21

Asymmetric Information

Learning Objectives

After reading this chapter, students should be able to:

- Understand how adverse selection impacts markets.
- Explain the concepts of signaling and screening.
- Understand the implications of competitive signaling and screening for resource allocation, and identify implications for government policy.
- Explain how moral hazard can impact a trading relationship.
- Describe how an incentive scheme can provide a trading partner with incentives to take favorable actions, and some of the potential problems with providing incentives.

During the final quarter of 2001, a series of revelations concerning widespread accounting fraud sent Enron Corporation, a highly diversified energy conglomerate and once one of the most respected and successful companies in the United States, spiraling toward the largest bankruptcy in history. The value of the company's outstanding stock, valued at roughly $65 billion in August 2000, dwindled to practically nothing over the course of a few turbulent weeks (see Add-On 2A). Yet even as the crisis deepened, Enron's management still held out hope that it could save the company through a merger with Dynegy, another prominent energy conglomerate. According to reports in early November 2001, Dynegy was negotiating to purchase Enron for $7 to $8 billion in stock, and to provide an immediate cash infusion of $1.5 billion to alleviate the short-term crisis. As November progressed, disturbing facts concerning Enron continued to surface. On November 28, Dynegy walked away from the Enron merger. Enron immediately suspended all inessential payments and filed for bankruptcy protection four days later.

Why did the merger between Enron and Dynegy fall apart? As Dynegy learned more about Enron's problems, why didn't it simply revise its offer downward to reflect Enron's lower value? With unfolding events eroding the credibility of Enron's management, Dynegy became increasingly concerned about what it didn't know, and what Enron might still be hiding. Recognizing that Enron had practically no incentive to be forthcoming in its desperation to consummate the deal, Dynegy was forced to assume the worst. The deal unraveled largely because of Dynegy's informational handicap.

The failed merger negotiation between Dynegy and Enron illustrates the problems that can arise when one party to a potential transaction is less well-informed than another. This chapter is devoted to the study of such *informational asymmetries.* It covers the following four topics.

1. *Adverse selection.* When one party to a transaction has more information than another, the informed party may be more willing to trade precisely when trading is less advantageous to the uninformed party. As a result, the uninformed party may be reluctant to trade. We'll see that this reluctance can cause markets to perform poorly.
2. *Signaling.* Informed individuals often undertake costly activities to convince others of particular facts. We'll explain why such actions can convey information, explore their effect on the efficiency of resource allocation, and discuss implications for the role of government.
3. *Screening.* Faced with an informational handicap, an uninformed party may establish a test that induces informed parties to self-select, thereby revealing what they know. We'll explain how those tests work and how the government can in principle improve upon the efficiency, fairness, and stability of screening in free markets.
4. *Incentives and moral hazard.* In many circumstances, the attributes of a good or service depend on unobservable actions taken by one or more of the trading parties. We'll explore the ways in which an uninformed party can provide a trading partner with incentives to take favorable actions.

21.1 ADVERSE SELECTION

Throughout most of this book, we've assumed that all parties to a transaction share the same information: no seller knows more about a product's characteristics than a buyer, and no buyer knows more about the product's costs than a seller.[1] That assumption is often unrealistic. If you've ever considered buying a used car, you've probably wondered about its quality. After all, the seller of the car may secretly have a good reason to get rid of it. A similar plight confronts an insurance company when it sells policies that protect against risks such as death or disability. An applicant may want insurance precisely because he knows—and the insurance company does not—that he faces unusually high risks. Evaluating the applicant's riskiness appropriately can mean the difference between a profitable sale and a substantial loss. Similarly, workers often know more about their own abilities than do prospective employers. When a firm contemplates hiring a new employee, it needs to consider the possibility that the applicant is exaggerating his skills and qualifications. Hiring and training an employee who turns out to be a poor performer can be a costly mistake.

[1]There are some exceptions; see Sections 12.5 and 18.3.

As these examples suggest, one party to a transaction often has more information than another about the characteristics of the good or service to be traded, a situation known as an **informational asymmetry**. The seller of a used car may know more about her car's quality than prospective buyers, an applicant for an insurance policy may understand her health or disability risk better than the companies that offer such policies, and a worker may be more familiar with her own abilities than potential employers. Such asymmetries can exacerbate the inefficiencies that sometimes arise when information is imperfect. If buyers can't distinguish good cars from bad ones, sellers may be inclined to unload lemons (cars of very low quality). If insurance companies have difficulty evaluating applicants' risks, they may find themselves disproportionately serving high-risk policyholders, who have greater incentives to purchase insurance at any given price. And if employers have trouble assessing the abilities of job applicants, they may find themselves with many poorly qualified workers, who are more willing to accept any given job offer. In each of these examples, the informed parties (used-car sellers, insurance buyers, workers) are more willing to trade when trading is less advantageous to the uninformed parties (used-car buyers, insurance companies, employers). That tendency is known as **adverse selection**. When uninformed parties realize that they face adverse selection, they may become reluctant to trade, causing markets to perform poorly.

An **informational asymmetry** is present when one party to a transaction has more information than another about the characteristics of the good or service to be traded.

Adverse selection is present if an informed individual is more willing to trade when trading is less advantageous to an uninformed trading partner.

To see an extreme example of how adverse selection can make markets fail, suppose an acquaintance mentions that he has just looked up the dollar-to-Euro exchange rate; then he proposes trading you $70 for 50 Euro. Would you accept this proposal? Clearly not. You can come out ahead only if your acquaintance comes out behind. But why would he knowingly make an offer that is contrary to his interests? With this extreme form of adverse selection—the informed party is willing to trade only when the uninformed party loses—no trade can take place. (In a similar vein, recall also the discussion of sports betting in Application 11.2.)

Adverse Selection and Lemons

The problems caused by adverse selection were first studied by economist George Akerlof, who shared the Nobel Prize in 2001 for his pioneering contributions. Akerlof showed how adverse selection can undermine the possibilities for trade in a used-car market.

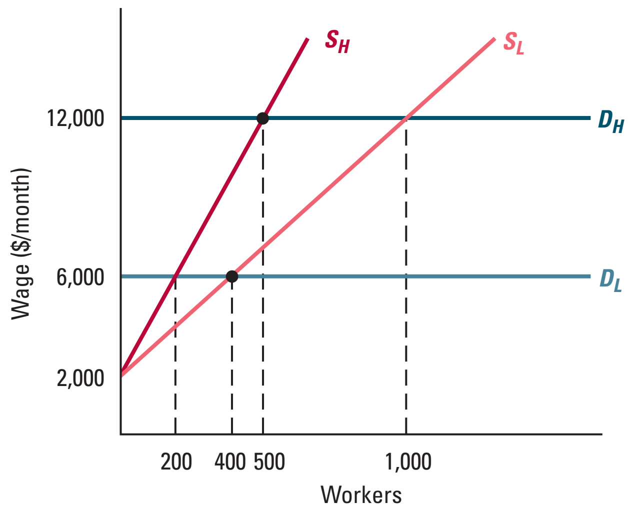
Because sellers want to sell lemons and keep good cars, buyers of used cars must be wary of quality. This consideration drives the price of a used car down and reduces the number of good cars owners are willing to sell. In some cases, adverse selection can drive good cars from the market completely. For example, suppose all owners of lemons are willing to sell their cars regardless of the price, but owners of good cars will become sellers only if the price of used cars is high enough. In that case, a reduction in price will raise the fraction of lemons among available used cars. This effect makes buyers even more wary of used cars and less willing to pay for them, which will drive the price down further. When the price falls, however, the fraction of lemons among the used-car supply will grow even larger. As a result of this vicious cycle, the price of a used car may be so low that no good cars are offered for sale.

George Akerlof, (1940–), who shared the 2001 Nobel Prize in Economics, showed how adverse selection can undermine the possibilities for trade.

Similar problems arise in many other contexts. In the rest of this section, we'll analyze the effects of adverse selection in a labor market. We'll assume that employers have difficulty assessing the abilities of job applicants, and that a less capable worker is more likely to accept a given job offer. In that case, low-ability workers can drive high-ability workers from the market, either partially or completely.

Figure 21.1

Demand and Supply for Software Programmers. The figure shows the demand and supply curves for high- and low-ability workers when ability is perfectly observable. In the competitive equilibrium, employers hire 400 low-ability workers at a wage of $6,000 per month, and 500 high-ability workers at a wage of $12,000 per month.



Adverse Selection in a Labor Market

Consider the labor market for entry-level software programmers in Palo Alto, California, in a given month. Suppose that each worker has either high or low ability. A high-ability worker generates $12,000 of profit per month, while a low-ability worker generates only $6,000 (in each case ignoring the worker's compensation). Figure 21.1 shows the demand curves for high- and low-ability workers when workers' abilities are observable to employers, labeled D_H and D_L respectively. The figure also shows the supply curves for high- and low-ability workers. The curve S_H reflects the supply decisions of high-ability workers. The higher the wage, the more high-ability workers are willing to accept employment.[2] The curve S_L reflects the supply decisions of low-ability workers.

When a worker's ability is perfectly observable, high- and low-ability workers receive different wages. In a perfectly competitive labor market, the wages of high- and low-ability workers adjust so that the quantity demanded equals the quantity supplied of each type. In the figure, employers hire 400 low-ability workers at a monthly wage of $6,000 and 500 high-ability workers at a monthly wage of $12,000.

Now suppose that firms can't tell whether a worker has high or low ability. This lack of knowledge poses a problem for each employer. How much should the firm be willing to pay a worker? To answer this question, the employer needs to determine the likelihood that a worker has high ability. If workers' abilities were perfectly observed, 400 out of the 900 workers hired would have low ability. But when employers can't discern a worker's ability, the likelihood of hiring a low-ability worker is greater. Figure 21.2 shows why. If employers cannot distinguish between high- and low-ability workers, then they must pay all workers the same wage, which will be between the values of the high- and low-ability workers. Employers will pay high-ability workers less than their value and low-ability

[2] Each worker may have other income-earning opportunities, such as starting his own firm or working in another city. In addition, some individuals may be more averse to work than others. Those with better alternatives and/or greater aversion to work will be less willing to accept employment at any given wage.

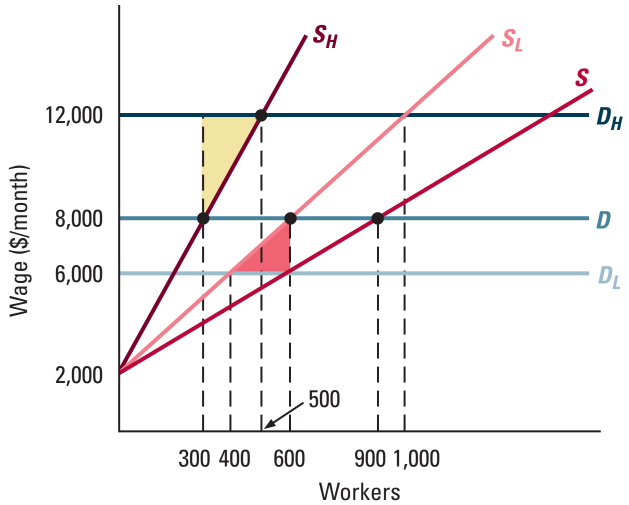


Figure 21.2

Demand and Supply for Software Programmers When Ability Is Unknown to Employers. The figure shows the market equilibrium when employers cannot observe workers' abilities. At each wage above $2,000 per month, two-thirds of the available workers have low ability. As a result, an employer is willing to pay workers $8,000 per month, leading to the demand curve labeled *D*. The aggregate labor supply curve is *S*, the horizontal sum of the supply curves for high-ability and low-ability workers, S_H and S_L respectively. In the equilibrium, employers hire 900 workers at a wage of $8,000 per month. Three hundred of those workers have high ability and 600 have low ability. The deadweight loss from asymmetric information is the sum of the yellow-shaded triangle (the loss from hiring too few high-ability workers) and the red-shaded triangle (the loss from hiring too many low-ability workers).

workers more than their value. As a result, low-ability workers will be more likely and high-ability workers less likely to accept employment than if employers could observe their ability; thus, a larger fraction of the available labor supply will have low ability.

In Figure 21.2, at every wage above $2,000, the number of low-ability workers willing to accept employment is exactly twice the number of available high-ability workers. For example, when the wage is $8,000 per month, 600 low-ability workers and 300 high-ability workers are willing to accept employment. When the wage is $12,000, 1,000 low-ability and 500 high-ability workers are willing to accept employment. So regardless of the wage rate, an employer should understand that he has a one-in-three chance of hiring a worker with high ability and a two-in-three chance of hiring one with low ability.

Figure 21.2 also shows the market equilibrium when employers can't observe a worker's ability. Because two-thirds of the available workers at any given wage have low ability, an employer should be willing to pay a worker $8,000 per month.[3] The resulting demand curve is labeled *D*. The curve labeled *S* is the market supply curve of workers, the horizontal sum of the high- and low-ability supply curves, S_H and S_L. The equilibrium wage equates market demand and supply, which occurs where the curves *D* and *S* intersect. The equilibrium wage is $8,000 per month and employers hire 900 workers. Of those, 600 have low ability and 300 have high ability.

