? There is ample evidence that people are very bad at performing calculations that include probabilities and are especially bad at conditioning probabilities on new information.[23] Therefore, we should be justifiably suspicious of equilibria that depend on the players' doing so.

Relative to this expectation, the findings of economists who have conducted laboratory experiments of signaling games are encouraging. Some surprisingly subtle refinements of *Bayesian-Nash* and *perfect Bayesian equilibria* are successfully observed, even though these refinements require not only updating of information by observing actions along the equilibrium path but also deciding how one would infer information from off-equilibrium actions that should never have been taken in the first place. However, the verdict of the experiments is not unanimous: much seems to depend on the precise details of the laboratory design of the experiment.[24]

## 5 SIGNALING IN THE LABOR MARKET

Many of you expect that when you graduate, you will work for an elite firm in finance or computing. These firms have two kinds of jobs. One kind requires high quantitative and analytical skills and capacity for hard work and offers high pay in return. The other kind of jobs are semiclerical, lower-skill, lower-pay jobs. Of course, you want the job with higher pay. You know your own qualities and skills far better than your prospective employer does. If you are highly skilled, you want your employer to know this about you, and he also wants to know. He can test and interview you, but what he can find out by these methods is limited by the available time and resources. You can tell him how skilled you are, but mere assertions about your qualifications are not credible. More objective evidence is needed, both for you to offer and for your employer to seek out.

[23] Deborah J. Bennett, *Randomness* (Cambridge, Mass.: Harvard University Press, 1998), pp. 2–3 and ch. 10. See also Paul Hoffman, *The Man Who Loved Only Numbers* (New York: Hyperion, 1998), pp. 233–40, for an entertaining account of how several probability theorists, as well as the brilliant and prolific mathematician Paul Erdös, got a very simple probability problem wrong and even failed to understand their error when it was explained to them.

[24] Douglas D. Davis and Charles A. Holt, *Experimental Economics* (Princeton: Princeton University Press, 1995), review and discuss these experiments in their chapter 7.

What items of evidence can the employer seek, and what can you offer? Recall from Section 2 of this chapter that your prospective employer will use *screening devices* to identify your qualities and skills. You will use *signals* to convey your information about those same qualities and skills. Sometimes similar or even identical devices can be used for either signaling or screening.

In this instance, if you have selected (and passed) particularly tough and quantitative courses in college, your course choices can be credible evidence of your capacity for hard work in general and of your quantitative and logical skills in particular. Let us consider the role of course choice as a screening device.

## A. Screening to Separate Types

To keep things simple, we approach this screening game using intuition and some algebra. Suppose college students are of just two types when it comes to the qualities most desired by employers: A (able) and C (challenged). Potential employers in finance or computing are willing to pay $160,000 a year to a type A and $60,000 to a type C. Other employment opportunities yield the A types a salary of $125,000 and the C types a salary of $30,000. These are just the numbers in the Citrus car example in Section 4.B above, but multiplied by a factor of 10 better to suit the reality of the job-market example. And just as in the used-car example where we supposed there was fixed supply and numerous potential buyers, we suppose here that there are many potential employers who have to compete with each other for a limited number of job candidates, so they have to pay the maximum amount that they are willing to pay. Because employers cannot directly observe any particular job applicant's type, they have to look for other credible means to distinguish among them.[25]

Suppose the types differ in their tolerance for taking a tough course rather than an easy one in college. Each type must sacrifice some party time or other activities to take a tougher course, but this sacrifice is smaller or easier to bear for the A types than it is for the C types. Suppose the A types regard the cost of each such course as equivalent to $3,000 a year of salary, while the C types regard it as $15,000 a year of salary. Can an employer use this differential to screen his applicants and tell the A types from the C types?

Consider the following hiring policy: anyone who has taken a certain number, $n$, or more of the tough courses will be regarded as an A and paid $160,000, and anyone who has taken fewer than $n$ will be regarded as a C and paid $60,000. The aim of this policy is to create natural incentives whereby only the A types will take the tough courses, and the C types will not. Neither wants to take more

[25] You may wonder whether the fact that the two types have different outside opportunities can be used to distinguish between them. For example, an employer may say, "Show me an offer of a job at $125,000, and I will accept you as type A and pay you $160,000." However, such a competing offer can be forged or obtained in cahoots with someone else, so it is not reliable.

of the tough courses than he has to, so the choice is between taking $n$ to qualify as an A or giving up and settling for being regarded as a C, in which case he may as well not take any of the tough courses and just coast through college.

To succeed, such a policy must satisfy two kinds of conditions. The first set of conditions requires that the policy gives each type of job applicant the incentive to make the choice that the firm wants him to make. In other words, the policy should be compatible with the incentives of the workers; therefore, the relevant conditions are called **incentive-compatibility conditions**. The second kind of conditions ensure that, with such an incentive-compatible choice, the workers get a better (at least, no worse) payoff from these jobs than they would get in their alternative opportunities. In other words, the workers should be willing to participate in this firm's offer; therefore, the relevant conditions are called the **participation conditions**. We will develop these conditions in the labor market context now. Similar conditions will appear in other examples later in this chapter and again in Chapter 13, where we develop the general theory of mechanism design.

**I. INCENTIVE COMPATIBILITY** The criterion that employers devise to distinguish an A from a C—namely, the number of tough courses taken—should be sufficiently strict that the C types do not bother to meet it but not so strict as to discourage even the A types from attempting it. The correct value of $n$ must be such that the true C types prefer to settle for being revealed as such and getting \$60,000, rather than incurring the extra cost of imitating the A type's behavior. That is, we need the policy to be incentive compatible for the C types, so[26]

$$60{,}000 \geq 160{,}000 - 15{,}000\, n, \quad \text{or} \quad 15\, n \geq 100, \quad \text{or} \quad n \geq 6.67.$$

Similarly, the condition that the true A types prefer to prove their type by taking $n$ tough courses is

$$160{,}000 - 3{,}000\, n \geq 60{,}000, \quad \text{or} \quad 3n \leq 100, \quad \text{or} \quad n \leq 33.33.$$

These incentive-compatibility conditions or, equivalently, **incentive-compatibility constraints**, align the job applicant's incentives with the employer's desires, or make it optimal for the applicant to reveal the truth about his skill through his action. The $n$ satisfying both constraints, because it is required to be an integer, must be at least 7 and at most 33.[27] The latter is not realistically

[26] We require merely that the payoff from choosing the option intended for one's type be at least as high as that from choosing a different option, not that it be strictly greater. However, it is possible to approach the outcome of this analysis as closely as one wants while maintaining a strict inequality, so nothing substantial hinges on this assumption.

[27] If in some other context the corresponding choice variable is not required to be an integer—for example, if it is a sum of money or an amount of time—then a whole continuous range will satisfy both incentive-compatibility constraints.

relevant in this example, as an entire college program is typically 32 courses, but in other examples it might matter.

What makes it possible to meet both conditions is the *difference* in the costs of taking tough courses between the two types: the cost is sufficiently lower for the "good" type that the employers wish to identify. When the constraints are met, the employer can use a policy to which the two types will respond differently, thereby revealing their types. This is called **separation of types** based on **self-selection**.

We did not assume here that the tough courses actually imparted any additional skills or work habits that might convert C types into A types. In our scenario, the tough courses serve only the purpose of identifying the persons who already possess these attributes. In other words, they have a pure screening function.

In reality, education does increase productivity. But it also has the additional screening or signaling function of the kind described here. In our example, we found that education might be undertaken solely for the latter function; in reality, the corresponding outcome is that education is carried further than is justified by the extra productivity alone. This extra education carries an extra cost—the cost of the information asymmetry.

**II. PARTICIPATION** When the incentive-compatibility conditions for the two types of jobs in this firm are satisfied, the A types take $n$ tough courses and get a payoff of $160{,}000 - 3{,}000n$, and the C types take no tough courses and get a payoff of 60,000. For the types to be willing to make these choices instead of taking their alternative opportunities, the participation conditions must be satisfied as well. So we need

$$160{,}000 - 3{,}000n \geq 125{,}000, \quad \text{and} \quad 60{,}000 \geq 30{,}000.$$

The C types' participation condition is trivially satisfied in this example (although that may not be the case in other examples); the A types' participation condition requires $n \leq 11.67$, or, since $n$ must be an integer, $n \leq 11$. Here, any $n$ that satisfies the A types' participation constraint of $n \leq 11$ also satisfies their incentive compatibility constraint of $n \leq 33$, so the latter becomes logically redundant, regardless of its realistic irrelevance.

The full set of conditions that are required to achieve separation of types in this labor market is then $7 \leq n \leq 11$. This restriction on possible values of $n$ combines the incentive-compatibility condition for the C types and the participation condition for the A types. The participation condition for the C types and the incentive-compatibility condition for the A types in this example are automatically satisfied when the other conditions hold.

When the requirement of taking enough tough courses is used for screening, the A types bear the cost. Assuming that only the minimum needed to achieve separation is used—namely, $n = 7$—the cost to each A type has the monetary equivalent of $7 \times \$3{,}000 = \$21{,}000$. This is the cost, in this context, of the information asymmetry. It would not exist if a person's type could be directly and objectively identified. Nor would it exist if the population consisted solely of A types. The A types have to bear this cost because there are some C types in the population from whom they (or their prospective employers) seek to distinguish themselves.[28]

## B. Pooling of Types

Rather than having the A types bear the cost of the information asymmetry, might it be better not to bother with the separation of types at all? With the separation, A types get a salary of $160,000 but suffer a cost, the monetary equivalent of $21,000, in taking the tough courses; thus, their net money-equivalent payoff is $139,000. And C types get the salary of $60,000. What happens to the two types if they are not separated?

If employers do not use screening devices, they have to treat every applicant as a random draw from the population and pay all the same salary. This is called **pooling of types**, or simply **pooling** when the sense is clear.[29] In a competitive job market, the common salary under pooling will be the population average of what the types are worth to an employer, and this average will depend on the proportions of the types in the population. For example, if 60% of the population is type A and 40% is type C, then the common salary with pooling will be

$$0.6 \times \$160{,}000 + 0.4 \times \$60{,}000 = \$120{,}000.$$

The A types will then prefer the situation with separation because it yields $139,000 instead of the $120,000 with pooling. But if the proportions are 80% A and 20% C, then the common salary with pooling will be $140,000, and the A types will be worse off under separation than they would be under pooling. The C types are always better off under pooling. The existence of the A types in the population means that the common salary with pooling will always exceed the C types' separation salary of $60,000.

However, even if both types prefer the pooling outcome, it cannot be an equilibrium when many employers or workers compete with each other in the screening or signaling process. Suppose the population proportions are 80–20 and there is an initial situation with pooling where both types are paid $140,000. An employer can announce that he will pay $144,000 for someone who takes just one tough course. Relative to the initial situation, the A types will find it

[28] In the terminology of economics, the C types in this example inflict a *negative external effect* on the A types. We will develop this concept in Chapter 11.