Finally, Figure 21.2 shows the deadweight losses due to asymmetric information. We know that the competitive equilibrium with perfect information, shown in Figure 21.1, maximizes aggregate surplus. With asymmetric information, employers hire 300 high-ability workers instead of 500, and 600 low-ability workers instead of 400. The yellow-shaded triangle in Figure 21.2 is the deadweight loss from hiring too few high-ability

[3]We assume here that employers are risk neutral. The expected productivity of a randomly selected worker is $(^2/_3 \times 6{,}000) + (^1/_3 \times 12{,}000) = \$8{,}000$. If employers are risk averse, they will be willing to pay less than $8,000. (See Section 11.2 for a discussion of risk aversion.)

workers, and the red-shaded triangle is the deadweight loss from hiring too many low-ability workers.

Worked-out problem 21.1 shows how to derive the market equilibrium and deadweight loss with adverse selection using algebra.

WORKED-OUT PROBLEM 21.1

The Problem Each entry-level software programmer in Palo Alto, California, has either high or low ability. All potential employers value a high-ability worker at \$12,000 per month and a low-ability worker at \$6,000. The supply of high-ability workers is $Q_H^s = 0.05(W - 2{,}000)$ and the supply of low-ability workers is $Q_L^s = 0.1(W - 2{,}000)$, where W is the monthly wage. (These are the supply functions that lead to the supply curves in Figures 21.1 and 21.2.) If workers' abilities are observable to employers, what are the equilibrium wages? How many workers of each type do employers hire? If workers' abilities are not observed by employers, what is the equilibrium wage? How many workers of each type do employers hire? What is the deadweight loss due to asymmetric information?

The Solution When workers' abilities are observable, the wage for a high-ability worker must equal \$12,000, his value to employers. We can calculate the number hired from the supply function: $Q_H = 0.05(12{,}000 - 2{,}000) = 500$. In a similar fashion, we find that the wage of a low-ability worker is \$6,000 and the number hired is $Q_L = 0.1(6{,}000 - 2{,}000) = 400$.

When workers' abilities are not observed by employers, there is a single wage, W. The aggregate labor supply function is

$$\begin{aligned} Q^s &= Q_H^s + Q_L^s \\ &= 0.005(W - 2{,}000) + 0.01(W - 2{,}000) \\ &= 0.015(W - 2{,}000) \end{aligned}$$

An employer's willingness to pay depends on the fraction of available workers who have high-ability, F_H. That fraction equals

$$F_H = \frac{Q_H^s}{Q_H^s + Q_L^s} = \frac{0.1(W - 2{,}000)}{0.05(W - 2{,}000) + 0.01(W - 2{,}000)} = \frac{0.05}{0.15} = \frac{1}{3}$$

Notice that this fraction is the same at every possible wage. So the expected value of a job applicant to an employer is $[(^1/_3)(\$12{,}000) + (^2/_3)(\$6{,}000)] = \$8{,}000$ regardless of the wage. The equilibrium wage is therefore \$8,000. We calculate the total number of workers hired from the supply function: $Q^s = 0.15(8{,}000 - 2{,}000) = 900$. The number of high-ability workers hired is one-third of this total (300); the rest (600) have low ability.

Employers hire too many low-ability workers (600 instead of 400) and too few high-ability workers (300 instead of 500). The deadweight loss from hiring too many low-ability workers is the red-shaded area between the demand and supply curves D_L and S_L in Figure 21.2, which equals \$200,000 per month. The deadweight loss from hiring too few high-ability workers is the yellow-shaded area between the demand and supply curves D_H and S_H in Figure 21.2, which equals \$400,000 per month. So the total deadweight loss is \$600,000 per month.

IN-TEXT EXERCISE 21.1 **Repeat worked-out problem 21.1, but assume that the supply function of low-ability workers is $Q_L^s = 0.015(W - 2{,}000)$.**

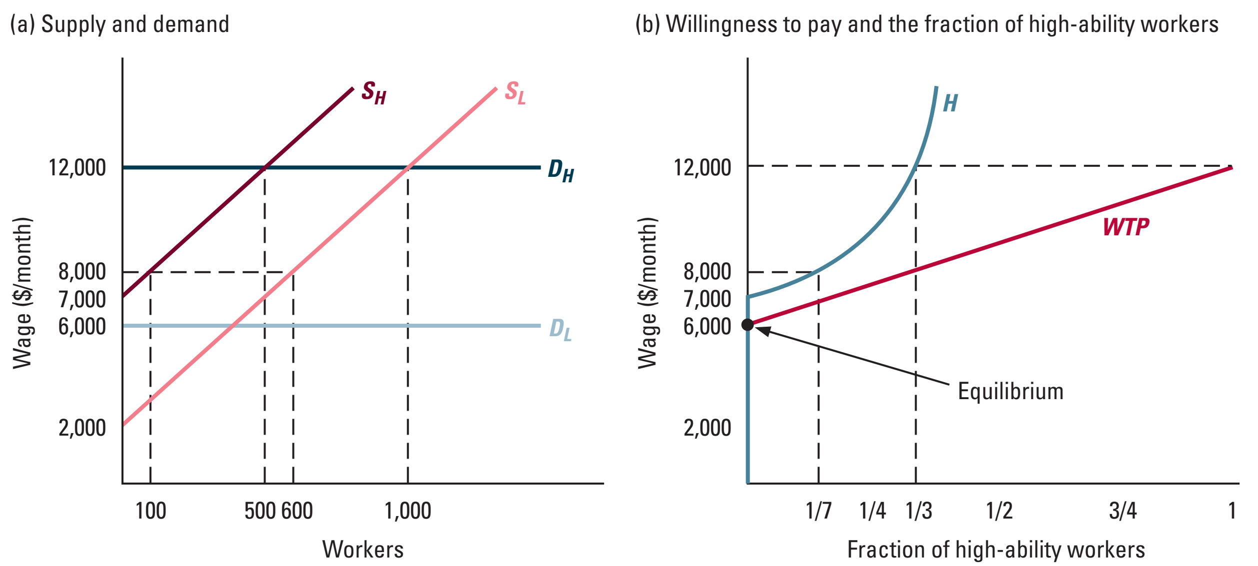
Market Unraveling In Figure 21.2, the fraction of workers willing to accept employment who have high ability is the same at every wage. Figure 21.3(a) shows a case in which that fraction is larger at higher wages. For example, if the wage is $12,000 per month, 1,000 low-ability workers and 500 high-ability workers are willing to accept employment, so one-third of the labor force has high ability, just as in Figure 21.2. But when the wage is $8,000 per month, 600 low-ability workers and 100 high-ability workers are willing to accept employment, so only one-seventh of the labor force has high ability. And if the wage falls below $7,000, *only* low-ability workers are willing to accept employment. In such cases, the presence of low-ability workers can chase high-ability workers out of the market entirely. This unfortunate outcome emerges even though employers would hire both high- and low-ability workers if information were perfect. This phenomenon is an example of **market unraveling**, which occurs in settings with adverse selection when the presence of unattractive trading partners drives attractive trading partners out of the market by altering the prices at which they can trade.

Market unraveling occurs in settings with adverse selection when the presence of unattractive trading partners drives attractive trading partners out of the market by altering the prices at which they can trade.

Why might a labor market unravel? Look again at Figure 21.3(a). We'll show that there is no wage above $7,000 (the lowest wage at which high-ability workers are willing to accept employment) at which the demand for labor equals the supply. Suppose the wage is $12,000. Then, as we've seen, two-thirds of the workers willing to accept employment have low ability. The presence of these low-ability workers means that an employer won't

Figure 21.3

Market Unraveling Due to Adverse Selection. In figure (a), the fraction of available workers who have high ability is greater the higher the wage. Figure (b) shows that there is no wage at which both high- and low-ability workers are hired; low-ability workers chase high-ability workers out of the market entirely. The equilibrium involves a wage of $6,000 per month.



be willing to pay more than \$8,000. But if the wage is \$8,000 or less, at least six-sevenths of the workers willing to accept employment would have low ability, so an employer would be unwilling to pay more than \$6,857 [since $\$6,857 = (^1/_7 \times 12,000) + (^6/_7 \times 6,000)$, rounded to the nearest dollar]. But with a wage less than \$7,000, *no high-ability workers will seek employment*! The equilibrium wage will be \$6,000 and employers will hire 400 low-ability workers. Although perfectly informed employers would hire 500 high-ability workers, asymmetric information drives all of those workers out of the labor market.

Figure 21.3(b) illustrates another approach to finding the market equilibrium. The vertical axis measures the wage and the horizontal axis measures the fraction of available workers who have high ability. The height of the curve labeled *WTP* shows, for each such fraction F_H, an employer's willingness to pay for workers: $WTP = (F_H \times 12,000) + [(1 - F_H) \times 6,000]$. The curve *H* shows, for each possible wage, the fraction of total labor that would be supplied by high-ability workers. That fraction is given by the formula:

$$F_H = \frac{S_H(W)}{S_L(W) + S_H(W)}$$

where $S_H(W)$ and $S_L(W)$ are the supplies of high- and low-ability workers, respectively, at a wage of *W*. An equilibrium occurs at the intersection of these two curves. In such an equilibrium the market wage *W* equals the average productivity of those workers seeking employment given that wage; employers are therefore willing to hire all of the workers who seek employment. In Figure 21.3(b), the curves intersect where the fraction of high-ability workers is zero: therefore, employers hire no high-ability workers.[4]

Responses to Informational Asymmetries

When asymmetric information leads to market failures, governments and private organizations often respond in ways that reduce the potential economic losses. Sometimes the government mandates minimum quality standards, which reduce the asymmetry of information. Many cities, for example, require restaurants to meet standards for cleanliness. Product liability laws also help reduce the effects of asymmetric information. These laws require manufacturers of defective products to compensate buyers for certain types of losses. By imposing prohibitive costs on very low quality firms, they reassure consumers that the available products meet minimum quality standards. Private organizations, such as Consumer Reports, also serve as quality certifiers. Because poorly informed parties are often willing to pay for better information, many such organizations see asymmetric information as a profit opportunity.

Even when governments and other organizations provide no remedies for asymmetric information, market participants themselves may respond in ways that reduce the potential economic losses. Experienced buyers may share information with those who are considering a firm's product. On eBay, for example, past purchasers can post reviews of a seller's performance. Poorly informed market participants may also gather information about prospective trading partners, usually at a cost. For example, a life insurance company may require applicants to have medical examinations, and a potential buyer of a used car may insist on bringing it to a mechanic for an inspection.

[4]The fact that the fraction of high-ability workers is larger at higher wages does not necessarily imply that the market will completely unravel. For an example, see exercise 21.4 at the end of the chapter.

In Sections 21.2 and 21.3, we'll discuss two other ways in which market participants respond to asymmetric information. The first, called *signaling,* involves efforts by some of the informed parties to reveal their information. The second, called *screening,* involves the creation of a test, designed by an uninformed party, that induces informed parties to self-select, thereby revealing what they know.

Application 21.1

Choosing among Health Plans

Many jobs, particularly at larger firms, include some form of health insurance. Often, an employer offers each employee a choice from among several health plans. The employer covers a significant fraction of the costs of these plans, with the employee absorbing the rest. Typically, an employee pays more for plans that offer more extensive coverage and/or give the employee more freedom in choosing her doctor or hospital. When designing these menus of plans, employers and health insurance companies must pay careful attention to the possibility of adverse selection.

Economists Daniel Altman, David Cutler, and Richard Zeckhauser have studied the health plan choices of state and local government employees in Massachusetts.[5] These employees could select from a traditional indemnity plan that allowed them to freely choose their doctor and hospital, a preferred-provider (PPO) plan that partially restricted their choice, and a number of health maintenance organization (HMO) plans that required them to use particular doctors and hospitals. Altman, Cutler, and Zeckhauser observed these employees' choices as well as their actual medical expenses in fiscal years 1994 and 1995.

Employees' choices reflected the presence of adverse selection. For example, between 1994 and 1995, employees who switched from an HMO to the more generous indemnity plan, on average, had incurred $1,651 of medical expenses in 1994 compared to an average of only $1,125 for those who stayed in an HMO plan.[6] In contrast, employees who switched from the indemnity plan to an HMO plan had incurred an average of $1,444 in medical expenses in 1994 compared with expenses of $2,252 for those who stayed in the indemnity plan. Thus, there is strong evidence that employees who chose the indemnity plan had a higher likelihood of filing claims than those workers who chose HMO plans.

Economists who have studied individuals' choices among insurance plans have not always found evidence of adverse selection, however.[7] Why not? First, in some contexts, people may not know much about their true risk levels. (For example, bad drivers may not know they are bad drivers.) Second, people may differ in ways that cause those with lower risk exposure to demand relatively more insurance, rather than less. For example, people who are more risk averse buy more insurance (see Section 11.3). But, they may also engage in less risky activities and therefore have fewer accidents. Because differences in risk aversion can create a negative relationship between the demand for insurance and the risk of an accident, insurers may not experience adverse selection.

[5] Daniel Altman, David M. Cutler, and Richard J. Zeckhauser, "Adverse Selection and Adverse Retention," *American Economic Review* 88, May 1998, pp. 122–126.

[6] These figures are age- and gender-adjusted. That is, they tell us that among workers of the same age and gender, those who switched to the indemnity plan had higher expenses than those who did not switch.