[29] It is the opposite of *separation of types*, described above where players differing in their characteristics get different outcomes, so the outcome reveals the type perfectly.

worthwhile because their cost of taking the course is only $3,000 and it raises their salary by $4,000, whereas C types will not find it worthwhile because their cost, $15,000, exceeds the benefit, $4,000. Because this particular employer selectively attracts the A types, each of whom is worth $160,000 to him but is paid only $144,000, he makes a profit by deviating from the pooling salary package.

But his deviation starts a process of adjustment by competing employers, and that causes the old pooling situation to collapse. As A types flock to work for him, the pool available to the other employers is of lower average quality, and eventually they cannot afford to pay $140,000 anymore. As the salary in the pool is lowered, the differential between that salary and the $144,000 offered by the deviating employer widens to the point where the C types also find it desirable to take that one tough course. But then the deviating employer must raise his requirement to two courses and must increase the salary differential to the point where two courses become too much of a burden for the C types, but the A types find it acceptable. Other employers who would like to hire some A types must use similar policies to attract them. This process continues until the job market reaches the separating equilibrium described earlier.

Even if the employers did not take the initiative to attract As rather than Cs, a type A earning $140,000 in a pooling situation might take a tough course, take his transcript to a prospective employer, and say, "I have a tough course on my transcript, and I am asking for a salary of $144,000. This should be convincing evidence that I am type A; no type C would make you such a proposition." Given the facts of the situation, the argument is valid, and the employer should find it very profitable to agree: the employee, being type A, will generate $160,000 for the employer but get only $144,000 in salary. Other A types can do the same. This starts the same kind of cascade that leads to the separating equilibrium. The only difference is in who takes the initiative. Now the type A workers choose to get the extra education as credible proof of their type; it becomes a case of signaling rather than screening.

The general point is that, even though the pooling outcome may be better for all, they are not choosing the one or the other in a cooperative, binding process. They are pursuing their own individual interests, which lead them to the separating equilibrium. This is like a prisoners' dilemma game with many players, and therefore there is something unavoidable about the cost of the information asymmetry.

## C. Many Types

We have considered an example with only two types, but the analysis generalizes immediately. Suppose there are several types: A, B, C, . . . , ranked in an order that is at the same time decreasing in their worth to the employer and increasing in the costs of extra education. Then it is possible to set up a sequence of requirements of successively higher and higher levels of education, such that the very worst type needs none, the next-worst type needs the lowest level, the type third from the bottom needs the next higher level, and so on, and the types will self-select the level that identifies them.

To finish this discussion, we provide one further point, or perhaps a word of warning, regarding signaling. You are the informed party and have available an action that would credibly signal good information (information whose credible transmission would work to your advantage). If you fail to send that signal, you will be assumed to have bad information. In this respect, signaling is like playing chicken: if you refuse to play, you have already played and lost.

You should keep this in mind when you have the choice between taking a course for a letter grade or on a pass/fail basis. The whole population in the course spans the whole spectrum of grades; suppose the average is B. A student is likely to have a good idea of his own abilities. Those reasonably confident of getting an A+ have a strong incentive to take the course for a letter grade. When they have done so, the average of the rest is less than B, say, B−, because the top end has been removed from the distribution. Now, among the rest, those expecting an A have a strong incentive to choose the letter-grade option. That in turn lowers the average of the rest. And so on. Finally, the pass/fail option is chosen by only those anticipating Cs and Ds. A strategically smart reader of a transcript (a prospective employer or the admissions officer for a professional graduate school) will be aware that the pass/fail option will be selected mainly by students in the lower portion of the grade distribution; such a reader will therefore interpret a Pass as a C or a D, not as the class-wide average B.

## 6 EQUILIBRIA IN TWO-PLAYER SIGNALING GAMES

Our analysis so far in this chapter has covered the general concept of incomplete information as well as the specific strategies of screening and signaling; we have also seen the possible outcomes of separation and pooling that can arise when these strategies are being used. We saw how adverse selection could arise in a market where many car owners and buyers came together and how signals and screening devices would operate in an environment where many employers and employees meet each other. However, we have not specified and solved a game in which just two players with differential information confront one another. Here we develop an example to show how that can be done using a game tree and payoff table as our tools of analysis. We will see that either separating or pooling can be an equilibrium and that a new type of **partially revealing** or **semiseparating equilibrium** can emerge.

## A. Basic Model and Payoff Structure

In this section, we analyze a game of market entry with asymmetric information; the players are two auto manufacturers, Tudor and Fordor. Tudor Auto Corporation currently enjoys a monopoly in the market for a particular kind of automobile, say a nonpolluting, fuel-efficient compact car. An innovator, Fordor, has a competing concept and is deciding whether to enter the market. But Fordor does not know how tough a competitor Tudor will prove to be. Specifically, Tudor's production cost, unknown to Fordor, may be high or low. If it is high, Fordor can enter and compete profitably; if it is low, Fordor's entry and development costs cannot be recouped by subsequent operating profits, and it will make a net loss if it enters.

The two firms interact in a sequential game. In the first stage of the game (period 1), Tudor sets a price (high or low, for simplicity) knowing that it is the only manufacturer in the market. In the next stage, Fordor makes its entry decision. Payoffs, or profits, are determined based on the market price of the automobile relative to each firm's production costs and, for Fordor, entry and development costs as well.

Tudor would of course prefer that Fordor not enter the market. It might therefore try to use its price in the first stage of the game as a signal of its cost. A low-cost firm would charge a lower price than would a high-cost firm. Tudor might therefore hope that if it keeps its period-1 price low, Fordor will interpret this as evidence that Tudor's cost is low and will stay out. (Once Fordor has given up and is out of the picture, in later periods Tudor can jack its price back up.) Just as a poker player might bet on a poor hand, hoping that the bluff will succeed and the opponent will fold, Tudor might try to bluff Fordor into staying out. Of course, Fordor is a strategic player and is aware of this possibility. The question is whether Tudor can bluff successfully in an equilibrium of their game. The answer depends on the probability that Tudor is genuinely low cost and on Tudor's cost of bluffing. We consider different cases below and show the resulting different equilibria.

In all the cases, the per-unit costs and prices are expressed in thousands of dollars, and the numbers of cars sold are expressed in hundreds of thousands, so the profits are measured in hundreds of millions. This will help us write the payoffs and tables in a relatively compact form that is easy to read. We calculate those payoffs using the same type of analysis that we used for the restaurant pricing game of Chapter 5, assuming that the underlying relationship between the price charged ($P$) and the quantity demanded ($Q$) is given by $P = 25 - Q$.[30]

[30] We do not supply the full calculations necessary to generate the profit-maximizing prices and the resulting firm profits in each case. You may do so on your own for extra practice, using the methods learned in Chapter 5.

To enter the market, Fordor must incur an up-front cost of 40 (this payment is in the same units as profits, or hundreds of millions, so the actual figure is $4 billion) to build its plant, launch an ad campaign, and so on. If it enters the market, its cost for producing and delivering each of its cars to the market will always be 10 (thousand dollars).

Tudor could be either a lumbering, old firm with a high unit production cost of 15 (thousand dollars) or a nimble, efficient producer with a lower unit cost. To start, we suppose that the lower cost is 5; this cost is less than what Fordor can achieve. Later in Sections 6.C and 6.D, we will investigate the effect of other cost levels. For now, suppose further that Tudor can achieve the lower unit cost with probability 0.4, or 40% of the time; therefore it has high unit cost with probability 0.6, or 60% of the time.[31]

Fordor's choices in the entry game will depend on how much information it has about Tudor's costs. We assume that Fordor knows the two possible levels of cost and therefore can calculate the profits associated with each case (as we do below). In addition, Fordor will form some belief about the probability that Tudor is the low-cost type. We are assuming that the structure of the game is common knowledge to both players. Therefore, although Fordor does not know the type of the specific Tudor it is facing, Fordor's prior belief exactly matches the probability with which Tudor has the lower unit cost; that is, Fordor's belief is that the probability of facing a low-cost Tudor is 40%.

If Tudor's cost is high, 15 (thousand), then under conditions of unthreatened monopoly it will maximize its profit by pricing its car at 20 (thousand). At that price it will sell 5 (hundred thousand) units and make a profit of 25 [$= 5 \times (20 - 15)$ hundred million, or 2.5 billion]. If Fordor enters and the two compete, then the Nash equilibrium of their duopoly game will yield operating profits of 3 to Tudor and 45 to Fordor. The operating profit exceeds Fordor's up-front cost of entry (40), so Fordor would choose to enter and earn a net profit of 5 if it knew Tudor to be high cost.

If Tudor's cost is low, 5, then in unthreatened monopoly it will price its car at 15, selling 10 and making a profit of 100. In the second-stage equilibrium following the entry of Fordor, the operating profits will be 69 for Tudor and 11 for Fordor. The 11 is less than Fordor's cost of entry of 40. Therefore, it would not enter and avoid incurring a loss of 29 if it knew Tudor to be low cost.

## B. Separating Equilibrium

If Tudor is actually high cost, but wants Fordor to think that it is low cost, Tudor must mimic the action of the low-cost type; that is, it has to price at 15. But that

[31] Tudor's probability of having low unit cost could be denoted with an algebraic parameter, $z$. The equilibrium will be the same regardless of the value of $z$, as you will be asked to show in Exercise S5 at the end of this chapter.

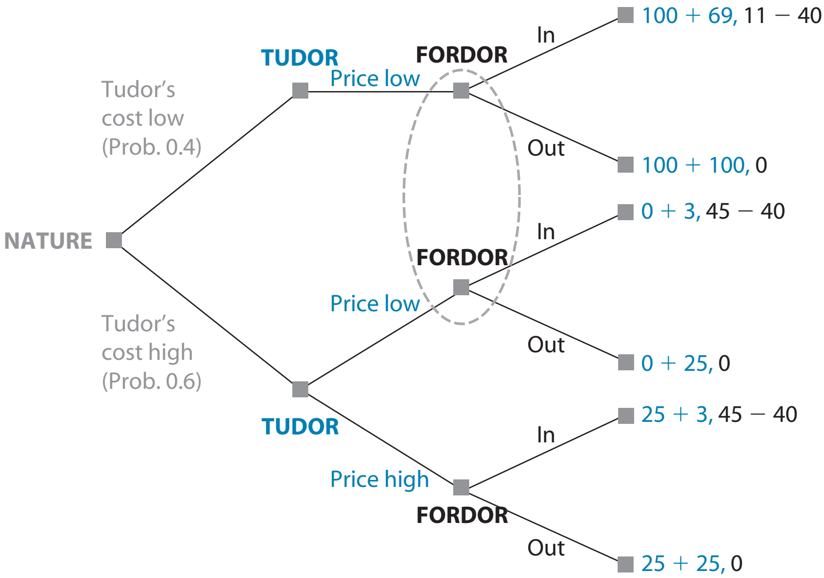


**FIGURE 8.7** Extensive Form of Entry Game: Tudor's Low Cost Is 5

price equals its cost in this case; it will make zero profit. Will this sacrifice of initial profit give Tudor the benefit of scaring Fordor off and enjoying the benefits of being a monopoly in subsequent periods?