[7] See, for example, James H. Cardon and Igal Hendel, "Asymmetric Information in the Health Insurance Market: Evidence from the NMES," *RAND Journal of Economics* 32, Autumn 2001, pp. 408–427 and Pierre-Andre Chiappori and Bernard Salanie, "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy* 108, February 2000, pp. 56–78.

21.2 SIGNALING

In 1888, Vincent van Gogh, the 19th century Dutch Post-Impressionist artist, famously cut off the lower portion of his left ear. Depending on which story one credits, his objective may have been to prove the depth of his love for a woman or to demonstrate remorse for attacking his friend and housemate, Paul Gauguin.[8] Van Gogh suffered from mental illness, and his self-mutilation was unquestionably excessive as a token of either love or remorse. Still, the *logic* of his act is familiar, if twisted. Talk is cheap; anyone can claim to be remorseful or in love. However, only a person who truly felt those emotions would be willing to cut off part of his ear to prove his sincerity.

Signaling occurs when an informed individual undertakes a costly activity to convince others of particular facts.

Van Gogh's tragic decision is an extreme example of a phenomenon known as signaling. **Signaling** occurs when an informed individual undertakes a costly activity to convince others of particular facts. A great deal of social and economic interaction involves some form of signaling. For instance, a love-struck man may not cut off his ear, but he may make other socially acceptable sacrifices, such as skipping poker night with his friends, at least in part to prove his devotion.

In many situations, signaling offers a partial solution to problems that arise from adverse selection. The market for used cars again serves as an excellent example. Many dealers sell used cars with warranties. The most obvious purpose of a warranty is to provide the buyer with some protection in the event the car turns out to be a lemon. But there is a second and equally important purpose: offering a warranty serves as a signal that the seller has a high quality car. Unlike the seller of a sound car, the seller of a lemon is likely to lose a great deal of money if he offers to fix the car free of charge. Therefore, sellers who believe their cars are sound can credibly convey that belief to buyers by providing warranties. Anyone can claim that a car is sound, but a warranty backs up the claim.

Many economists believe that educational achievement serves, at least in part, as a signal to potential employers of raw intellectual ability. With sufficient work, most people could finish college with good grades, or even earn higher degrees. Yet those with less academic talent are less likely to do so because they find school more difficult and less enjoyable. Therefore, students who are intellectually talented can credibly convey that fact to future employers by obtaining more education. Anyone can claim to be intellectually capable, but getting more education backs up the claim.

In the rest of this section, we'll explain in more detail how education can serve as a signal of ability. We'll see that greater education can lead to higher earnings even if it adds absolutely nothing to a worker's productivity. Employers may pay higher wages to more highly educated workers simply because those workers tend to be more capable to begin with.

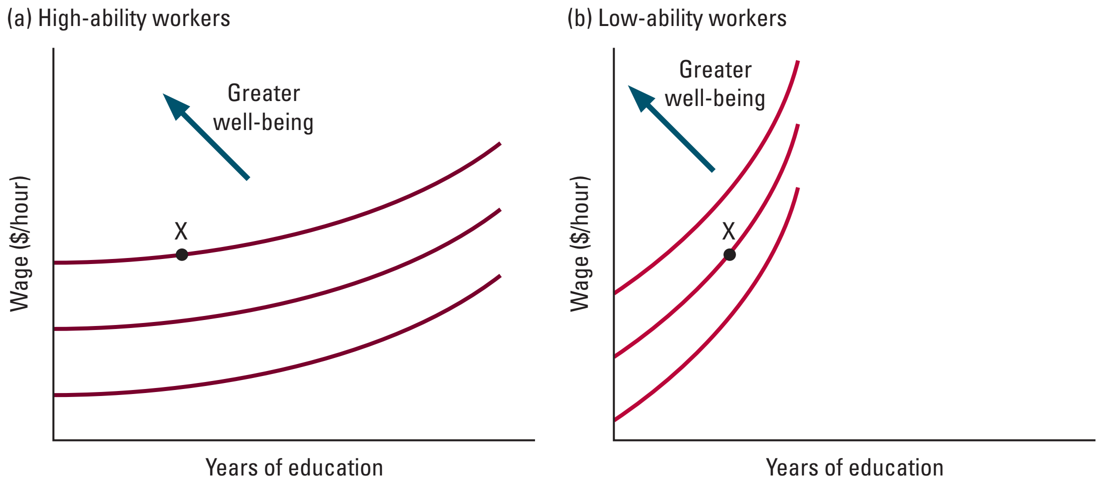
A Simple Model of Educational Attainment

To keep our analysis simple, we'll assume that there are two types of workers, those with high intellectual ability and those with low ability. Both types are equally numerous, so there's a 50 percent chance that a randomly chosen worker will have high ability. Each

[8]Historians have questioned both accounts. According to some reports, van Gogh wrapped the severed portion of his ear in a cloth, took it to a nearby bordello, presented it to a prostitute, and asked her to keep it for him. Alternatively, Gauguin may have cut off part of van Gogh's ear in a fight, and then blamed van Gogh.

Figure 21.4

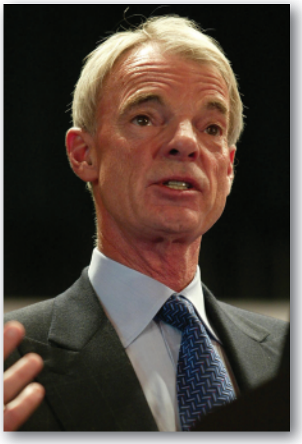
Indifference Curves for Education and Wages. Each worker cares about her wage and the amount of education she obtains. Fixing her education, she prefers a higher wage because it enables her to purchase more goods. Fixing her wage, she prefers less education because the costs of time and effort outweigh the pleasure of learning. Therefore, shifting her bundle to the northwest (in the direction of the blue arrow) improves her well-being, and her indifference curves slope upward. In addition, at any given point (such as point X in figures (a) and (b)), the indifference curve of a high-ability worker is flatter than that of a low-ability worker.



worker cares about her wage and the amount of education she obtains. Fixing her education, she prefers a higher wage because it enables her to purchase more goods. Fixing her wage, she prefers less education because the costs of time and effort outweigh the pleasure of learning.[9]

We illustrate the preferences of high-ability workers in Figure 21.4(a) and low-ability workers in Figure 21.4(b) by drawing indifference curves. Three features of these figures deserve emphasis. First, because the worker likes a high wage and dislikes schooling, shifting her bundle to the northwest (in the direction of the blue arrow) improves her well-being. Second, the indifference curves slope upward: to compensate the worker for suffering through more years of schooling, we must increase her wage. Third, at any given point (such as point X in both figures), the indifference curve of the high-ability worker is flatter than that of the low-ability worker. In other words, the greater the worker's ability, the smaller is the wage increase required to compensate for an increase in education. That pattern reflects an assumption that education is both easier and more pleasant for those with greater academic talent.

[9]In reality, many people enjoy education. But as long as there is some level beyond which the costs of time and effort exceed the nonmonetary benefits of schooling, our analysis would be essentially unchanged.



Michael Spence (1943–), who shared the 2001 Nobel Prize in Economics, is credited with developing the theory of signaling.

We'll assume that employing an additional high-ability worker adds \$50 per hour to a firm's revenue regardless of her educational attainment or the number of workers employed, while employing an additional low-ability worker adds \$20 per hour. This assumption has two important components. First, high-ability workers are more productive than low-ability workers. Second, schooling has absolutely no effect on a worker's productivity. In this simple model, the purpose of education is simply to make students jump through hoops; it imparts no useful knowledge. We adopt this rather extreme assumption to highlight the role of education as a signal rather than a determinant of ability. (Perhaps the assumption will strike some readers as plausible. Obviously, *this* course is jammed with valuable pearls of wisdom, but the same may not be true of your *other* classes!)

Let's suppose that the labor market is competitive. Potential employers are numerous, and every employer is willing to pay each worker the value of her marginal product. If each worker's ability is known both to the worker and to potential employers, then the equilibrium wage rates will be \$50 per hour for high-ability workers and \$20 per hour for low-ability workers. (We made a similar point at the start of Section 21.1.) Moreover, because education is costly and unproductive, everyone will receive only the minimum amount of schooling required by law, which we'll take to be 10 years.

Now we turn to the more interesting case, in which each worker's ability is initially known to the worker but not to potential employers. The next two subsections examine two types of equilibria, *separating equilibria* and *pooling equilibria.* In a **separating equilibrium**, people with different information choose different alternatives. In a **pooling equilibrium**, people with different information choose the same alternative.

In a **separating equilibrium**, people with different information choose different alternatives. In a **pooling equilibrium**, people with different information choose the same alternative.

To keep our discussion relatively simple, we'll assume throughout that every worker is willing to accept a full-time job at any positive wage. Therefore, in contrast to Section 21.1, everyone will choose to work in equilibrium. If signaling were impossible, adverse selection would not cause this market to unravel. Later, we'll briefly explain how signaling can help to resolve the problems arising from adverse selection when labor force participation depends on the wage rate.

Equilibrium with Separation

In a separating equilibrium, high-ability workers choose one level of education, E_H, and low-ability workers choose a different level, $E_L \neq E_H$. Each employer understands this relationship between educational choices and ability; he assumes that a job applicant with E_H years of education has high ability, and that one with E_L years of education has low ability. Because employers can accurately infer a worker's ability from her educational achievement, competition between employers will guarantee that each worker is paid the value of her marginal product. Firms will pay \$50 per hour (the value of the marginal product created by a high-ability worker) to those with E_H years of schooling, and \$20 per hour (the value of the marginal product created by a low-ability worker) to those with E_L years of schooling.

Low-ability workers will obtain only the minimum amount of education required by law (10 years). To understand why, let's imagine that there's an equilibrium in which low-ability workers obtain more schooling than required (12 years instead of 10), and see what goes wrong with it. Though workers with low ability are expected to obtain 12 years of education, any particular worker could choose the minimum instead. What would employers, ignorant of this nonconformist's ability, be willing to pay her? An employer

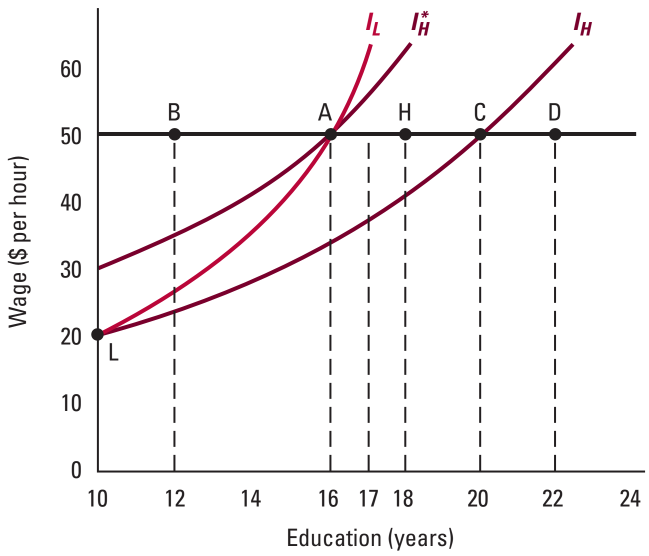


Figure 21.5

Equilibrium with Separation. In a separating equilibrium, low-ability workers end up at point L. They obtain the minimum amount of education required by law (10 years) and are paid the value of their marginal product ($20 per hour). High-ability workers end up at a point such as H, between points A and C. They obtain enough education to discourage imitation by low-ability workers (no less than 16 years), but not so much that they want to imitate those with low ability (no more than 20 years). They too are paid the value of their marginal product ($50 per hour).

can safely assume that the value of her marginal product is no less than $20 per hour; certainly she can't be less productive than a low-ability worker, and she might be more productive. Therefore, competition between employers guarantees that her wage will be no less than $20 per hour. Clearly, she would rather obtain 10 years of education and receive no less than $20 per hour, than obtain 12 years of education and receive exactly $20 per hour. That is why we cannot have a separating equilibrium in which low-ability workers attend school for 12 years.

Figure 21.5 illustrates everything we've learned so far about separating equilibria. Those who obtain $E_L = 10$ years of schooling receive $20 per hour (point L). Though we haven't yet determined the value of E_H, we know that it's possible to earn a wage of $50 per hour by obtaining E_H years of schooling. That outcome corresponds to some point on the horizontal black line.

To distinguish themselves from low-ability workers, high-ability workers must obtain more education than the minimum required by law. How much more? In a separating equilibrium, low-ability workers must not have an incentive to masquerade as high-ability workers by obtaining E_H years of schooling to receive $50 per hour, and high-ability workers must not have an incentive to masquerade as low-ability workers by obtaining E_L years of education, with the expectation that they would then receive $20 per hour.

In Figure 21.5, we illustrate the implications of these requirements by drawing two indifference curves through point L. The one labeled I_L belongs to a low-ability worker, and the one labeled I_H belongs to a high-ability worker. The indifference curve I_L intersects the black horizontal line at point A, which corresponds to 16 years of schooling and a wage of $50 per hour. It follows that E_H is no less than 16. To understand why, suppose for the moment that E_H , the level of education that will be chosen by high-ability workers, is 12, so that those who obtain 12 years of education are paid $50 per hour (point B). Because point B lies above the indifference curve I_L, low-ability workers prefer it to point L. In other words, if low-ability workers can earn $50 per hour by obtaining 12 years of education, they won't settle for $20 per hour and 10 years of education. Because they'll masquerade as high-ability workers, we don't have a separating equilibrium.

Similarly, the indifference curve I_H intersects the black horizontal line at point C, which corresponds to 20 years of schooling and a wage of $50 per hour. It follows that E_H is no greater than 20. To understand why, suppose for the moment that E_H , the level of education that will be chosen by high-ability workers, is 22, so that those who obtain 22 years of education are paid $50 per hour (point D). Because point D lies below the indifference curve I_H, high-ability workers don't like it as much as point L. In other words, if high-ability workers can earn $20 per hour by obtaining 10 years of education, they won't be willing to obtain 22 years of education, even for $50 per hour. Because they'll masquerade as low-ability workers, we don't have a separating equilibrium.

However, as long as E_H is between 16 and 20, we do have a separating equilibrium. To illustrate, let's suppose that E_H , the level of education that will be chosen by high-ability workers, is 18. In other words, workers who receive 18 years of education are paid $50 per hour. That bundle corresponds to point H in Figure 21.5. Notice that point H is below the indifference curve I_L and above the indifference curve I_H. Therefore, low-ability workers don't like it as much as point L, and high-ability workers like it better. If those with 10 years of education are paid $20 per hour (point L) while those with 18 years of education are paid $50 per hour (point H), low-ability workers will obtain 10 years of schooling while high-ability workers will obtain 18. That pattern confirms employers' expectations and justifies both wage rates.

Might workers choose some level of education other than 10 or 18 years? Not if employers believe that anyone receiving less than 18 years of schooling has low ability, in which case they will pay such workers $20 per hour—the same wage received by those with 10 years of education.[10] As long as all workers continue to choose either 10 or 18 years of education, no employer's experience will ever contradict those beliefs. Arguably, the beliefs may therefore persist as part of the equilibrium. (We will revisit this issue later in this section.)

Three features of separating equilibria deserve emphasis. First, though employers cannot observe a worker's ability directly, they can infer his ability from his schooling. Employers rely on workers to *self-select* into observably differentiated groups. (We previously encountered the concept of self-selection when studying price discrimination in Section 18.3.)

Second, for any worker, additional education leads to higher pay. Upon observing that pattern, a casual observer might be tempted to conclude that education enhances a worker's value to employers, presumably by imparting valuable knowledge that promotes productivity. That conclusion would be mistaken. In this model, schooling is assumed to have no beneficial effects whatsoever. Educational attainment is correlated with productivity but does not contribute to it.

Third, compared to the outcome with full information, separating equilibria are Pareto inefficient (see Section 16.3, page 588). High-ability workers would be better off if they obtained 10 years of schooling and received $50 per hour, and neither low-ability workers nor employers would be no worse off.

Even though all separating equilibria are inefficient, those with lower values of E_H are more efficient than those with higher values because they involve less wasteful education.

[10]Because a wage of $50 per hour reflects the highest possible level of productivity, no employer will be willing to pay more than $50 per hour to any worker, even if he obtains more than 18 years of education. Because workers can obtain $50 per hour by attending school for 18 years, they have no incentive to remain in school past that point.

In Figure 21.5, the most efficient separating equilibrium involves points L and A; those obtaining 10 years of education are paid \$20 per hour, and those obtaining 16 years of education are paid \$50 per hour. If the value of E_H were any lower, low-ability workers would masquerade as high-ability workers.

Equilibrium with Pooling

In a pooling equilibrium, all workers choose the same level of education, E_P, regardless of ability. Because an employer knows that a job applicant with E_P years of schooling is equally likely to have low ability and high ability, he is willing to pay her $(1/2 \times \$20) + (1/2 \times \$50) = \$35$ per hour.[11] Competition between employers guarantees that each worker who chooses E_P actually receives that wage.

In one type of pooling equilibrium, all workers obtain the minimum amount of education required by law ($E_P = 10$). If a worker attends school longer than required, employers continue to assume that she is equally likely to have low ability and high ability, and therefore remain willing to pay her \$35 per hour. Because school attendance is costly and doesn't lead to higher earnings, all workers are content with 10 years of schooling. Therefore, we have a pooling equilibrium.

Figure 21.6 illustrates another pooling equilibrium, in which all workers attend school for 12 years and end up at point P. If a worker receives 12 or more years of education, then (as above) employers assume she is equally likely to have low ability and high ability and are therefore willing to pay her \$35 per hour. Since extra education has costs and no benefits, no one is tempted to stay in school beyond 12 years. If a worker has less than 12 years of education, employers assume that she has low ability and are therefore willing

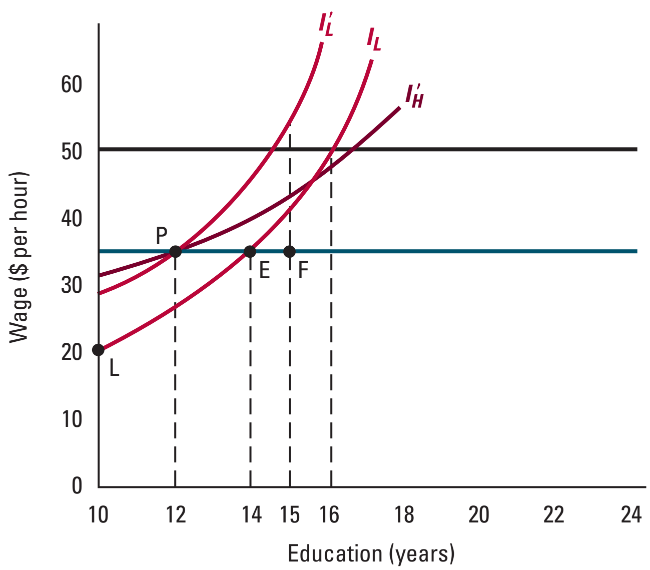


Figure 21.6

Equilibrium with Pooling. In a pooling equilibrium, high- and low-ability workers receive the same amount of education and are paid the expected value of the marginal product for a randomly selected worker (\$35 per hour). Workers must not obtain so much education that those with low ability would instead choose to obtain the minimum level of education and receive \$20 per hour. Therefore, workers end up at a point on the blue line, such as P, between the vertical axis and point E.

[11] We assume here that employers are risk neutral; see Section 11.2, p. 380.

to pay her only $20 per hour. Among those choices, 10 years of schooling is plainly the most tempting. When will neither type of worker prefer to obtain 10 years of education rather than 12? In the figure, we've drawn two indifference curves through point P. The one labeled I'_L belongs to a low-ability worker, and the one labeled I'_H belongs to a high-ability worker. Because both indifference curves intersect the vertical axis above point L, neither type of worker will be tempted to attend school for less than 12 years and settle for a wage of $20 per hour. Therefore, we have a pooling equilibrium.

Are there other pooling equilibria? Figure 21.6 also includes a horizontal blue line that intersects the vertical axis at $35, as well as the indifference curve I_L from Figure 21.5 (which passes through point L). The point at which the blue line intersects I_L, labeled E, corresponds to 14 years of education and a wage of $35 per hour. As long as E_P doesn't exceed 14 years, there is a pooling equilibrium just like the one described in the previous paragraph. However, no pooling equilibrium involves more than 14 years of schooling. Why not? Workers with low ability prefer point L to all points on the horizontal blue line to the right of point E. Take point F, which corresponds to 15 years of education and a wage of $35 per hour. As is clear from the figure, low-ability workers would rather attend school for 10 years, knowing that they can't possibly end up with a wage lower than $20 per hour. Therefore, there is no pooling equilibrium with 15 years of schooling.

Pooling equilibria with lower values of E_P are more efficient than those with higher values because they involve less wasteful education. The most efficient pooling equilibrium involves the minimum level of education required by law (10 years). That equilibrium is actually Pareto efficient in our simple model, taking the legal minimum as given.[12] With no wasteful education, it is impossible to make low-ability workers better off without hurting high-ability workers, and vice versa.

Comparing Separating Equilibria and Pooling Equilibria

In some cases, every worker would be better off with the most efficient pooling equilibrium than with the most efficient separating equilibrium. That's true, for example, in Figure 21.5. The figure includes an indifference curve for high-ability workers, labeled I^*_H, that runs through point A, the outcome chosen by high-ability workers in the most efficient separating equilibrium. That same indifference curve hits the vertical axis at $30. Therefore, a high-ability worker would be better off with 10 years of education and a wage of $35 per hour, than with 16 years of education and wage of $50 per hour. Obviously, a low-ability worker would be better off with a wage of $35 per hour than a wage of $20 per hour, assuming he has 10 years of education in either case. Therefore, both high-ability and low-ability workers are better off with the pooling equilibrium.

In fact, both types of workers always prefer the pooling equilibrium when more than one-third of workers have high ability, because in all such cases the productivity of the average worker exceeds $30 per hour (the vertical intercept of I^*_H). The superiority of the pooling equilibrium is most evident when only a very tiny fraction of workers, say one in a million, has low ability. In that case, the hourly wage in a pooling equilibrium will be only a fraction of a cent below $50. In contrast, in the most efficient separating equilibrium, the hourly wage of high-ability workers (who make up almost the entire popula-

[12] You should not conclude from this observation that the most efficient pooling equilibrium is always Pareto efficient. It need not be Pareto efficient if labor force participation depends on the wage rate, a case we discuss briefly below.

tion) will be only slightly higher, even though they will bear the costs of six additional years in school.

Which Equilibrium Will Prevail?

We have seen that our simple model of educational attainment has many separating equilibria and many pooling equilibria. Which one will prevail? Some economists think that the most efficient separating equilibrium is the only plausible competitive outcome, and that all other equilibria are unstable. In this section, we explain why that is so.

Consider the relatively inefficient separating equilibrium in which high-ability workers obtain 18 years of education (discussed above and illustrated in Figure 21.5). Suppose some high-ability worker obtains 17 years of education instead of 18. What would employers be willing to pay her? As we suggested previously, the answer depends on the employers' beliefs about her ability.

According to Figure 21.5, a low-ability worker would obtain 17 years of education rather than 10 years only if she expected to receive more than $50 per hour. Because no employee can reasonably hope for a wage that exceeds the value of the most productive worker, an employer can rule out the possibility that the nonconformist has low ability. In contrast, a high-ability worker would choose 17 years of education instead of 18 as long as she expected to receive a wage close to $50 per hour. Therefore, an employer cannot rule out the possibility that the nonconformist has high ability. Putting these inferences together, the employer concludes that the nonconformist must have high ability and is therefore worth $50 per hour.[13]

Even if an employer fails to reach that conclusion on his own, a high-ability worker with 17 years of schooling could argue as follows:

> I decided to obtain seventeen years of education and then convince you that I have high ability so that you would pay me $50 per hour. Here's why you should believe me. If I was a low-ability worker, it would have been silly for me to obtain seventeen years of education rather than ten, even if I had confidence that I could then trick you into paying me $50 per hour. However, as a high ability worker, I had every incentive to obtain seventeen years of education rather than eighteen, given my confidence that you would find this argument persuasive and pay me $50 per hour.

Because a high-ability worker can anticipate that his competitive wage will be $50 per hour if he gets 17 years of education, he has no reason to get 18 years. Sooner or later, some high-ability worker will be willing to take that risk, at which point the inefficient separating equilibrium will collapse.

For similar reasons, the pooling equilibria may also be unstable. Consider the equilibrium in which all workers obtain 12 years of schooling (discussed above and illustrated in Figure 21.6). Suppose some high-ability worker obtains 15 years of education instead of 12. Arguably, an employer should be willing to pay her $50 per hour. According to Figure 21.6, a low-ability worker would obtain 15 years of education rather than her equilibrium outcome only if she expected to receive more than $50 per hour, which is impossible. In

[13]This argument involves a criterion known as *equilibrium dominance,* which holds that a worker will not take an action if the outcome would definitely be worse (regardless of how employers respond) than his equilibrium outcome. The equilibrium dominance criterion is more controversial than the notion of dominance discussed in Section 12.2, p. 407.

contrast, a high-ability worker would choose 15 years of education over her equilibrium outcome as long as she expected to receive a wage close to $50 per hour, which is possible. Therefore, an employer should conclude that the nonconformist must have high ability and is therefore worth $50 per hour. Even if an employer fails to reach that conclusion on his own, a high-ability worker with 15 years of education might convince him of her ability through an argument similar to the one given above.[14] Sooner or later, some high-ability worker will be willing to take that risk, at which point the pooling equilibrium will collapse.

A Possible Role for the Government

Because signaling equilibria are often inefficient, it's natural to wonder whether some form of government intervention might improve the allocation of resources. If the government were perfectly informed about each worker's ability, it could easily resolve the problems arising from adverse selection. However, in practice, it is probably no better informed, and indeed may be less well informed, than employers. The question we need to address is whether there are beneficial public policies that require no more information than the government is likely to possess.

Suppose that competition tends to produce the most efficient separating equilibrium (as we argued in the last section), but that everyone would be better off with the most efficient pooling equilibrium. Then the government may be able to improve the allocation of resources by promoting the pooling equilibrium. In our simple model, it could compel pooling by banning education beyond 10 years. Because in reality education does impart useful knowledge, we shouldn't take that particular policy prescription seriously.