We show the full game in extensive form in Figure 8.7. Note that we use the fictitious player called Nature, as in Section 3, to choose Tudor's cost type at the start of the game. Then Tudor makes its pricing decision. We assume that if Tudor has low cost, it will not choose a high price.[32] But if Tudor has high cost, it may choose either the high price or the low price if it wants to bluff. Fordor cannot tell apart the two situations in which Tudor prices low; therefore its entry choices at these two nodes are enclosed in one information set. Fordor must choose either In at both or Out at both.

At each terminal node, the first payoff entry (in blue) is Tudor's profit, and the second entry (in black) is Fordor's profit. Tudor's profit is added over two periods, the first period when it is the sole producer, and the second period when

[32] This seems obvious: Why choose a price different from the profit-maximizing price? Charging the high price when you have low cost not only sacrifices some profit in period 1 (if the low-cost Tudor charges 20, its sales will drop by so much that it will make a profit of only 75 instead of the 100 it gets by charging 15), but also increases the risk of entry and so lowers period-2 profits as well (competing with Fordor, the low-cost Tudor would have a profit of only 69 instead of the 100 it gets under monopoly). However, game theorists have found strange equilibria where a high period-1 price for Tudor is perversely interpreted as evidence of low cost, and they have applied great ingenuity in ruling out these equilibria. We leave out these complications, as we did in our analysis of cheap talk equilibria earlier, but refer interested readers to In-Koo Cho and David Kreps, "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, vol. 102, no. 2 (May 1987), pp. 179–222.

| | | FORDOR | |
|---|---|---|---|
| | | Regardless (II) | Conditional (OI) |
| TUDOR | Bluff (LL) | $169 \times 0.4 + 3 \times 0.6 = 69.4$, $-29 \times 0.4 + 5 \times 0.6 = -8.6$ | $200 \times 0.4 + 25 \times 0.6 = 95$, $0$ |
| | Honest (LH) | $169 \times 0.4 + 28 \times 0.6 = 84.4$, $-29 \times 0.4 + 5 \times 0.6 = -8.6$ | $200\times 0.4 + 28 \times 0.6 = 96.8$, $5 \times 0.6 = 3$ |

**FIGURE 8.8** Strategic Form of Entry Game: Tudor's Low Cost Is 5

it may be a monopolist or a duopolist, depending on whether Fordor enters. Fordor's profit covers only the second period and is non-zero only when it has chosen to enter.

Using one step of rollback analysis, we see that Fordor will choose In at the bottom node where Tudor has chosen the high price, because $45 - 40 = 5 > 0$. Therefore, we can prune the Out branch at that node. Then each player has just two strategies (complete plans of action). For Tudor the strategies are Bluff, or choose the low price in period 1 regardless of cost (LL in the shorthand notation of Chapter 3), and Honest, or choose the low price in period 1 if cost is low and the high price if cost is high (LH). For Fordor, the two strategies are Regardless, or enter irrespective of Tudor's period-1 price (II, for In-In), and Conditional, or enter only if Tudor's period-1 price is high (OI).

We can now show the game in strategic (normal) form. Figure 8.8 shows each player with two possible strategies; payoffs in each cell are the expected profits to each firm, given the probability (40%) that Tudor's cost is low. The calculations are similar to those we performed to fill in the table in Figure 8.6. As in that example, you may find the calculations easier if you label the terminal nodes in the tree and determine which ones are relevant for each cell of the table.

This is a simple dominance-solvable game. For Tudor, Honest dominates Bluff. And Fordor's best response to Tudor's dominant strategy of Honest is Conditional. Thus (Honest, Conditional) is the only (subgame-perfect) Nash equilibrium of the game.

The equilibrium found in Figure 8.8 is separating. The two cost types of Tudor charge different prices in period 1. This action reveals Tudor's type to Fordor, which then makes its entry decision appropriately.

The key to understanding why Honest is the dominant strategy for Tudor can be found in the comparison of its payoffs against Fordor's Conditional strategy. These are the outcomes when Tudor's bluff "works": Fordor enters if Tudor charges the high price in period 1 and stays out if Tudor charges the low price in period 1. If Tudor is truly low cost, then its payoffs against Fordor playing Conditional are the same whether it chooses Bluff or Honest. But when Tudor is actually high cost, the results are different.

If Fordor's strategy is Conditional and Tudor is high cost, Tudor can use Bluff successfully. However, even the successful bluff will be too costly. If Tudor charged its best monopoly (Honest) price in period 1, it would make a profit of 25; the bluffing low price reduces this period-1 profit drastically, in this instance all the way to 0. The higher monopoly price in period 1 would encourage Fordor's entry and reduce period-2 profit for Tudor, from the monopoly level of 25 to the duopoly level of 3. But Tudor's period-2 benefit from charging the low (Bluff) price and keeping Fordor out ($25 - 3 = 22$) is less than the period-1 cost imposed by bluffing and giving up its monopoly profits ($25 - 0 = 25$). As long as there is some positive probability that Tudor is high cost, then the benefits from choosing Honest will outweigh those from choosing Bluff, even when Fordor's choice is Conditional.

If the low price were not so low, then a truly high-cost Tudor would sacrifice less by mimicking the low-cost type. In such a case, Bluff might be a more profitable strategy for a high-cost Tudor. We consider exactly this possibility in the analysis below.

## C. Pooling Equilibrium

Let us now suppose that the lower of the production costs for Tudor is 10 per car instead of 5. With this cost change, the high-cost Tudor still makes profit of 25 under monopoly if it charges its profit-maximizing price of 20. But the low-cost Tudor now charges 17.5 as a monopolist (instead of 15) and makes a profit of 56. If the high-cost type mimics the low-cost type and also charges 17.5, its profit is now 19 (rather than the 0 it earned in this case before); the loss of profit from bluffing is now much smaller: $25 - 19 = 6$, rather than 25. If Fordor enters, then the two firms' profits in their duopoly game are 3 for Tudor and 45 for Fordor if Tudor has high costs (as in the previous section). Duopoly profits are now 25 for each firm if Tudor has low costs; in this situation, Fordor and the low-cost Tudor have identical unit costs of 10.

Suppose again that the probability of Tudor being the low-cost type is 40% (0.4) and Fordor's belief about the low-cost probability is correct. The new game tree is shown in Figure 8.9. Because Fordor will still choose In when Tudor prices High, the game again collapses to one in which each player has exactly two complete strategies; those strategies are the same ones we described in Section 6.B above. The payoff table for the normal form of this game is then the one illustrated in Figure 8.10.

This is another dominance-solvable game. Here it is Fordor with a dominant strategy, however; it will always choose Conditional. And given the dominance of Conditional, Tudor will choose Bluff. Thus, (Bluff, Conditional) is the unique (subgame-perfect) Nash equilibrium of this game. In all other cells of the table, one firm gains by deviating to its other action. We leave it to you

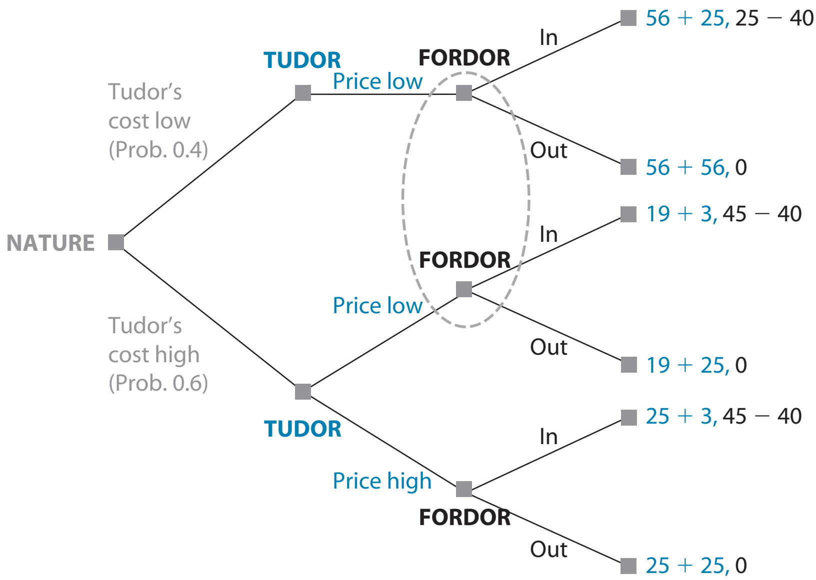


**FIGURE 8.9** Extensive Form of Entry Game: Tudor's Low Cost Is 10

to think about the intuitive explanations of why each of these deviations is profitable.

The equilibrium found using Figure 8.10 involves pooling. Both cost types of Tudor charge the same (low) price and, seeing this, Fordor stays out. When both types of Tudor charge the same price, observation of that price does not convey any information to Fordor. Its estimate of the probability of Tudor's cost being low stays at 0.4, and it calculates its expected profit from entry to be $-3 < 0$, so it does not enter. Even though Fordor knows full well that Tudor is bluffing in equilibrium, the risk of calling the bluff is too great because the probability of Tudor's cost actually being low is sufficiently great.

What if this probability were smaller—say, 0.1—and Fordor was aware of this fact? If all the other numbers remain unchanged, then Fordor's expected profit from its Regardless strategy is $-15 \times 0.1 + 5 \times 0.9 = 4.5 - 1.5 = 3 > 0$.

| | | FORDOR | |
|---|---|---|---|
| | | Regardless (II) | Conditional (OI) |
| TUDOR | Bluff (LL) | $81 \times 0.4 + 22 \times 0.6 = 45.6$, $-15 \times 0.4 + 5 \times 0.6 = -3$ | $112 \times 0.4 + 44 \times 0.6 = 71.2$, $0$ |
| | Honest (LH) | $81 \times 0.4 + 28 \times 0.6 = 49.2$, $-15 \times 0.4 + 5 \times 0.6 = -3$ | $112 \times 0.4 + 28 \times 0.6 = 61.6$, $5 \times 0.6 = 3$ |

**FIGURE 8.10** Strategic Form of Entry Game: Tudor's Low Cost Is 10

Then Fordor will enter no matter what price Tudor charges, and Tudor's bluff will not work. Such a situation results in a new kind of equilibrium; we consider its features below.

## D. Semiseparating Equilibrium

Here we consider the outcomes in the entry game when Tudor's probability of achieving the low production cost of 10 is small, only 10% (0.1). All of the cost and profit numbers are the same as in the previous section; only the probabilities have changed. Therefore, we do not show the game tree (Figure 8.9) again. We show only the payoff table as Figure 8.11.