It also important to remember that, in some cases, signaling can help market participants overcome the problems associated with adverse selection, which we discussed in Section 21.1. Banning signaling can then be particularly counterproductive. Suppose, for example, that high-ability workers drop out of the labor force in disproportionate numbers as the wage rate declines. Without signaling, the market may unravel, driving away all high-ability workers. Because separating equilibria induce each worker to reveal her ability through her choices, they are immune to such problems. By preventing a market from unraveling either partially or completely, signaling can therefore enhance social efficiency.

Even though signaling can serve a socially beneficial purpose, taxes on signaling activities can be relatively efficient sources of government revenue. In our simple model, an education tax would reduce the amount of wasteful schooling required to separate high-ability workers from those with low ability, while also providing resources for public projects. (For a numerical example, work through exercise 21.8 at the end of the chapter.) Although there are good reasons to avoid taxing education (see, for example, Application 20.5, page 779), similar observations have potentially important implications for tax policy in other contexts. Consider the taxation of corporate dividends. Many economists believe that corporations pay shareholders regular cash dividends to signal high profitability. A company can convince investors of its profitability by paying dividends because an unprofitable company would find that activity too costly; for example, it may risk bankruptcy if it depletes its cash reserves. If dividends serve as costly signals of profitability, then the taxation of dividends may involve little or no deadweight loss.

[14]The argument is essentially the same, except that the worker would substitute "fifteen" for "seventeen," "twelve" for "ten," and "twelve" for "eighteen."

WORKED-OUT PROBLEM 21.2

The Problem The value of a worker's marginal product is \$40 per hour for high-ability workers and \$10 per hour for low-ability workers. The preferences of high-ability workers correspond to the utility function $U_H(E, W) = W - 5E$, and the preferences of low-ability workers correspond to the utility function $U_L(E, W) = W - 10E$. By law, everyone is required to attend at least ten years of school. Two-thirds of the population has low ability, and one-third has high ability. What can you say about the educational attainment of each type of worker in a separating equilibrium, and in a pooling equilibrium? Are workers better off with the most efficient separating equilibrium, or with the most efficient pooling equilibrium?

The Solution We'll start with separating equilibrium. For the same reasons given in the text, firms will pay \$40 per hour (the value of the marginal product created by a high-ability worker) to those with E_H years of schooling, and \$10 per hour (the value of the marginal product created by a low-ability worker) to those with E_L years of schooling. Also, low-ability workers will obtain only the minimum amount of education required by law ($E_L = 10$). Thus, low-ability workers end up at the point labeled L in Figure 21.7.

At point L, the utility of a high-ability worker is $U_H(10, 10) = 10 - (10 \times 5) = -40$, while the utility of a low-ability worker is $U_L(10, 10) = 10 - (10 \times 10) = -90$. We therefore obtain the following formulas for indifference curves through point L: $W = -40 + 5E$ for high-ability workers, and $W = -90 + 10E$ for low-ability workers. We've drawn those indifference curves in Figure 21.7. The one labeled I_H belongs to a high-ability worker; and the one labeled I_L belongs to a low-ability worker. We have also drawn a black horizontal line that intersects the vertical axis at \$40. High-ability workers must end up at a point on the thick segment of that

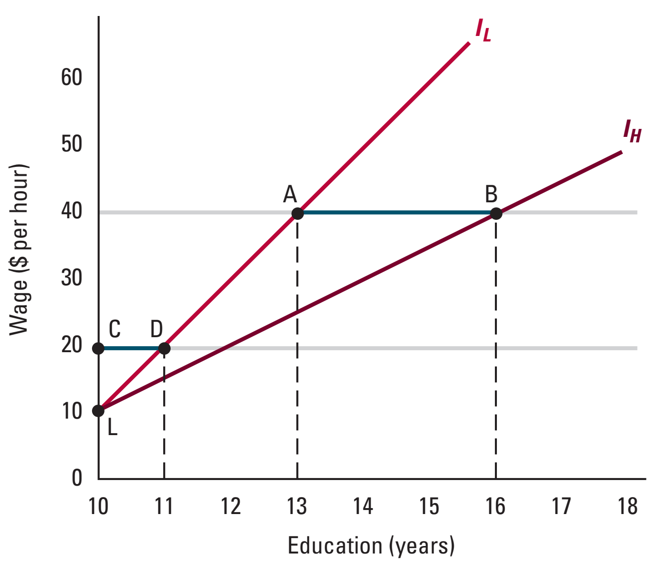


Figure 21.7:
Solution to Worked-Out Problem 21.2. In a separating equilibrium, low-ability workers end up at point L, and high-ability workers end up at a point on the thick segment of the black line between points A and B. In a pooling equilibrium, all workers end up at a point on the thick segment of the blue line between points C and D.

line between point A, where the black line intersects I_L, and B, where the black line intersects I_H. To calculate the horizontal coordinate of point A, we solve the equation $-90 + 10E = 40$, which implies $E = 13$. To calculate the horizontal coordinate of point B, we solve the equation $-40 + 5E = 40$, which implies $E = 16$. Therefore, E_H can be any value between 13 and 16.

Now we turn to pooling equilibria. The marginal product of an average worker is $(2/3 \times \$10) + (1/3 \times \$40) = \$20$ per hour. We have drawn a horizontal blue line in Figure 21.7 that intersects the vertical axis at \$20 (point C). The indifference curve labeled I_L intersects that line at point D. All workers must end up on the thick segment of blue line between points C and D. To calculate the horizontal coordinate of point D, we solve the equation $-90 + 10E = 20$, which implies $E = 11$. Therefore, E_P can be any value between 10 and 11.

In the most efficient separating equilibrium, low-ability workers end up at point L, and high-ability workers end up at point A. In the most efficient pooling equilibrium, all workers end up at point C. Clearly, low-ability workers prefer the pooling equilibrium (point C) to the separating equilibrium (point L). For high-ability workers, utility at point A is $U_H(13, 40) = 40 - (5 \times 13) = -25$, and utility at point C is $U_H(10, 20) = 20 - (5 \times 10) = -30$. Therefore, high-ability workers prefer the separating equilibrium (point A) to the pooling equilibrium (point C).

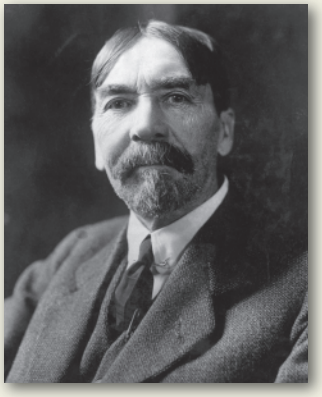
IN-TEXT EXERCISE 21.2 **The value of a worker's marginal product is \$55 per hour for high-ability workers and \$15 per hour for low-ability workers. The preferences of high-ability workers correspond to the utility function $U_H(E, W) = W - 4E$, and the preferences of low-ability workers correspond to the utility function $U_L(E, W) = W - 8E$. By law, everyone is required to attend at least 10 years of school. One-quarter of the population has low ability, and three-quarters have high ability. What can you say about the educational attainment of each type of worker in a separating equilibrium, and in a pooling equilibrium? Are workers better off with the most efficient separating equilibrium, or with the most efficient pooling equilibrium?**

Application 21.2

Conspicuous Consumption

In his celebrated late-19th century treatise, *A Theory of the Leisure Class,* the economist Thorstein Veblen argued that wealthy individuals often consume highly conspicuous goods and services to advertise their wealth, thereby achieving greater social status.[15] Veblen wrote that, "to gain and to hold the esteem of men, wealth must be put in evidence, for esteem is awarded only on evidence." By social custom, the evidence consists of unduly costly goods that fall into "accredited canons of conspicuous consumption, the effect of which is to hold the consumer up to a standard of expensiveness

[15]Thorstein Veblen. *The Theory of the Leisure Class: An Economic Study of Institutions.* London: Unwin Books, 1899; reprinted New York: Dover Publications, 1994.



Thorstein Veblen (1857–1929) argued that wealthy people purchase conspicuous luxury goods to advertise their wealth and thereby achieve greater social status.

and wastefulness in his consumption of goods and his employment of time and effort."

The details of Veblen's arguments naturally invite the interpretation that conspicuous consumption reflects signaling. In a signaling equilibrium, wealthy individuals would differentiate themselves from those with less money by consuming costly and conspicuous goods, an activity which Veblen dubbed "invidious comparison." This display of affluence would be convincing only if the costs incurred were large enough to discourage imitation—in Veblen's terms, "pecuniary emulation"—by those with less wealth. Signaling is possible because those with less wealth must make more painful sacrifices to finance conspicuously wasteful expenditures.

Today, many manufacturers of luxury goods acknowledge the validity of Veblen's observations. According to one executive at LVMH (a French conglomerate that owns Louis Vuitton, Moët et Chandon, and Christian Dior), for many individuals buying luxury goods "is all about demonstration."[16] Not surprisingly, marketing strategies often tout the status-enhancing effects of luxury products. For example, one Jaguar advertisement asks you to visualize yourself arriving at your high school reunion in a flashy convertible; "you can almost see the heads turn as your classmates ask, 'Isn't that. . .?'"

"It may be vulgar, but it shows who has the loot."

www.CartoonStock.com

Because effective signaling requires the consumption of expensive goods, luxury brand producers can often maintain prices well in excess of production costs, even when their goods are easily imitated. According to one marketing manager, "Our customers do not want to pay less. If we halved the price of all our products, we would double our sales for six months and then we would sell nothing."[17] Similarly, *The Wall Street Journal* has observed that "a BMW in every driveway might thrill investors in the short run but ultimately could dissipate the prestige that lures buyers to these luxury cars."[18] In some cases, companies sell two nearly identical versions of the same good for vastly different prices simply to ensure that the more expensive one carries greater prestige. For example, in 1993, the Rolls-Royce Silver Spur III sold for $25,800 more than the long-wheelbase Bentley Brooklands, even though the cars were virtual twins (apart from the style of the grille and headlights).

If the consumption of conspicuous luxury goods serves as a signal of wealth, then taxes on those goods may be relatively efficient sources of revenue. Indeed, a study by economist Laurie Hodrick and the first author of this book suggests that such taxes may involve little or no deadweight loss.[19] This observation is of particular interest in light of the Omnibus Budget Reconciliation Act of 1990, which for a time established substantial federal taxes on the sale of various conspicuous goods, including expensive automobiles, yachts, jewelry, and aircraft.

21.3 SCREENING

Asymmetric information often provides people with incentives to communicate with each other. In the previous section, we saw that informed parties can initiate communication by

[16]Quoted in "The Luxury Good Trade," *The Economist,* January 8, 1993, p. 97.

[17]Quoted in "The Luxury Good Trade," *The Economist,* January 8, 1993, p. 96.

[18]Appel, Timothy, "BMW, Despite Success, Is Acting Like It's Under Siege," *The Wall Street Journal,* February 19, 1992, p. B4.

[19]Laurie Simon Bagwell and B. Douglas Bernheim, "Veblen Effects in a Theory of Conspicuous Consumption," *American Economic Review* 86, June 1996, pp. 349–373.

signaling what they know. *Uninformed* parties can also initiate the transfer of information by testing either the informed parties or the goods those parties seek to trade.

In some cases, the informed party is at most a passive participant in the test. Consider once again the market for used cars. The potential buyer of a car can learn about its quality in various ways, such as researching its accident record or bringing it to a mechanic for a complete evaluation. Likewise, a life insurance company can learn about the health of an insurance applicant by obtaining the applicant's medical records, contacting his current physician, and subjecting him to a physical examination.

Screening occurs when an uninformed party establishes a test that induces informed parties to self-select, thereby revealing what they know.

In other cases, the informed party is an active participant in the test. Indeed, the test is designed to induce that party to self-select, thereby revealing what she knows. This process, known as **screening**, has much in common with signaling. To illustrate the similarities and differences between these approaches, let's compare Vincent van Gogh's ear-slashing with the conventions of courtly love that were practiced among the nobility of Europe and celebrated in the poetry of troubadours during the High Middle Ages. According to those conventions, a lady would bestow her love upon a knight only after he proved his love for her by performing one or more difficult tasks of *her* design. That test of love is an example of screening. If van Gogh's (potentially fictional) love interest had *asked* him to cut off his ear to prove his love, then like the knights of the Middle Ages, he would have been involved in screening, rather than signaling.

Screening is widely observed in real markets. If you've purchased automobile insurance, you know that insurance companies allow you to customize your policy in a number of ways. For example, you can choose the size of your deductible (the amount of any loss you pay before your insurance coverage begins). Because careful drivers are less likely to have accidents than reckless drivers, they may be more willing to consider high deductibles. Therefore, an insurance company can potentially learn much about a driver's riskiness from her choice of a deductible. In effect, the company sets a task for the insurance applicant: prove to us that you're a safe driver by accepting a high deductible, and we'll price your policy more attractively.

Screening also helps employers learn about workers' characteristics. An employer who wants to hire a certain type of worker for a particular position may design the position in a way that attracts applicants with the desired characteristics and repels others. For example, because large law firms benefit from hiring workaholics, many require new associates to put in extremely long hours. As a result, those firms attract workaholic applicants. Application 21.3 on page 21-31 elaborates on this example.