In this new situation, the game illustrated in Figure 8.11 has no equilibrium in pure strategies. From (Bluff, Regardless), Tudor gains by deviating to Honest; from (Honest, Regardless), Fordor gains by deviating to Conditional; from (Honest, Conditional), Tudor gains by deviating to Bluff; and from (Bluff, Conditional), Fordor gains by deviating to Regardless. Once again, we leave it to you to think about the intuitive explanations of why each of these deviations is profitable.

So now we need to look for an equilibrium in mixed strategies. We suppose Tudor mixes Bluff and Honest with probabilities $p$ and $(1 - p)$, respectively. Similarly, Fordor mixes Regardless and Conditional with probabilities $q$ and $(1 - q)$, respectively. Tudor's $p$-mix must keep Fordor indifferent between its two pure strategies of Regardless and Conditional; therefore we need

$$3p + 3\,(1 - p) = 0p + 4.5\,(1 - p), \quad \text{or} \quad 4.5\,(1 - p) = 3, \quad \text{or}$$

$$1 - p = 2/3, \quad \text{or} \quad p = 1/3.$$

And Fordor's $q$-mix must keep Tudor indifferent between its two pure strategies of Bluff and Honest; therefore we need

$$27.9q + 50.8\,(1 - q) = 33.3q + 36.4\,(1 - q), \quad \text{or} \quad 5.4q = 14\,(1 - q), \quad \text{or}$$

$$q = 14.4/19.8 = 16/22 = 0.727.$$

The mixed-strategy equilibrium of the game then entails Tudor playing Bluff one-third of the time and Honest two-thirds of the time, while Fordor

| | | FORDOR | |
|---|---|---|---|
| | | Regardless (II) | Conditional (OI) |
| TUDOR | Bluff (LL) | $81 \times 0.1 + 22 \times 0.9 = 27.9$, $-15 \times 0.1 + 5 \times 0.9 = 3$ | $112 \times 0.1 + 44 \times 0.9 = 50.8$, $0$ |
| | Honest (LH) | $81 \times 0.1 + 28 \times 0.9 = 33.3$, $-15 \times 0.1 + 5 \times 0.9 = 3$ | $112 \times 0.1 + 28 \times 0.9 = 36.4$, $5 \times 0.9 = 4.5$ |

**FIGURE 8.11** Strategic Form of Entry Game: Tudor's Low Cost Is 10 with Probability 0.1

| | | TUDOR'S PRICE | | Sum of row |
|---|---|---|---|---|
| | | Low | High | |
| TUDOR'S COST | Low | 0.1 | 0 | 0.1 |
| | High | $0.9 \times 1/3 = 0.3$ | $0.9 \times 2/3 = 0.6$ | 0.9 |
| Sum of column | | 0.4 | 0.6 | |

**FIGURE 8.12** Applying Bayes' Theorem to the Entry Game

plays Regardless sixteen twenty-seconds of the time and Conditional six twenty-seconds of the time.

In this equilibrium, the Tudor types are only partially separated. The low-cost-type Tudor always prices Low in period 1, but the high-cost-type Tudor mixes and will also charge the low price one-third of the time. If Fordor observes a high price in period 1, it can be sure that Tudor has high cost; in that case, it will always enter. But if Fordor observes a low price, it does not know whether it faces a truly low-cost Tudor or a bluffing, high-cost Tudor. Then Fordor also plays a mixed strategy, entering 72.7% of the time. Thus, a high price conveys full information, but a low price conveys only partial information about Tudor's type. Therefore, this kind of equilibrium is labeled *semiseparating*.

To understand better the mixed strategies of each firm and the semiseparating equilibrium, consider how Fordor can use the partial information conveyed by Tudor's low price. If Fordor sees the low price in period 1, it will use this observation to update its belief about the probability that Tudor is low cost; it does this updating using Bayes' theorem.[33] The table of calculations is shown as Figure 8.12; this table is similar to Figure 8A.3 in the appendix.

The table shows the possible types of Tudor in the rows and the prices Fordor observes in the columns. The values in the cells represent the overall probability that a Tudor of the type shown in the corresponding row chooses the price shown in the corresponding column (incorporating Tudor's equilibrium mixture probability); the final row and column show the total probabilities of each type and of observing each price, respectively.

Using Bayes' rule, when Fordor observes Tudor charging a low period-1 price, it will revise its belief about the probability of Tudor being low cost by taking the probability that a low-cost Tudor is charging the low price (the 0.1 in the top-left cell) and dividing that by the total probability of the two types of Tudor choosing the low price (0.4, the column sum in the left column). This calculation yields Fordor's updated belief about the probability that Tudor has low costs to

[33] We provide a thorough explanation of Bayes' theorem in the appendix to this chapter. Here, we simply apply the analysis found there to our entry game.

be $0.1/0.4 = 0.25$. Then Fordor also updates its expected profit from entry to be $-15 \times 0.25 + 5 \times 0.75 = 0$. Thus, Tudor's equilibrium mixture is exactly right for making Fordor indifferent between entering and not entering when it sees the low period-1 price. This outcome is exactly what is needed to keep Tudor willing to mix in the equilibrium.

The original probability 0.1 of Tudor being low cost was too low to deter Fordor from entering. Fordor's revised probability of 0.25, after observing the low price in period 1, is higher. Why? Precisely because the high-cost-type Tudor is not always bluffing. If it were, then the low price would convey no information at all. Fordor's revised probability would equal 0.1 in that case, whereupon it would enter. But when the high-cost-type Tudor bluffs only sometimes, a low price is more likely to be indicative of low cost.

We developed the equilibria in this entry game in an intuitive way, but we now look back and think systematically about the nature of those equilibria. In each case, we first ensured that each player's (and each type's) strategy was optimal, given the strategies of everyone else; we applied the Nash concept of equilibrium. Second, we ensured that players drew the correct inference from their observations; this required a probability calculation using Bayes' theorem, most explicitly in the semiseparating equilibrium. The combination of concepts necessary to identify equilibria in such asymmetric information games justifies giving them the label **Bayesian Nash equilibria**. Finally, although this was a rather trivial part of this example, we did a little bit of rollback, or subgame perfectness, reasoning. The use of rollback justifies calling it the **perfect Bayesian equilibrium (PBE)** as well. Our example was a simple instance of all of these equilibrium concepts: you will meet some of them again in slightly more sophisticated forms in later chapters and in much fuller contexts in further studies of game theory.

## SUMMARY

When facing imperfect or incomplete information, game players with different attitudes toward risk or different amounts of information can engage in strategic behavior to control and manipulate the risk and information in a game. Players can reduce their risk with payment schemes or by sharing the risk with others, although the latter is complicated by *moral hazard* and *adverse selection*. Risk can sometimes be manipulated to a player's benefit, depending on the circumstances within the game.

Players with private information may want to conceal or reveal that information, while those without the information try to elicit it or avoid it. Actions speak louder than words in the presence of asymmetric information. To reveal information, a credible *signal* is required. In some cases, mere words may be sufficient to convey information credibly, and then a *cheap talk* equilibrium can arise. The extent to which player interests are aligned plays an important role in achieving such equilibria. When the information content of a player's words is ignored, the game has a *babbling equilibrium*.

More generally, specific actions taken by players convey information. *Signaling* works only if the signal action entails different costs to players with different information. To obtain information, when questioning is not sufficient to elicit truthful information, a *screening* scheme that looks for a specific action may be required. Screening works only if the *screening device* induces others to reveal their *types* truthfully; there must be *incentive compatibility* to get *separation*. At times, credible signaling or screening may not be possible; then the equilibrium can entail *pooling* or there can be a complete collapse of the market or transaction for one of the types. Many examples of signaling and screening games can be found in ordinary situations such as the labor market or in the provision of insurance. The evidence on players' abilities to achieve perfect Bayesian equilibria seems to suggest that, despite the difficult probability calculations necessary, such equilibria are often observed. Different experimental results appear to depend largely on the design of the experiment.

In the equilibrium of a game with asymmetric information, players must not only use their best actions given their information, but must also draw correct inferences (update their information) by observing the actions of others. This type of equilibrium is known as a *Bayesian Nash equilibrium*. When the further requirement of optimality at all nodes (as in rollback analysis) must be imposed, the equilibrium becomes a *perfect Bayesian equilibrium*. The outcome of such a game may entail pooling, separation, or *partial separation*, depending on the specifics of the payoff structure and the specified updating processes used by players. In some parameter ranges, such games may have multiples types of perfect Bayesian equilibria.

## KEY TERMS

**adverse selection (295)**
**babbling equilibrium (283)**
**Bayesian Nash equilibrium (319)**
**cheap talk equilibrium (281)**
**incentive-compatability condition (constraint) (306)**
**moral hazard (272)**
**negatively correlated (273)**
**partially revealing equilibrium (310)**
**participation condition (constraint) (306)**
**perfect Bayesian equilibrium (PBE) (319)**
**pooling (308)**
**pooling of types (308)**
**pooling equilibrium (281)**
**positively correlated (274)**
**screening (281)**
**screening device (281)**
**self-selection (307)**
**semiseparating equilibrium (310)**
**separating equilibrium (281)**
**separation of types (307)**

## SOLVED EXERCISES

**S1.** In the risk-trading example in Section 1, you had a risky income that was \$160,000 with good luck (probability 0.5) and \$40,000 with bad luck (probability 0.5). When your neighbor had a sure income of \$100,000, we derived a scheme in which you could eliminate all of your risk while raising his expected utility slightly. Assume that the utility of each of you is still the square root of the respective income. Now, however, let the probability of good luck be 0.6. Consider a contract that leaves you with exactly \$100,000 when you have bad luck. Let $x$ be the payment that you make to your neighbor when you have good luck.

**(a)** What is the minimum value of $x$ (to the nearest penny) such that your neighbor slightly prefers to enter into this kind of contract rather than no contract at all?

**(b)** What is the maximum value of $x$ (to the nearest penny) for which this kind of contract gives you a slightly higher expected utility than no contract at all?

**S2.** A local charity has been given a grant to serve free meals to the homeless in its community, but it is worried that its program might be exploited by nearby college students, who are always on the lookout for a free meal. Both a homeless person and a college student receive a payoff of 10 for a free meal. The cost of standing in line for the meal is $t^2/320$ for a homeless person and $t^2/160$ for a college student, where $t$ is the amount of time in line measured in minutes. Assume that the staff of the charity cannot observe the true type of those coming for free meals.

**(a)** What is the minimum wait time $t$ that will achieve separation of types?

**(b)** After a while, the charity finds that it can successfully identify and turn away college students half of the time. College students who are turned away receive no free meal and, further, incur a cost of 5 for their time and embarrassment. Will the partial identification of college students reduce or increase the answer in part (a)? Explain.