In the rest of this section, we'll explain in greater detail how employers can screen workers by designing jobs that appeal to people with particular characteristics. (Readers who are interested in a detailed discussion of screening in insurance markets should consult Add-On 21A.) We'll see that screening may not function well in competitive markets, and we'll discuss the implications for government policy.

A Simple Model of Workplace Responsibilities

To keep matters relatively simple, we'll assume as in Section 21.2 that there are two types of workers, those with high ability and those with low ability. In this case, each worker cares about her income and the number of tasks she performs at her job during a day. Fixing her tasks, she prefers a higher wage because it allows her to purchase more goods. Fixing her income, she prefers to perform fewer tasks because work is tiring.

We illustrate the preferences of high-ability workers in Figure 21.8(a) and low-ability workers in Figure 21.8(b) by drawing indifference curves. These figures resemble Figures 21.4(a) and (b) (page 21-11). Because each worker likes a high wage and dislikes performing tasks, her indifference curves slope upward, and shifting her bundle to the northwest (in the direction of the blue arrow) improves her well-being. We will assume that her indifference curves bow to the right, as shown.[20] If we also assume that more able workers find tasks less tiring, then at any given point (such as point X in the figure), the indifference curve of a high-ability worker will be flatter than that of the low-ability worker, as shown. (Can you explain why?)

We'll assume that high-ability workers perform tasks more effectively than low-ability workers. Employing an additional worker adds $10 in profit for each task if the worker has high ability, and $5 per task if the worker has low ability (in each case ignoring the worker's compensation). We'll also assume that the labor market is competitive; potential employers are numerous, and every employer is willing to pay each worker the value of her marginal product.

If each worker's ability were known both to the worker and to potential employers, then the equilibrium wage rates would be $10 per task for high-ability workers and $5 per task for low-ability workers. Employers could fix these rates and allow workers to choose their own tasks. Low-ability workers could then pick any point along the line labeled B_L in Figure 21.9 (the slope of which is $5 per task), and high-ability workers could pick any point on the line labeled B_H (the slope of which is $10 per task). To determine how many tasks each type of worker would choose, we look for points of tangency between these lines and the workers' indifference curves. According to the figure, high-ability workers would perform 50 tasks and earn $500 per day (point A), while low-ability workers would perform 25 tasks and earn $125 per day (point B). This outcome is efficient (see Chapters 14 and 16).

The problem becomes interesting when each worker knows her own ability, but her employer does not (unless the worker reveals it through her actions). We will assume that an employer can easily monitor the completion of tasks but cannot easily determine how well a task was performed. For example, when a worker repairs a machine, her employer can verify that the machine operates but may not know whether the repair was excessively costly or insufficiently durable. Therefore, employers must compensate workers based only on tasks performed rather than on either their ability or the quality of their performance (at least in the near term).

Asymmetric information prevents a competitive market from achieving the efficient outcome shown in Figure 21.9 (the one that would prevail with symmetric information between employers and workers). To understand why, suppose employers offer jobs that allow workers to earn $500 while performing 50 tasks per day (point A), as well as jobs that allow workers to earn $125 while performing 25 tasks per day (point B). As we've drawn the figure, both types of workers would choose the jobs associated with point A over those associated with point B, and the firms offering $500 for 50 tasks per day would lose money. This can't be a competitive equilibrium. What then does competition deliver, if not the efficient outcome? The next two subsections address that question. As in Section 21.2, our discussion will focus on separating equilibria and pooling equilibria.

Joseph Stiglitz (1943–), top, who shared the 2001 Nobel Prize in Economics, Michael Rothschild (1942–), middle, and Charles Wilson (1948–), bottom, are credited with developing the theory of competitive screening.

[20]If we think of the individual as choosing an income level and the absence of tasks (in other words, two goods rather than a good and a bad), then this assumption means that indifference curves have declining marginal rates of substitution.

Figure 21.8

Indifference Curves for Tasks and Income. Each worker cares about her income and the amount of tasks she performs. Fixing her tasks, she prefers higher income because it enables her to purchase more goods. Fixing her income, she prefers fewer tasks because work is tiring. Therefore, shifting her bundle to the northwest (in the direction of the blue arrow) improves her well-being, and her indifference curves slope upward. In addition, at any given point [such as point X in figures (a) and (b)], the indifference curve of the high-ability worker is flatter than that of the low-ability worker. Indifference curves also bow to the right, as shown.

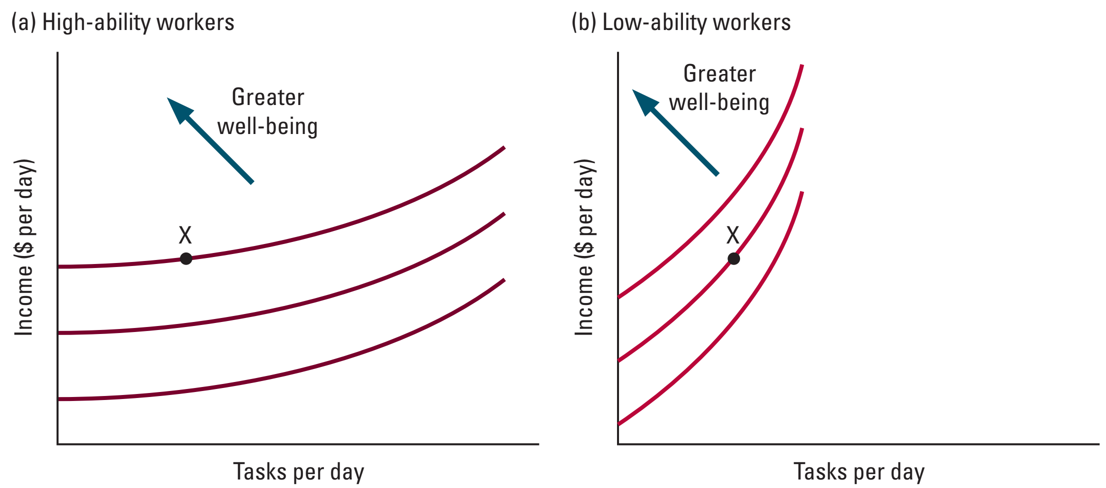
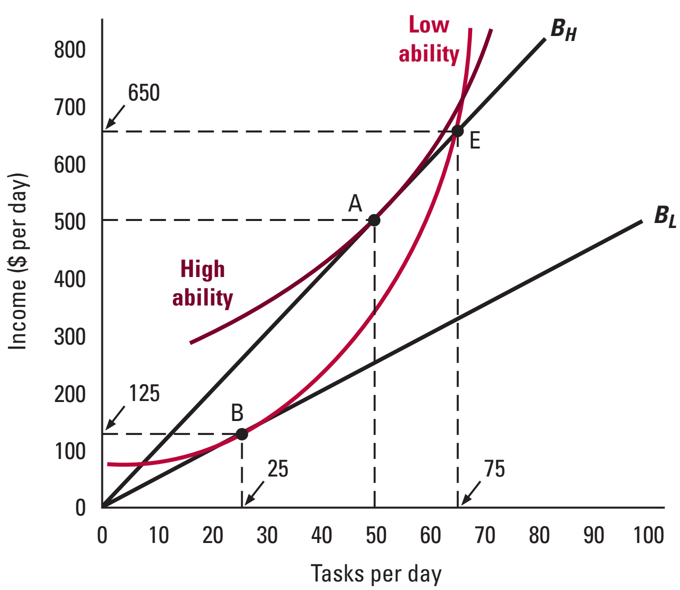


Figure 21.9

Income and Tasks When Employers Know Workers' Abilities. If each worker's ability were known both to the worker and to potential employers, then the equilibrium wage rates would be $10 per task for high-ability workers and $5 per task for low-ability workers. Employers could fix these rates and allow workers to choose their own tasks. High-ability workers would perform 50 tasks per day and earn $500 (point A), while low-ability workers would perform 25 tasks per day and earn $125 (point B).



Equilibrium with Separation

If an employer offers workers a choice between two different rates of compensation per task, one high and one low, and places no requirements on the number of tasks performed, all workers will obviously choose the higher rate of compensation. However, by assigning workplace responsibilities appropriately, an employer can induce different types of workers to sort themselves into different types of jobs. Figure 21.10 illustrates this point. Here we assume that the employer creates two types of jobs. One requires the employee to perform 60 tasks and pays $480 per day (a rate of $8 per task). The other requires the employee to perform 20 tasks and pays $120 per day (a rate of $6 per task). With these job offerings (and no other employers), employees must choose between points C and D. As shown, the workers self-select into different jobs: high-ability workers choose point C while low-ability workers choose point D. The low-ability workers settle for lower compensation because they are more averse to work.

In a separating equilibrium, workers of different abilities sort themselves into different types of jobs, much as in Figure 21.10. Employers offer one type of job requiring the completion of T_H tasks and paying $\$Y_H$ per day, and another requiring the completion of T_L tasks (where $T_L < T_H$) and paying $\$Y_L$ per day. They expect high-ability workers to self-select into jobs of the first type, and low-ability workers to self-select into jobs of the second type. Neither type of job can generate positive profits; otherwise, new employers would have an incentive to enter the labor market and create jobs of that type. Nor can either type of job generate negative profits; without an offsetting source of positive profits, any employer offering the unprofitable type of job would necessarily lose money and shut down. Therefore, the type of job selected by high-ability workers must pay $Y_H = 10T_H$, and the type of job selected by low-ability workers must pay $Y_L = 5T_L$. In other words, high-ability workers will end up on the line labeled B_H in Figure 21.10, and low-ability workers will end up on the line labeled B_L. (Thus, while points C and D in

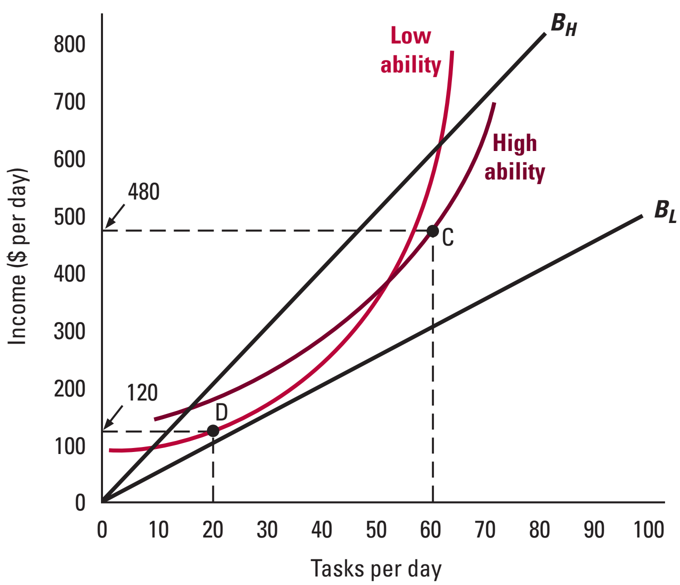


Figure 21.10

Voluntary Sorting by Workers. Suppose there are two types of jobs available, one that requires workers to complete 60 tasks and pays $480 per day (point C), and another that requires workers to complete 20 tasks and pays $120 per day (point D). Then high-ability workers will apply for the first type of job, and low-ability workers will apply for the second.

Figure 21.10 may induce high- and low-ability workers to make different choices, competitive firms will not offer the jobs associated with those points.)

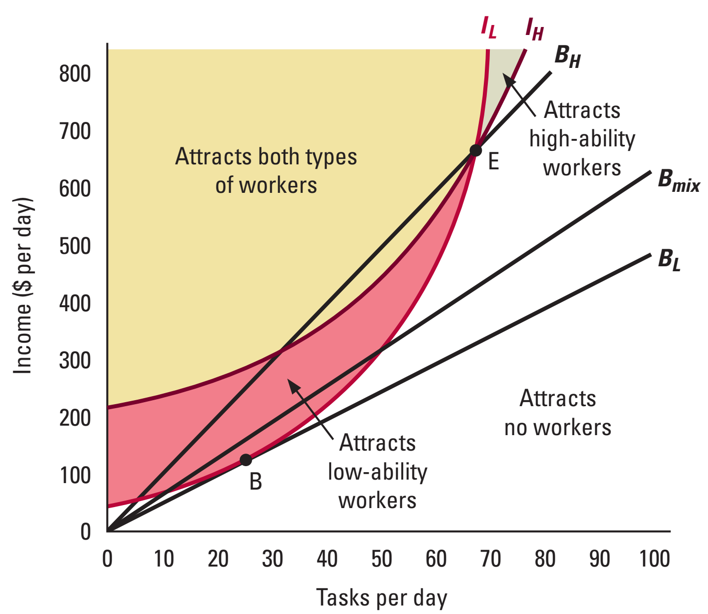
A Separating Equilibrium Look again at Figure 21.9. Point E lies at the intersection of the line B_H and the indifference curve labeled "low ability" that runs through point B. Under certain conditions (which we identify below), competition between employers leads to a separating equilibrium in which the terms of the job chosen by low-ability employees, T_L and Y_L, correspond to point B, and the terms of the job chosen by high ability employees, T_H and Y_H, correspond to point E. Low-ability workers are willing to choose point B even when point E is available. Because high-ability workers have flatter indifference curves than low-ability workers, they will choose point E over point B. Given those choices, both types of jobs generate zero profits for employers. To determine whether this combination of jobs survives against open competition, we need to assess whether a new employer can enter this labor market and earn positive profits by offering some other type of job. Sometimes it can, and sometimes it can't.