**S3.** Consider the used-car market for the 2011 Citrus described in Section 4.B. There is now a surge in demand for used Citruses; buyers would now be willing to pay up to \$18,000 for an orange and \$8,000 for a lemon. All else remains identical to the example in Section 4.B.

(a) What price would buyers be willing to pay for a 2011 Citrus of unknown type if the fraction of oranges in the population, $f$, were 0.6?
(b) Will there be a market for oranges if $f = 0.6$? Explain.
(c) What price would buyers be willing to pay if $f$ were 0.2?
(d) Will there be a market for oranges if $f = 0.2$? Explain.
(e) What is the minimum value of $f$ such that the market for oranges does not collapse?
(f) Explain why the increase in the buyers' willingness to pay changes the threshold value of $f$, where the market for oranges collapses.

**S4.** Suppose electricians come in two types: competent and incompetent. Both types of electricians can get certified, but for the incompetent types certification takes extra time and effort. Competent ones have to spend $C$ months preparing for the certification exam; incompetent ones take twice as long. Certified electricians can earn 100 (thousand dollars) each year working on building sites for licensed contractors. Uncertified electricians can earn only 25 (thousand dollars) each year in freelance work; licensed contractors won't hire them. Each type of electrician gets a payoff equal to $\sqrt{S} - M$, where $S$ is the salary measured in thousands of dollars and $M$ is the number of months spent getting certified. What is the range of values of $C$ for which a competent electrician will choose to signal with this device but an incompetent one will not?

**S5.** Return to the Tudor-Fordor example in Section 6.A, when Tudor's low per-unit cost is 5. Let $z$ be the probability that Tudor actually has a low per-unit cost.
(a) Rewrite the table in Figure 8.8 in terms of $z$.
(b) How many pure-strategy equilibria are there when $z = 0$? Explain.
(c) How many pure-strategy equilibria are there when $z = 1$? Explain.
(d) Show that the Nash equilibrium of this game is always a separating equilibrium for any value of $z$ between 0 and 1 (inclusive).

**S6.** Looking at Tudor and Fordor again, assume that the old, established company Tudor is risk averse, whereas the would-be entrant Fordor (which is planning to finance its project through venture capital) is risk neutral. That is, Tudor's utility is always the square root of its total profit over both periods. Fordor's utility is simply the amount of its profit—if any—during the second period. Assume that Tudor's low per-unit cost is 5, as in Section 6.A.
(a) Redraw the extensive-form game shown in Figure 8.7, giving the proper payoffs for a risk-averse Tudor.

(b) Let the probability that Tudor is low cost, $z$, be 0.4. Will the equilibrium be separating, pooling, or semiseparating? (Hint: Use a table equivalent to Figure 8.8.)

(c) Repeat part (b) with $z = 0.1$.

**S7.** Return to a risk-neutral Tudor, but with a low per-unit cost of 6 (instead of 5 or 10 as in Section 6). If Tudor's cost is low, 6, then it will earn 90 in a profit-maximizing monopoly. If Fordor enters, Tudor will earn 59 in the resulting duopoly while Fordor earns 13. If Tudor is actually high cost (that is, its per-unit cost is 15) and prices as if it were low cost (that is, with a per-unit cost of 6), then it earns 5 in a monopoly situation.

(a) Draw a game tree for this game equivalent to Figure 8.7 or 8.9 in the text, changing the appropriate payoffs.

(b) Write the normal form of this game, assuming that the probability that Tudor is low price is 0.4.

(c) What is the equilibrium of the game? Is it separating, pooling, or semiseparating? Explain why.

**S8.** Felix and Oscar are playing a simplified version of poker. Each makes an initial bet of 8 dollars. Then each separately draws a card, which may be High or Low with equal probabilities. Each sees his own card but not that of the other.

Then Felix decides whether to Pass or to Raise (bet an additional 4 dollars). If he chooses to pass, the two cards are revealed and compared. If the outcomes are different, the one who has the High card collects the whole pot. The pot has 16 dollars, of which the winner himself contributed 8, so his winnings are 8 dollars. The loser's payoff is $-8$ dollars. If the outcomes are the same, the pot is split equally and each gets his 8 dollars back (payoff 0).

If Felix chooses Raise, then Oscar has to decide whether to Fold (concede) or See (match with his own additional 4 dollars). If Oscar chooses Fold, then Felix collects the pot irrespective of the cards. If Oscar chooses See, then the cards are revealed and compared. The procedure is the same as that in the preceding paragraph, but the pot is now bigger.

(a) Show the game in extensive form. (Be careful about information sets.)

If the game is instead written in the normal form, Felix has four strategies: (1) Pass always (PP for short), (2) Raise always (RR), (3) Raise if his own card is High and Pass if it is Low (RP), and (4) the other way round (PR). Similarly, Oscar has four strategies: (1) Fold always (FF), (2) See always (SS), (3) See if his own card is High and Fold if it is Low (SF), and (4) the other way round (FS).

(b) Show that the table of payoffs to Felix is as follows:

| | | OSCAR | | | |
|---|---|---|---|---|---|
| | | FF | SS | SF | FS |
| FELIX | PP | 0 | 0 | 0 | 0 |
| | RR | 8 | 0 | 1 | 7 |
| | RP | 2 | 1 | 0 | 3 |
| | PR | 6 | –1 | 1 | 4 |

(In each case, you will have to take an expected value by averaging over the consequences for each of the four possible combinations of the card draws.)

(c) Eliminate dominated strategies as far as possible. Find the mixed-strategy equilibrium in the remaining table and the expected payoff to Felix in the equilibrium.

(d) Use your knowledge of the theory of signaling and screening to explain intuitively why the equilibrium has mixed strategies.

**S9.** Felix and Oscar are playing another simplified version of poker called Stripped-Down Poker. Both make an initial bet of one dollar. Felix (and only Felix) draws one card, which is either a King or a Queen with equal probability (there are four Kings and four Queens). Felix then chooses whether to Fold or to Bet. If Felix chooses to Fold, the game ends, and Oscar receives Felix's dollar in addition to his own. If Felix chooses to Bet, he puts in an additional dollar, and Oscar chooses whether to Fold or to Call.

If Oscar Folds, Felix wins the pot (consisting of Oscar's initial bet of one dollar and two dollars from Felix). If Oscar Calls, he puts in another dollar to match Felix's bet, and Felix's card is revealed. If the card is a King, Felix wins the pot (two dollars from each of the roommates). If it is a Queen, Oscar wins the pot.

(a) Show the game in extensive form. (Be careful about information sets.)

(b) How many strategies does each player have?

(c) Show the game in strategic form, where the payoffs in each cell reflect the expected payoffs given each player's respective strategy.

(d) Eliminate dominated strategies, if any. Find the equilibrium in mixed strategies. What is the expected payoff to Felix in equilibrium?

**S10.** Wanda works as a waitress and consequently has the opportunity to earn cash tips that are not reported by her employer to the Internal Revenue Service. Her tip income is rather variable. In a good year (G), she earns a high income, so her tax liability to the IRS is \$5,000. In a bad year (B), she earns a low income, and her tax liability to the IRS is \$0. The IRS knows that the probability of her having a good year is 0.6, and the probability of her having a bad year is 0.4, but it doesn't know for sure which outcome has resulted for her this tax year.

In this game, first Wanda decides how much income to report to the IRS. If she reports high income (H), she pays the IRS \$5,000. If she reports low income (L), she pays the IRS \$0. Then the IRS has to decide whether to audit Wanda. If she reports high income, they do not audit, because they automatically know they're already receiving the tax payment Wanda owes. If she reports low income, then the IRS can either audit (A) or not audit (N). When the IRS audits, it costs the IRS \$1,000 in administrative costs, and also costs Wanda \$1,000 in the opportunity cost of the time spent gathering bank records and meeting with the auditor. If the IRS audits Wanda in a bad year (B), then she owes nothing to the IRS, although she and the IRS have each incurred the \$1,000 auditing cost. If the IRS audits Wanda in a good year (G), then she has to pay the \$5,000 she owes to the IRS, in addition to her and the IRS each incurring the cost of auditing.

**(a)** Suppose that Wanda has a good year (G), but she reports low income (L). Suppose the IRS then audits her (A). What is the total payoff to Wanda, and what is the total payoff to the IRS?

**(b)** Which of the two players has an incentive to bluff (that is, to give a false signal) in this game? What would bluffing consist of?

**(c)** Show this game in extensive form. (Be careful about information sets.)

**(d)** How many pure strategies does each player have in this game? Explain your reasoning.

**(e)** Write down the strategic-form game matrix for this game. Find all of the Nash equilibria to this game. Identify whether the equilibria you find are separating, pooling, or semiseparating.

**(f)** Let $x$ equal the probability that Wanda has a good year. In the original version of this problem, we had $x = 0.6$. Find a value of $x$ such that Wanda always reports low income in equilibrium.

**(g)** What is the full range of values of $x$ for which Wanda always reports low income in equilibrium?

**S11.** The design of a health-care system concerns matters of information and strategy at several points. The users—potential and actual patients—

have better information about their own state of health, lifestyle, and so forth—than the insurance companies can find out. The providers—doctors, hospitals, and so forth—know more about what the patients need than do either the patients themselves or the insurance companies. Doctors also know more about their own skills and efforts, and hospitals about their own facilities. Insurance companies may have some statistical information about outcomes of treatments or surgical procedures from their past records. But outcomes are affected by many unobservable and random factors, so the underlying skills, efforts, or facilities cannot be inferred perfectly from observation of the outcomes. The pharmaceutical companies know more about the efficacy of drugs than do the others. As usual, the parties do not have natural incentives to share their information fully or accurately with others. The design of the overall scheme must try to face these matters and find the best feasible solutions.

Consider the relative merits of various payment schemes—fee for service versus capitation fees to doctors, comprehensive premiums per year versus payment for each visit for patients, and so forth—from this strategic perspective. Which are likely to be most beneficial to those seeking health care? To those providing health care? Think also about the relative merits of private insurance and coverage of costs from general tax revenues.

**S12.** In a television commercial for a well-known brand of instant cappuccino, a gentleman is entertaining a lady friend at his apartment. He wants to impress her and offers her cappuccino with dessert. When she accepts, he goes into the kitchen to make the instant cappuccino—simultaneously tossing take-out boxes into the trash and faking the noises made by a high-class (and expensive) espresso machine. As he is doing so, a voice comes from the other room: "I want to see the machine . . . ."

Use your knowledge of games of asymmetric information to comment on the actions of these two people. Pay attention to their attempts to use signaling and screening, and point out specific instances of each strategy. Offer an opinion about which player is the better strategist.

**S13.** (**Optional, requires appendix**) In the genetic test example, suppose the test comes out negative ($Y$ is observed). What is the probability that the person does not have the defect ($B$ exists)? Calculate this probability by applying Bayes' rule, and then check your answer by doing an enumeration of the 10,000 members of the population.

**S14.** (**Optional, requires appendix**) Return to the example of the 2011 Citrus in Section 4.B. The two types of Citrus—the reliable orange and the hapless lemon—are outwardly indistinguishable to a buyer. In the example, if the fraction $f$ of oranges in the Citrus population is less than 0.65, the seller of an orange will not be willing to part with the car for the maximum price buyers are willing to pay, so the market for oranges collapses.

But what if a seller has a costly way to signal her car's type? Although oranges and lemons are in nearly every respect identical, the defining difference between the two is that lemons break down much more frequently. Knowing this, owners of oranges make the following proposal. On a buyer's request, the seller will in one day take a 500-mile round-trip drive in the car. (Assume this trip will be verifiable via odometer readings and a time-stamped receipt from a gas station 250 miles away.) For the sellers of both types of Citrus, the cost of the trip in fuel and time is \$0.50 per mile (that is, \$250 for the 500-mile trip). However, with probability $q$ a lemon attempting the journey will break down. If a car breaks down, the cost is \$2 per mile of the total length of the attempted road trip (that is, \$1,000). Additionally, breaking down will be a sure sign that the car is a lemon, so a Citrus that does so will sell for only \$6,000.

Assume that the fraction of oranges in the Citrus population, $f$, is 0.6. Also, assume that the probability of a lemon breaking down, $q$, is 0.5 and that owners of lemons are risk neutral.

(a) Use Bayes' rule to determine $f_{\text{updated}}$, the fraction of Citruses that have successfully completed a 500-mile road trip that are oranges. Assume that all Citrus owners attempt the trip. Is $f_{\text{updated}}$ greater than or less than $f$? Explain why.

(b) Use $f_{\text{updated}}$ to determine the price, $p_{\text{updated}}$, that buyers are willing to pay for a Citrus that has successfully completed the 500-mile road trip.

(c) Will an owner of an orange be willing to make the road trip and sell her car for $p_{\text{updated}}$? Why or why not?

(d) What is the expected payoff of attempting the road trip to the seller of a lemon?

(e) Would you describe the outcome of this market as pooling, separating, or semiseparating? Explain.

## UNSOLVED EXERCISES

**U1.** Jack is a talented investor, but his earnings vary considerably from year to year. In the coming year he expects to earn either \$250,000 with good luck or \$90,000 with bad luck. Somewhat oddly, given his chosen profession, Jack is risk averse, so that his utility is equal to the square root of his income. The probability of Jack's having good luck is 0.5.

(a) What is Jack's expected utility for the coming year?

(b) What amount of certain income would yield the same level of utility for Jack as the expected utility in part (a)?

Jack meets Janet, whose situation is identical in every respect. She's an investor who will earn \$250,000 in the next year with good luck and \$90,000 with bad, she's risk averse with square-root utility, and her probability of having good luck is 0.5. Crucially, it turns out that Jack and Janet invest in such a way that their luck is completely independent. They agree to the following deal. Regardless of their respective luck, they will always pool their earnings and then split them equally.

(c) What are the four possible luck-outcome pairs, and what is the probability of reaching each one?

(d) What is the expected utility for Jack or Janet under this arrangement?

(e) What amount of certain income would yield the same level of utility for Jack and Janet as in part (d)?

Incredibly, Jack and Janet then meet Chrissy, who is also identical to Jack and Janet with respect to her earnings, utility, and luck. Chrissy's probability of good luck is independent from either Jack's or Janet's. After some discussion, they decide that Chrissy should join the agreement of Jack and Janet. All three of them will pool their earnings and then split them equally three ways.

(f) What are the eight possible luck-outcome triplets, and what is the probability of reaching each of them?

(g) What is the expected utility for each of the investors under this expanded arrangement?

(h) What amount of certain income would yield the same level of utility as in part (g) for these risk-averse investors?

**U2.** Consider again the case of the 2011 Citrus. Almost all cars depreciate over time, and so it is with the Citrus. Every month that passes, all sellers of Citruses—regardless of type—are willing to accept \$100 less than they were the month before. Also, with every passing month, buyers are maximally willing to pay \$400 less for an orange than they were the previous month and \$200 less for a lemon. Assume that the example in the text takes place in month 0. Eighty percent of the Citruses are oranges, and this proportion never changes.

**(a)** Fill out three versions of the following table for month 1, month 2, and month 3:

| | Willingness to accept of sellers | Willingness to pay of buyers |
|---|---|---|
| Orange | | |
| Lemon | | |

**(b)** Graph the willingness to accept of the sellers of oranges over the next 12 months. On the same figure, graph the price that buyers are willing to pay for a Citrus of unknown type (given that the proportion of oranges is 0.8). (Hint: Make the vertical axis range from 10,000 to 14,000.)

**(c)** Is there a market for oranges in month 3? Why or why not?

**(d)** In what month does the market for oranges collapse?

**(e)** If owners of lemons experienced no depreciation (that is, they were never willing to accept anything less than \$3,000), would this affect the timing of the collapse of the market for oranges? Why or why not? In what month does the market for oranges collapse in this case?

**(f)** If buyers experienced no depreciation for a lemon (that is, they were always willing to pay up to \$6,000 for a lemon), would this affect the timing of the collapse of the market for oranges? Why or why not? In what month does the market for oranges collapse in this case?

**U3.** An economy has two types of jobs, Good and Bad, and two types of workers, Qualified and Unqualified. The population consists of 60% Qualified and 40% Unqualified. In a Bad job, either type of worker produces 10 units of output. In a Good job, a Qualified worker produces 100 units, and an Unqualified worker produces 0. There is enough demand for workers that for each type of job, companies must pay what they expect the appointee to produce.

Companies must hire each worker without observing his type and pay him before knowing his actual output. But Qualified workers can signal their qualification by getting educated. For a Qualified worker, the cost of getting educated to level $n$ is $n^2/2$, whereas for an Unqualified worker, it is $n^2$. These costs are measured in the same units as output, and $n$ must be an integer.

**(a)** What is the minimum level of $n$ that will achieve separation?

**(b)** Now suppose the signal is made unavailable. Which kind of jobs will be filled by which kinds of workers and at what wages? Who will gain and who will lose from this change?

**U4.** You are the Dean of the Faculty at St. Anford University. You hire Assistant Professors for a probationary period of 7 years, after which they come up for tenure and are either promoted and gain a job for life or turned down, in which case they must find another job elsewhere.

Your Assistant Professors come in two types, Good and Brilliant. Any types worse than Good have already been weeded out in the hiring process, but you cannot directly distinguish between Good and Brilliant types. Each individual Assistant Professor knows whether he or she is Brilliant or merely Good. You would like to tenure only the Brilliant types.

The payoff from a tenured career at St. Anford is \$2 million; think of this as the expected discounted present value of salaries, consulting fees, and book royalties, plus the monetary equivalent of the pride and joy that the faculty member and his or her family would get from being tenured at St. Anford. Anyone denied tenure at St. Anford will get a faculty position at Boondocks College, and the present value of that career is \$0.5 million.

Your faculty can do research and publish the findings. But each publication requires effort and time and causes strain on the family; all these are costly to the faculty member. The monetary equivalent of this cost is \$30,000 per publication for a Brilliant Assistant Professor and \$60,000 per publication for a Good one. You can set a minimum number, N, of publications that an Assistant Professor must produce in order to achieve tenure.

(a) Without doing any math, describe, as completely as you can, what would happen in a separating equilibrium to this game.

(b) There are two potential types of pooling outcomes to this game. Without doing any math, describe what they would look like, as completely as you can.

(c) Now please go ahead and do some math. What is the set of possible N that will accomplish your goal of screening the Brilliant professors out from the merely Good ones?

**U5.** Return to the Tudor-Fordor problem from Section 6.C, when Tudor's low per-unit cost is 10. Let $z$ be the probability that Tudor actually has a low per-unit cost.

(a) Rewrite the table in Figure 8.10 in terms of $z$.

(b) How many pure-strategy equilibria are there when $z = 0$? What type of equilibrium (separating, pooling, or semiseparating) occurs when $z = 0$? Explain.

(c) How many pure-strategy equilibria are there when $z = 1$? What type of equilibrium (separating, pooling, or semiseparating) occurs when $z = 1$? Explain.

**(d)** What is the lowest value of $z$ such that there is a pooling equilibrium?

**(e)** Explain intuitively why the pooling equilibrium cannot occur when the value of $z$ is too low.

**U6.** Assume that Tudor is risk averse, with square-root utility over its total profit (see Exercise S6), and that Fordor is risk neutral. Also, assume that Tudor's low per-unit cost is 10, as in Section 6.C.

**(a)** Redraw the extensive-form game shown in Figure 8.9, giving the proper payoffs for a risk-averse Tudor.

**(b)** Let the probability that Tudor is low cost, $z$, be 0.4. Will the equilibrium be separating, pooling, or semiseparating? (Hint: Use a table equivalent to Figure 8.10.)

**(c)** Repeat part (b) with $z = 0.1$.

**(d)** **(Optional)** Will Tudor's risk aversion change the answer to part (d) of Exercise U5? Explain why or why not.

**U7.** Return to the situation in Exercise S7, where Tudor's low per-unit cost is 6.

**(a)** Write the normal form of this game in terms of $z$, the probability that Tudor is low price.

**(b)** What is the equilibrium when $z = 0.1$? Is it separating, pooling, or semiseparating?

**(c)** Repeat part (b) for $z = 0.2$.

**(d)** Repeat part (b) for $z = 0.3$.

**(e)** Compare your answers in parts (b), (c), and (d) of this problem with part (d) of Exercise U5. When Tudor's low cost is 6 instead of 10, can pooling equilibria be achieved at lower values of $z$? Or are higher values of $z$ required for pooling equilibria to occur? Explain intuitively why this is the case.

**U8.** Corporate lawsuits may sometimes be signaling games. Here is one example. In 2003, AT&T filed suit against eBay, alleging that its Billpoint and PayPal electronic-payment systems infringed on AT&T's 1994 patent on "mediation of transactions by a communications system."

Let's consider this situation from the point in time when the suit was filed. In response to this suit, as in most patent-infringement suits, eBay can offer to settle with AT&T without going to court. If AT&T accepts eBay's settlement offer, there will be no trial. If AT&T rejects eBay's settlement offer, the outcome will be determined by the court.

The amount of damages claimed by AT&T is not publicly available. Let's assume that AT&T is suing for \$300 million. In addition, let's assume that if the case goes to trial, the two parties will incur court costs (paying lawyers and consultants) of $10 million each.

Because eBay is actually in the business of processing electronic payments, we might think that eBay knows more than AT&T does about its probability of winning the trial. For simplicity, let's assume that eBay knows for sure whether it will be found innocent (i) or guilty (g) of patent infringement. From AT&T's point of view, there is a 25% chance that eBay is guilty (g) and a 75% chance that eBay is innocent (i).