Figure 21.11 reproduces the lines B_L and B_H, the points B and E, and the indifference curve of a low-ability worker that runs through those points (now labeled I_L). We have added the indifference curve for a high-ability worker that runs through point E (labeled I_H). Let's evaluate the profit opportunities available to a new employer. Obviously, jobs associated with points B and E will attract workers, but generate zero profits. What other alternatives are available?

First consider points in the unshaded portion of the figure. Because those points are below both I_H and I_L, the corresponding jobs will not attract any workers. Therefore, they aren't profitable.

Figure 21.11

A Separating Equilibrium. In a separating equilibrium, high-ability workers end up at point E and low-ability workers end up at point B. Employers break even on both types of jobs and cannot earn positive profits by offering another type of job. Jobs corresponding to points in the unshaded area attract no workers. Jobs corresponding to points in the green-shaded area attract only high-ability workers but pay more than the value of their marginal product. Jobs corresponding to points in the red-shaded area attract only low-ability workers but pay more than the value of their marginal product. Jobs corresponding to points in the yellow-shaded area attract both types of workers but pay more than the expected value of a randomly selected worker's marginal product.



Next consider points in the green-shaded area. Because those points are above I_H and below I_L, the corresponding jobs will attract only high-ability workers. However, because those points also lie above B_H, the jobs would pay more than $10 per task, which means that the employer would lose money.

Next consider points in the red-shaded area. Because those points are above I_L and below I_H, the corresponding jobs will attract only low-ability workers. Because those points also lie above B_L, the jobs would pay more than $5 per task, which means that the employer would lose money.

Finally, consider points in the yellow-shaded area. Because those points are above both I_L and I_H, the corresponding jobs will attract all workers. Will they be profitable? Notice that we've added a new line to Figure 21.11, labeled B_{mix}. The slope of B_{mix} is the average value of a task performed by a randomly selected worker. For example, if high- and low-ability workers are equally numerous, then the slope of B_{mix} is $(^1/_2 \times \$5) + (^1/_2 \times \$10) = \$7.50$ per task. A job that attracts all workers isn't profitable unless it lies below B_{mix}. As the mix of workers shifts from low ability to high ability, B_{mix} rotates upward from B_L to B_H. Therefore, if low-ability workers are sufficiently numerous, B_{mix} passes below the yellow-shaded area, as shown in the figure. In that case, an employer who offered a job corresponding to any point in the yellow-shaded area would lose money. However, if high-ability workers were sufficiently numerous, B_{mix} would pass through the yellow-shaded area. In that case, points below B_{mix} and above both indifference curves correspond to jobs that would attract both types of workers while generating a profit.

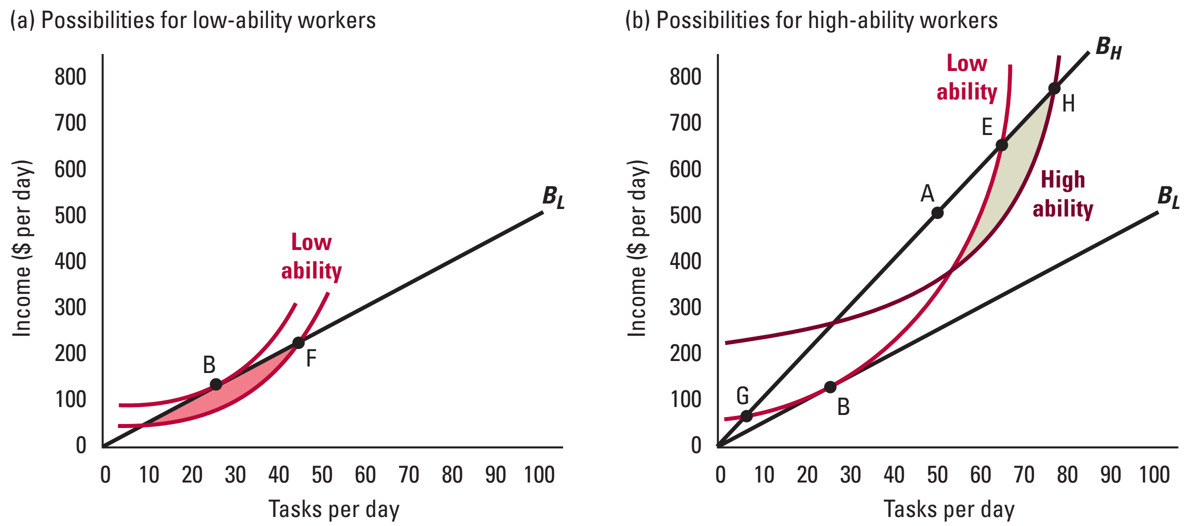
What have we learned? If (and only if) low-ability workers are sufficiently numerous, there is a separating equilibrium in which employers offer the combination of jobs shown in Figure 21.11 (points B and E). Low-ability workers perform the same number of tasks and receive the same wages regardless of whether employers know each worker's ability. The burden of asymmetric information falls on high-ability workers, who perform more tasks when employers are uninformed (point E lies to the right of point A), while receiving the same compensation per task (points E and A both lie on the line B_H). In the separating equilibrium, employers screen workers by presenting them with the following test: "if you want me to believe that you have high ability and pay you $10 per task instead of $5, then prove your ability by agreeing to perform a large number of extra tasks."

Are There Other Separating Equilibria? As it turns out, there are no separating equilibria other than the one described in the preceding section. Figure 21.12(a) shows why low-ability workers must end up at point B. Let's suppose that the job chosen by those workers corresponded to some other point on the line B_L, such as point F. The indifference curve of a low-ability worker that runs through any such point must pass below point B, as shown. If a new employer entered this market and created a job with characteristics corresponding to any point in the red-shaded area of the figure (below B_L and above the indifference curve that runs through point F), it would be able to attract workers—certainly those with low ability, and potentially those with high ability as well—while paying them less than $5 per task. Because that strategy permits the new entrant to earn a profit, the market isn't in a competitive equilibrium.

What about high-ability workers? Recall that they must end up at a point on B_H. In Figure 21.12(b), we've reproduced the low-ability worker's indifference curve from Figure 21.11. That curve intersects the line B_H at point G, as well as point E. To convince ourselves that high-ability workers must end up at point E, let's rule out the alternatives. High-ability workers cannot end up at or to the left of point G, because then they would

Figure 21.12

Possibilities that Do Not Survive Competition. Figure (a) shows that, in a competitive separating equilibrium, low-ability workers must end up at point B. If low-ability workers ended up at another point such as F, an employer could offer a job corresponding to a point in the red-shaded area, attract low-ability workers (and possibly high-ability workers), and earn a profit. Figure (b) shows that high-ability workers must end up at point E. If high-ability workers ended up at a point to the right of point E, such as H, an employer could offer a job corresponding to a point in the green-shaded area, attract only high-ability workers, and earn a profit. We can also rule out points between points G and E, as those would attract low-ability workers, as well as points to the left of point G, as those would fail to attract high-ability workers given that point B is available.



choose the job intended for low-ability workers (point B) instead.[21] They cannot end up between points E and G, because then low-ability workers would choose the type of job intended for high-ability workers. Finally, high-ability workers can't end up at any point to the right of point E, like point H. Why not? The indifference curve of a high-ability worker through any such point must pass below point E, as shown.[22] If a new employer entered this market and created a job with characteristics corresponding to any point in the green-shaded area of the figure (below B_H and between the two indifference curves), it would be able to attract high-ability workers (because the point is above a high-ability worker's indifference curve through point H) but no low-ability workers (because the point is below a low-ability worker's indifference curve through point B), while paying them less than \$10 per task. Because that strategy permits the new entrant to earn a profit, the market isn't in a competitive equilibrium.

[21]As the figure shows, low-ability workers are indifferent between points B and G, and prefer point B to all points to the left of G. Because high-ability workers are even more willing to take on additional tasks for greater compensation, they must prefer point B to point G, as well as to all points to the left of G.

[22]If the indifference curve through point H passed above point E, then it would also pass above point A (because indifference curves bow to the right). But that cannot be, because point A is a high-ability worker's favorite point on the line B_H.

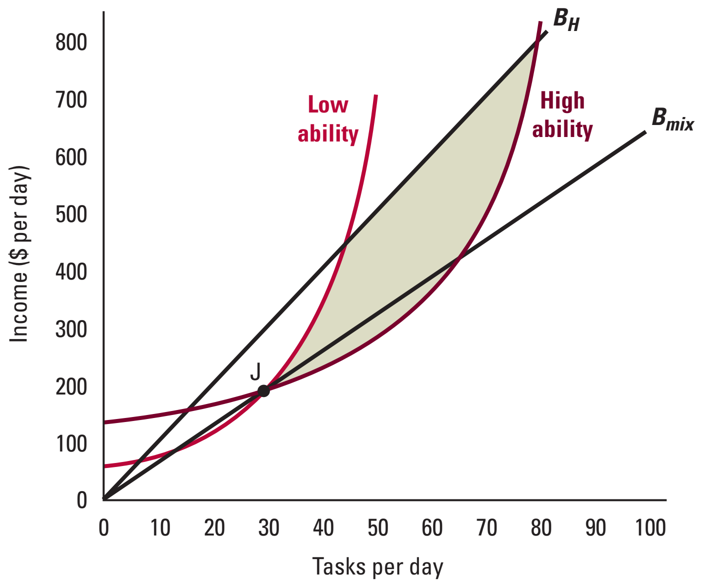


Figure 21.13

Pooling Does Not Survive Competition. If all available jobs correspond to point J, an employer can offer a job corresponding to a point in the green-shaded area, attract only high-ability workers, and earn a positive profit. Therefore, the market is not in equilibrium.

Are There Pooling Equilibria?

In a competitive pooling equilibrium, employers would offer only one type of job, requiring the completion of T_P tasks and paying $\$Y_P$ per day. However, as it turns out, there are never any equilibria of this type. To understand why, let's consider the following two possibilities.

First, we can rule out the possibility that (T_P, Y_P) is either above or below the line labeled B_{mix} in Figure 21.11, which we've reproduced in Figure 21.13. If it were above that line, existing employers would earn positive profits, so new employers would continue to enter the labor market, creating an unlimited number of jobs. If it were below that line, employers would turn away applicants to avoid losing money. In either case, the market wouldn't be in equilibrium.

Second, we can rule out the possibility that (T_P, Y_P) lies on the line B_{mix}. Consider, for example, point J in Figure 21.13. We have drawn two indifference curves through point J, one for each type of worker. If an employer created a job with characteristics corresponding to any point in the green-shaded area of the figure (below B_H and the low-ability worker's indifference curve, but above the high-ability worker's indifference curve), it would attract only high-ability workers, while paying them less than \$10 per task. Because that strategy permits the employer to earn a profit, the market wouldn't be in equilibrium.[23] Having ruled out every conceivable possibility, we conclude that there is no pooling equilibrium.

[23]In reaching this conclusion, we have assumed that employers cannot observe each others' job offers. If employers have the opportunity to turn away job applicants after observing what others have offered, then a pooling equilibrium will exist precisely when the separating equilibrium fails to exist, and vice versa. The jobs offered in the pooling equilibrium will correspond to the best point on the line B_{mix} from the perspective of the high-ability workers. We explain this point in Add-On 21A, in the context of insurance markets.

A Possible Role for the Government

We've seen that when people are asymmetrically informed, competitive markets do not necessarily allocate resources efficiently, even when screening is possible. In some cases competition leads to separating equilibria, which are inefficient. In other cases there are no equilibria with either separation or pooling, and competition is likely to produce unstable and potentially unpredictable outcomes. Therefore, government intervention *may* be justified. However, as in the context of signaling, it is important to bear in mind that the government is probably no better informed (and indeed may be less well informed) than private individuals. The key question is whether there are beneficial public policies that require no more information than the government is likely to possess.

Social insurance is insurance provided by the government.

Adverse selection is perhaps most frequently cited as a justification for government intervention in insurance markets. In practice, governments usually provide citizens with various types of insurance. Government-provided insurance is known as **social insurance**. Examples of social insurance programs in the United States include Medicare, which provides health insurance to elderly individuals; Social Security, which requires all workers to invest in life annuities (see Application 11.3, p. 390); and unemployment insurance, which is typically provided at the state level. Do the market failures associated with adverse selection and competitive screening justify these types of programs?

In an insurance market, **cross-subsidization** occurs when one type of policy generates losses, while another type of policy generates profits, and all policies collectively break even.