Let's also suppose that eBay has two possible actions: a generous settlement offer (G) of $200 million or a stingy settlement offer (S) of $20 million. If eBay offers a generous settlement, assume that AT&T will accept, thus avoiding a costly trial. If eBay offers a stingy settlement, then AT&T must decide whether to accept (A) and avoid a trial or reject and take the case to court (C). In the trial, if eBay is guilty, it must pay AT&T $300 million in addition to paying all the court costs. If eBay is found innocent, it will pay AT&T nothing, and AT&T will pay all the court costs.

**(a)** Show the game in extensive form. (Be careful to label information sets correctly.)

**(b)** Which of the two players has an incentive to bluff (that is, to give a false signal) in this game? What would bluffing consist of? Explain your reasoning.

**(c)** Write the strategic-form game matrix for this game. Find all of the Nash equilibria to this game. What are the expected payoffs to each player in equilibrium?

**U9.** For the Stripped-Down Poker game that Felix and Oscar play in Exercise S9, what does the mix of Kings and Queens have to be for the game to be fair? That is, what fraction of Kings will make the expected payoff of the game zero for both players?

**U10.** Bored with Stripped-Down Poker, Felix and Oscar now make the game more interesting by adding a third card type: Jack. Four Jacks are added to the deck of four Kings and four Queens. All rules remain the same as before, except for what happens when Felix Bets and Oscar Calls. When Felix Bets and Oscar Calls, Felix wins the pot if he has a King, they "tie" and each gets his money back if Felix is holding a Queen, and Oscar wins the pot if the card is a Jack.

**(a)** Show the game in extensive form. (Be careful to label information sets correctly.)

**(b)** How many pure strategies does Felix have in this game? Explain your reasoning.

**(c)** How many pure strategies does Oscar have in this game? Explain your reasoning.

(d) Represent this game in strategic form. This should be a matrix of *expected* payoffs for each player, given a pair of strategies.

(e) Find the unique pure-strategy Nash equilibrium of this game.

(f) Would you call this a pooling equilibrium, a separating equilibrium, or a semiseparating equilibrium?

(g) In equilibrium, what is the expected payoff to Felix of playing this game? Is it a fair game?

**U11.** Consider Spence's job-market signaling model with the following specifications. There are two types of workers, 1 and 2. The productivities of the two types, as functions of the level of education $E$, are

$$W_1(E) = E \quad \text{and} \quad W_2(E) = 1.5E.$$

The costs of education for the two types, as functions of the level of education, are

$$C_1(E) = E^2/2 \quad \text{and} \quad C_2(E) = E^2/3.$$

Each worker's utility equals his or her income minus the cost of education. Companies that seek to hire these workers are perfectly competitive in the labor market.

(a) If types are public information (observable and verifiable), find expressions for the levels of education, incomes, and utilities of the two types of workers.

Now suppose each worker's type is his or her private information.

(b) Verify that if the contracts of part (a) are attempted in this situation of information asymmetry, then type 2 does not want to take up the contract intended for type 1, but type 1 does want to take up the contract intended for type 2, so "natural" separation cannot prevail.

(c) If we leave the contract for type 1 as in part (a), what is the range of contracts (education-wage pairs) for type 2 that can achieve separation?

(d) Of the possible separating contracts, which one do you expect to prevail? Give a verbal but not a formal explanation for your answer.

(e) Who gains or loses from the information asymmetry? How much?

**U12.** "Mr. Robinson pretty much concludes that business schools are a sifting device—M.B.A. degrees are union cards for yuppies. But perhaps the most important fact about the Stanford business school is that all meaningful sifting occurs before the first class begins. No messy weeding is done within the walls. 'They don't want you to flunk. They want you to become a rich alum who'll give a lot of money to the school.' But one wonders: If corporations are abdicating to the Stanford admissions office the responsibility for selecting young managers, why don't they simply replace their personnel departments with Stanford admissions officers, and eliminate the spurious education? Does the very act of throwing away a lot of money and two years of one's life demonstrate a commitment to business that employers find appealing?" (From the review by Michael Lewis of Peter Robinson's *Snapshots from Hell: The Making of an MBA*, in the *New York Times*, May 8, 1994, Book Review section.) What answer to Lewis's question can you give, based on our analysis of strategies in situations of asymmetric information?

**U13.** **(Optional, requires appendix)** An auditor for the IRS is reviewing Wanda's latest tax return (see Exercise S10), on which she reports having had a bad year. Assume that Wanda is playing according to her equilibrium strategy and that the auditor knows this.

(a) Using Bayes' rule, find the probability that Wanda had a good year given that she reports having had a bad year.

(b) Explain why the answer in part (a) is more or less than the baseline probability of having a good year, 0.6.

**U14.** **(Optional, requires appendix)** Return to Exercise S14. Assume, reasonably, that the probability of a lemon's breaking down increases over the length of the road trip. Specifically, let $q = m/(m + 500)$, where $m$ is the number of miles in the round trip.

(a) Find the minimum integer number of miles, $m$, necessary to avoid the collapse of the market for oranges. That is, what is the smallest $m$ such that the seller of an orange is willing to sell her car at the market price for a Citrus that has successfully completed the road trip? (Hint: Remember to calculate $f_{\text{updated}}$ and $p_{\text{updated}}$.)

(b) What is the minimum integer number of miles, $m$, necessary to achieve complete separation between functioning markets for oranges and lemons? That is, what is the smallest $m$ such that the owner of a lemon will never decide to attempt the road trip?

# Appendix: Risk Attitudes and Bayes' Theorem

## 1 ATTITUDES TOWARD RISK AND EXPECTED UTILITY

In Chapter 2, we pointed out a difficulty about using probabilities to calculate the average or expected payoff for players in a game. Consider a game where players gain or lose money, and suppose we measure payoffs simply in money amounts. If a player has a 75% chance of getting nothing and a 25% chance of getting \$100, then the expected payoff is calculated as a *probability-weighted average;* the expected payoff is the average of the different payoffs with the probabilities of each as weights. In this case, we have \$0 with a probability of 75%, which yields $0.75 \times 0 = 0$ on average, added to \$100 with a probability of 25%, which yields $0.25 \times 100 = 25$ on average. That is the same payoff as the player would get from a simple nonrandom outcome that guaranteed him \$25 every time he played. People who are indifferent between two alternatives with the same average monetary value but different amounts of risk are said to be **risk-neutral**. In our example, one prospect is riskless (\$25 for sure), while the other is risky, yielding either \$0 with a probability of 0.75 or \$100 with a probability of 0.25, for the same average of \$25. In contrast are **risk-averse** people—those who, given a pair of alternatives each with the same average monetary value, would prefer the less risky option. In our example, they would rather get \$25 for sure than face the risky \$100-or-nothing prospect and, given the choice, would pick the safe prospect. Such risk-averse behavior is quite common; we should therefore have a theory of decision making under uncertainty that takes it into account.

We also said in Chapter 2 that a very simple modification of our payoff calculation can get us around this difficulty. We said that we could measure payoffs not in money sums but by using a nonlinear rescaling of the dollar amounts. Here we show explicitly how that rescaling can be done and why it solves our problem for us.

Suppose that, when a person gets $D$ dollars, we define the payoff to be something other than just $D$, perhaps $\sqrt{D}$. Then the payoff number associated with \$0 is 0, and that for \$100 is 10. This transformation does not change the way in which the person rates the two payoffs of \$0 and \$100; it simply rescales the payoff numbers in a particular way.

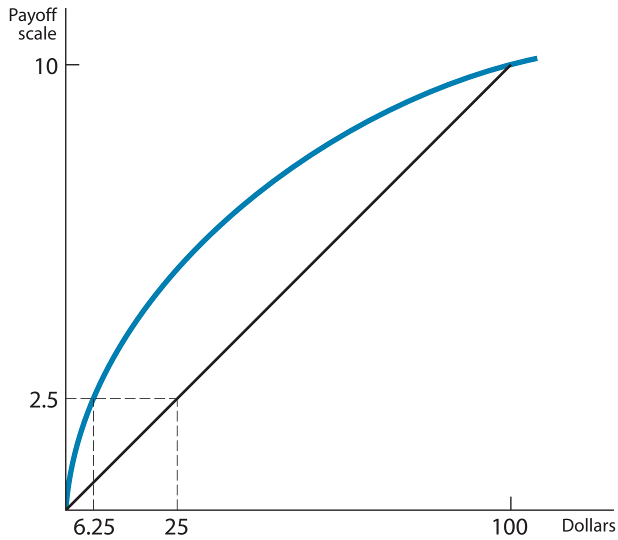


**FIGURE 8A.1** Concave Scale: Risk Aversion

Now consider the risky prospect of getting \$100 with probability 0.25 and nothing otherwise. After our rescaling, the expected payoff (which is the average of the two payoffs with the probabilities as weights) is $(0.75 \times 0) + (0.25 \times 10) = 2.5$. This expected payoff is equivalent to the person's getting the dollar amount whose square root is 2.5; because $2.5 = \sqrt{6.25}$, a person getting \$6.25 for sure would also receive a payoff of 2.5. In other words, the person with our square-root payoff scale would be just as happy getting \$6.25 for sure as he would getting a 25% chance at \$100. This indifference between a guaranteed \$6.25 and a 1 in 4 chance of \$100 indicates quite a strong aversion to risk; this person is willing to give up the difference between \$25 and \$6.25 to avoid facing the risk. Figure 8A.1 shows this nonlinear scale (the square root), the expected payoff, and the person's indifference between the sure prospect and the gamble.

What if the nonlinear scale that we use to rescale dollar payoffs is the cube root instead of the square root? Then the payoff from \$100 is 4.64, and the expected payoff from the gamble is $(0.75 \times 0) + (0.25 \times 4.64) = 1.16$, which is the cube root of 1.56. Therefore, a person with this payoff scale would accept only \$1.56 for sure instead of a gamble that has a money value of \$25 on average; such a person is extremely risk-averse indeed. (Compare a graph of the cube root of $x$ with a graph of the square root of $x$ to see why this should be so.)

And what if the rescaling of payoffs from $x$ dollars is done by using the function $x^2$? Then the expected payoff from the gamble is $(0.75 \times 0) + (0.25 \times 10{,}000) = 2{,}500$, which is the square of 50. Therefore, a person with this payoff scale would be indifferent between getting \$50 for sure and the gamble with an

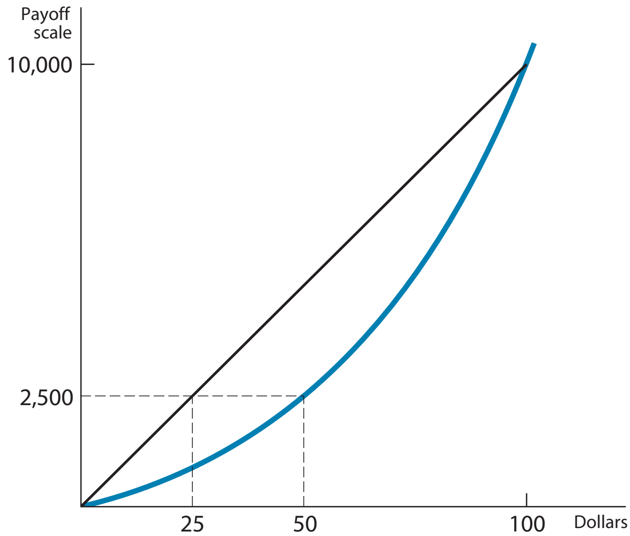


**FIGURE 8A.2** Convex Scale: Risk Loving

expected money value of only \$25. This person must be a risk lover because he is not willing to give up any money to get a reduction in risk; on the contrary, he must be given an extra \$25 in compensation for the loss of risk. Figure 8A.2 shows the nonlinear scale associated with a function such as $x^2$.

So by using different nonlinear scales instead of pure money payoffs, we can capture different degrees of risk-averse or risk-loving behavior. A concave scale like that of Figure 8A.1 corresponds to risk aversion, and a convex scale like that of Figure 8A.2 corresponds to risk-loving behavior. You can experiment with different simple nonlinear scales—for example, logarithms, exponentials, and other roots and powers—to see what they imply about attitudes toward risk.[34]

This method of evaluating risky prospects has a long tradition in decision theory; it is called the expected utility approach. The nonlinear scale that gives payoffs as functions of money values is called the **utility function**; the square root, cube root, and square functions referred to earlier are simple examples. Then the mathematical expectation, or probability-weighted average, of the utility values of the different money sums in a random prospect is called the **expected utility** of that prospect. And different random prospects are compared with one another in terms of their expected utilities; prospects with higher expected utility are judged to be better than those with lower expected utility.

[34] Additional information on the use of expected utility and risk attitudes of players can be found in many intermediate microeconomic texts; for example, Hal Varian, *Intermediate Microeconomics*, 7th ed. (New York: W. W. Norton & Company, 2006), ch. 12; Walter Nicholson and Christopher Snyder, *Microeconomic Theory*, 10th ed. (New York: Dryden Press, 2008), ch. 7.

Almost all of game theory is based on the expected utility approach, and it is indeed very useful, although it is not without flaws. We will adopt it in this book, leaving more detailed discussions to advanced treatises.[35]

## 2 INFERRING PROBABILITIES FROM OBSERVING CONSEQUENCES

When players have different amounts of information in a game, they will try to use some device to ascertain their opponents' private information. As we saw in Section 3 of this chapter, it is sometimes possible for direct communication to yield a cheap talk equilibrium. But more often, players will need to determine one another's information by observing one another's actions. They then must estimate the probabilities of the underlying information by using those actions or their observed consequences. This estimation requires some relatively sophisticated manipulation of the rules of probability, and we examine this process in detail here.

The rules given in the appendix to Chapter 7 for manipulating and calculating the probability of events, particularly the combination rule, prove useful in our calculations of payoffs when individual players are differently informed. In games of asymmetric information, players try to find out the other's information by observing their actions. Then they must draw inferences about the likelihood of—estimate the probabilities of—the underlying information by exploiting the actions or consequences that are observed.

The best way to understand this is by example. Suppose 1% of the population has a genetic defect that can cause a disease. A test that can identify this genetic defect has a 99% accuracy: when the defect is present, the test will fail to detect it 1% of the time, and the test will also falsely find a defect when none is present 1% of the time. We are interested in determining the probability that a person with a positive test result really has the defect. That is, we cannot directly observe the person's genetic defect (underlying condition), but we can observe the results of the test for that defect (consequences)—except that the test is not a perfect indicator of the defect. How certain can we be, given our observations, that the underlying condition does in fact exist?

We can do a simple numerical calculation to answer the question for our particular example. Consider a population of 10,000 persons in which 100 (1%) have the defect and 9,900 do not. Suppose they all take the test. Of the 100 persons with the defect, the test will be (correctly) positive for 99. Of the 9,900

[35] See R. Duncan Luce and Howard Raiffa, *Games and Decisions* (New York: John Wiley & Sons, 1957), ch. 2 and app. 1, for an exposition; and Mark Machina, "Choice Under Uncertainty: Problems Solved and Unsolved," *Journal of Economic Perspectives*, vol. 1, no. 1 (Summer 1987), pp. 121–54, for a critique and alternatives. Although decision theory based on these alternatives has made considerable progress, it has not yet influenced game theory to any significant extent.

without the defect, it will be (wrongly) positive for 99. That is 198 positive test results of which one-half are right and one-half are wrong. If a random person receives a positive test result, it is just as likely to be because the test is indeed right as because the test is wrong, so the risk that the defect is truly present for a person with a positive result is only 50%. (That is why tests for rare conditions must be designed to have especially low error rates of generating "false positives.")

For general questions of this type, we use an algebraic formula called **Bayes' theorem** to help us set up the problem and do the calculations. To do so, we generalize our example, allowing for two alternative underlying conditions, $A$ and $B$ (genetic defect or not, for example), and two observable consequences, $X$ and $Y$ (positive or negative test result, for example). Suppose that, in the absence of any information (over the whole population), the probability that $A$ exists is $p$, so the probability that $B$ exists is $(1 - p)$. When $A$ exists, the chance of observing $X$ is $a$, so the chance of observing $Y$ is $(1 - a)$. (To use the language that we developed in the appendix to Chapter 7, $a$ is the probability of $X$ conditional on $A$, and $(1 - a)$ is the probability of $Y$ conditional on $A$.) Similarly, when $B$ exists, the chance of observing $X$ is $b$, so the chance of observing $Y$ is $(1 - b)$.

This description shows us that four alternative combinations of events could arise: (1) $A$ exists and $X$ is observed, (2) $A$ exists and $Y$ is observed, (3) $B$ exists and $X$ is observed, and (4) $B$ exists and $Y$ is observed. Using the modified multiplication rule, we find the probabilities of the four combinations to be, respectively, $pa$, $p(1 - a)$, $(1 - p)b$, and $(1 - p)(1 - b)$.

Now suppose that $X$ is observed: a person has the test for the genetic defect and gets a positive result. Then we restrict our attention to a subset of the four preceding possibilities—namely, the first and third, both of which include the observation of $X$. These two possibilities have a total probability of $pa + (1 - p)b$; this is the probability that $X$ is observed. Within this subset of outcomes in which $X$ is observed, the probability that $A$ *also* exists is just $pa$, as we have already seen. So we know how likely we are to observe $X$ alone and how likely it is that both $X$ and $A$ exist.

But we are more interested in determining how likely it is that $A$ exists, given that we have observed $X$—that is, the probability that a person has the genetic defect, given that the test is positive. This calculation is the trickiest one. Using the modified multiplication rule, we know that the probability of both $A$ and $X$ happening equals the product of the probability that $X$ alone happens times the probability of $A$ conditional on $X$; it is this last probability that we are after. Using the formulas for "$A$ and $X$" and for "$X$ alone," which we just calculated, we get:

$$\text{Prob}(A \text{ and } X) = \text{Prob}(X \text{ alone}) \times \text{Prob}(A \text{ conditional on } X)$$

$$pa = [\,pa + (1 - p)b\,] \times \text{Prob}(A \text{ conditional on } X)$$

$$\text{Prob}(A \text{ conditional on } X = \frac{pa}{pa + (1 - p)b}.$$

This formula gives us an assessment of the probability that $A$ has occurred, given that we have observed $X$ (and have therefore conditioned everything on this fact). The outcome is known as *Bayes' theorem* (or rule or formula).

In our example of testing for the genetic defect, we had Prob($A$) $= p = 0.01$, Prob($X$ conditional on $A$) $= a = 0.99$, and Prob($X$ conditional on $B$) $= b = 0.01$. We can substitute these values into Bayes' formula to get

Probability defect exists given that test is positive = Prob($A$ conditional on $X$)

Probability defect exists given that test is positive = Prob($A$ conditional on $X$)

$$= \frac{(0.01)(0.99)}{(0.01)(0.99) + (1 - 0.01)(0.01)}$$

$$= \frac{0.0099}{0.0099 + 0.0099}$$

$$= 0.5$$

The probability algebra employing Bayes' rule confirms the arithmetical calculation that we used earlier, which was based on an enumeration of all of the possible cases. The advantage of the formula is that, once we have it, we can apply it mechanically; this saves us the lengthy and error-susceptible task of enumerating every possibility and determining each of the necessary probabilities.

We show Bayes' rule in Figure 8A.3 in tabular form, which may be easier to remember and to use than the preceding formula. The rows of the table show the alternative true conditions that might exist, for example, "genetic defect" and "no genetic defect." Here, we have just two, $A$ and $B$, but the method generalizes immediately to any number of possibilities. The columns show the observed events—for example, "test positive" and "test negative."

Each cell in the table shows the overall probability of that combination of the true condition and the observation; these are just the probabilities for the four alternative combinations listed above. The last column on the right shows the sum across the first two columns for each of the top two rows. This sum is

| | | OBSERVATION | | Sum of row |
|---|---|---|---|---|
| | | $X$ | $Y$ | |
| TRUE CONDITION | $A$ | $pa$ | $p(1-a)$ | $p$ |
| | $B$ | $(1-p)b$ | $(1-p)(1-b)$ | $1-p$ |
| Sum of column | | $pa + (1-p)b$ | $p(1-a) + (1-p)(1-b)$ | |

**FIGURE 8A.3** Bayes' Rule

the total probability of each true condition (so, for instance, $A$'s probability is $p$, as we have seen). The last row shows the sum of the first two rows in each column. This sum gives the probability that each observation occurs. For example, the entry in the last row of the $X$ column is the total probability that $X$ is observed, either when $A$ is the true condition (a true positive in our genetic test example) or when $B$ is the true condition (a false positive).

To find the probability of a particular condition, given a particular observation, then, Bayes' rule says that we should take the entry in the cell corresponding to the combination of that condition and that observation and divide it by the column sum in the last row for that observation. As an example, Prob ($B$ given $X$) = $(1 - p)b/[pa + (1 - p)b]$.

## SUMMARY

Judging consequences by taking expected monetary payoffs assumes *risk-neutral* behavior. *Risk aversion* can be allowed by using the *expected utility* approach, which requires the use of a *utility function*, which is a concave rescaling of monetary payoffs, and taking its probability-weighted average as the measure of expected payoff.

If players have asymmetric information in a game, they may try to infer probabilities of hidden underlying conditions from observing actions or the consequences of those actions. *Bayes' theorem* provides a formula for inferring such probabilities.

## KEY TERMS

**Bayes' theorem** (339)
**expected utility** (337)
**risk-averse** (335)
**risk-neutral** (335)
**utility function** (337)