Even if the government knows absolutely nothing about individual policyholders' risks, it has one critical advantage over private insurers: the power to compel participation. It can use that power to accomplish two objectives that are not achievable through competitive markets. First, it can mandate a break-even outcome with pooling by requiring everyone to join the pool. As long as it prevents low-risk individuals from opting out, it will escape the consequences of adverse selection. Second, it can induce one type of policyholder to cross-subsidize another. In an insurance market, **cross-subsidization** occurs when one type of policy generates losses, while another type of policy generates profits, and all policies collectively break even. In a competitive market, cross-subsidization is impossible because each policy must break even individually. (We made a similar point in the context of labor markets; see Add-On 21A for a full explanation in the context of insurance.) In contrast, the government can induce one type of individual to cross-subsidize another by offering a menu of policies and requiring everyone to select one of them. For example, it can offer one policy that is chosen by high-risk individuals and loses money, and another that is chosen by low-risk individuals and earns a profit, while still breaking even overall. In that case, low-risk individuals would cross-subsidize high-risk individuals. Unlike private insurers, the government needn't worry that a competitor might try to steal its profitable low-risk customers.

Why might government-mandated pooling or cross-subsidization be advantageous? One potential advantage involves fairness. If competition leads to stable separation of low-risk and high-risk individuals, the high-risk individuals may pay extremely high premiums. That outcome may strike us as particularly unfair in certain contexts. For example, the costs of medical insurance can be astronomical for people who, through no fault of their own, have congenital health problems. Universal pooling spreads those high costs across the entire population.

Less obviously, cross-subsidization can lead to a more efficient allocation of resources. To understand why, look again at Figure 21.11 on page 21-26. (Though the figure pertains to a labor market, the same principles apply in the context of insurance markets; see Add-On 21A.) Suppose that instead of offering jobs corresponding to points B and E, employers offer one job corresponding to a point slightly above B (with the same horizontal coordinate), and another corresponding to a point slightly to the southwest of E, in the yellow-shaded area between B_H and I_H. If low-ability workers end up at the point near B and high-ability workers end up at the point near E, both will be better off. As long as we choose the new points carefully, workers will self-select in exactly that way. Employers will lose money on each job filled by a low-ability worker (because the corresponding point lies above B_L), but earn profits on each job chosen by a high-ability worker (because the corresponding point lies below B_H). However, if there are enough high-ability workers, employers will at least break even overall. Therefore, cross-subsidization can in principle improve upon the separating equilibrium, making everyone better off.

For similar reasons, government-mandated pooling can also lead to more efficient resource allocation. To understand why, suppose the government requires everyone to purchase a policy that provides partial insurance, and permits them to supplement that policy with private insurance. (Such supplementation is quite common in the case of Medicare.) Suppose also that competition among private insurers leads to a separating equilibrium for supplemental insurance. In that case, the government will profit on low-risk individuals and lose money on high-risk individuals, while private insurers will break even on both types of individuals. By varying the size of the social insurance component, the government can, in effect, fine-tune the overall degree to which low-risk individuals cross-subsidize high-risk individuals, and thereby make everyone better off.

A final potential advantage of government intervention involves stability. In some contexts, it is important to make sure that insurance is always available, and that premiums are reasonably steady. If there is no separating equilibrium, then social insurance leads to a stable, predictable outcome, whereas private markets may not.

Application 21.3

The Rat Race

As associates at most large law firms, recent law school graduates receive high salaries but work brutally long hours. Many firms require associates to bill 40 or more hours per week, and some make it clear that young attorneys are expected to exceed those requirements by wide margins. Because the typical attorney spends one hour on nonbillable tasks for every two billable hours, associates frequently find themselves working 60 to 80 hours per week.

What is the point of such burdensome requirements? Surely two well-rested associates, each working 40 hours per week, would be more productive than one overtired, stressed-out associate struggling through 80-hour workweeks. But the senior members of a law firms aren't just interested in their associates' short-term productivity. Ultimately, they must decide whether to invite each associate to join the firm as a partner.

Why does an existing partner care whether an associate is promoted? Each partner's compensation depends in large part on the firm's overall success, not just on her own performance. Because the money-making activities of individual partners are often difficult to observe, and because the economic value of any particular activity can be hard to measure, such income-sharing arrangements are unavoidable. Two important implications follow. First, each partner benefits when the others work hard. Second, the monetary incentives for any given partner to work hard are relatively weak; to some extent, any partner can free ride on the others. (You learned about the problems associated with free-riding in Section 20.5.) To counteract the free-riding problem, existing partners may seek to promote associates who work extremely hard for nonmonetary reasons—for example, because they crave professional success and respect, love their work, have few outside interests, or are simply driven. To determine whether an associate has these characteristics, the firm may confront her with a test: "if you want us to believe that you are the type of person who will continue to work hard after making partner, then prove it to us by agreeing to work an insane number of hours—far more than you would like—for the next several years." The test is effective because workaholics find long hours less unpleasant than those who are less motivated.

A study by economists Renée Landers, James Rebitzer, and Lowell Taylor provides empirical support for the preceding explanation of associates' work hours.[24] Based on data gathered from two large law firms, the study found that nearly two-thirds of associates would prefer to work fewer hours for proportionately less compensation, just as one would expect in a separating equilibrium. Attorneys rated an associate's willingness to work long hours when required as one of the top two factors in promotion decisions, along with quality of work. They attached much less *direct* importance to the number of hours actually billed. However, that factor was *indirectly* important because attorneys construed it as an indicator of the associate's willingness to work hard. Consequently, billable hours mattered a great deal in hypothetical promotion decisions.

Similar considerations may explain why many universities promote faculty to tenured positions only if they have demonstrated high productivity in addition to insight, creativity, and/or brilliance. Because tenure eliminates many of the incentives for expending effort, universities wish to retain the type of professors who will continue to work hard for other reasons. Similar practices are also found in management consulting.

21.4 INCENTIVES AND MORAL HAZARD

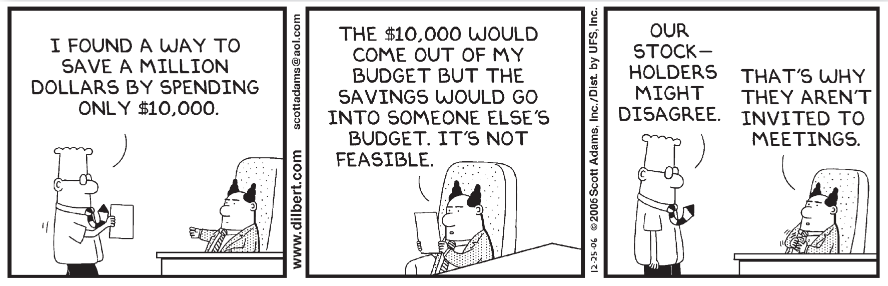
Moral hazard is present when one party to a transaction takes actions that a trading partner cannot observe, and that affect the benefits the partner receives from the trade.

In many situations, one party to a transaction takes actions that a trading partner cannot observe, and that affect the benefits the partner receives from the trade. This form of asymmetric information is known as **moral hazard**.[25] For example, automobile manufacturers are free to vary the quality of inputs and the care with which their cars are assembled. Customers are profoundly affected by those decisions, but rarely can observe them. Likewise, an insured individual's risk of death or disability may be high because of his behavior—smoking, excessive drinking, or a lack of exercise. However, while the company from which he purchases life or disability insurance is obviously affected by those decisions, it is likely to have difficulty monitoring his behavior and adjusting its premiums accordingly.

Moral hazard often arises in employment settings. The profits earned by the owner of a firm depend on the effort of her employees. Often, that effort is difficult to observe. Many employees exploit their employer's informational handicap by exerting less effort than the employer would like. Think of a salesperson who works in the field visiting potential customers; if he relaxes between sales calls with an extra cup of coffee, his employer will never know. Moral hazard is also prevalent in large corporations, where individual managers may take actions that further their own interests at the expense of the firms' owners (shareholders).

[24] Renée M. Landers, James B. Rebitzer, and Lowell J. Taylor, "Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms," *American Economic Review* 86, June 1996, pp. 329–348.

[25] With adverse selection, the benefits of trade depend on a fixed attribute of the good or service being exchanged. One party knows that attribute, while the other does not. With moral hazard, one party has some control over the unobserved attribute.



DILBERT: © Scott Adams/Dist. by United Feature Syndicate, Inc.

In a setting with moral hazard, the uninformed party wants to ensure that her trading partner takes actions that promote her interests. Ideally, she would write a contract that specifies the actions her partner must take. Unfortunately, if neither she nor a court can observe those actions, such a contract would be unenforceable. Nonetheless, there may be observable measures of performance that are affected by the trading partner's actions. For instance, though an employer cannot observe a salesperson's level of effort, she can observe an important consequence of that effort: success in making sales. In such cases, there is a solution to the moral hazard problem, albeit often an imperfect one: provide the trading partner with incentives to take desirable actions by writing a contract or adopting a compensation policy that ties rewards and punishments to observable measures of performance. That is why salespeople are usually compensated at least in part through commissions.

An **incentive scheme** is a contract or compensation policy that ties rewards or punishments to performance, designed in a manner to induce desirable behavior. Many examples of incentive schemes are observed in professional sports. Baseball players' contracts often include incentives tied to measurable aspects of individual and team-oriented performance. These may include bonuses for batting over 0.300 (30 percent) and for reaching the playoffs. Such performance measures do not perfectly capture the amount of effort the player expended to ensure success. Instead, the player's pay is conditioned on outcomes that are related to his effort, but that are also affected by factors outside of his control, such as the abilities of other teams and luck. Nonetheless, tying pay to these outcomes creates incentives for the player to exert effort.

An **incentive scheme** is a contract or compensation policy that ties rewards or punishments to performance, designed in a manner to induce desirable behavior.

In the rest of this section we'll examine the use of incentive schemes, as well as their limitations, in more detail.

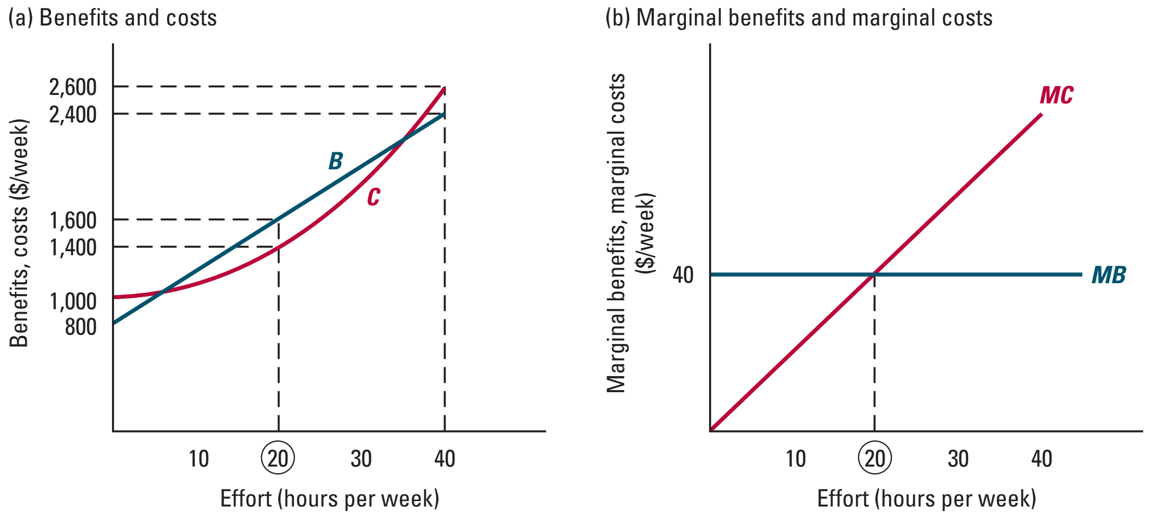
Efficiency and Incentive Pay

Consider the case of a salesperson whose effort may result in profitable sales for his employer, the owner of the car dealership. Other things equal, the salesperson would prefer to relax and not exert much effort. If he does so, however, he probably won't sell many cars. Unfortunately, the owner can't watch him every minute. The salesperson might work harder, though, if the owner structures his compensation so that he earns more if he sells more cars.

Figure 21.14(a) illustrates the weekly costs and benefits of the salesperson's decision. The horizontal axis measures his effort in terms of hours per week. (He is on the job for

Figure 21.14

The Benefits and Costs of a Salesperson's Effort. Figure (a) shows the weekly costs that a salesperson incurs by working hard at a car dealership, as well as the benefits to his employer. Figure (b) shows the corresponding marginal benefit and marginal costs of effort. The efficient outcome entails the salesperson working hard for 20 hours.



40 hours per week, but may not work hard all of the time.) The curve labeled *C* shows the relationship between his effort and the personal costs he incurs while working at the dealership. The curve starts at $1,000, his opportunity cost of accepting employment (for example, this cost could reflect his best alternative job offer). Because effort is costly, the curve is upward-sloping. Figure 21.14(b) shows the corresponding marginal cost curve, labeled *MC*. Notice that the marginal cost of effort increases as the salesperson works harder; as he grows tired, extra effort becomes increasingly costly.

The curve labeled *B* in Figure 21.14(a) shows the owner's benefit from employing the worker. The curve hits the vertical axis at $800; even if the salesperson doesn't work hard at all he can watch the sales floor and lock up at night. The owner's benefit rises as the salesperson expends more effort. In drawing this curve, we've assumed that each hour of hard work creates a 4 percent chance of selling a car. Each sale generates a profit of $1,000, so each additional hour of high effort generates $40, on average, in profit for the dealership. For example, if the salesperson works hard for all 40 hours, he generates, on average, an additional $1,600 in profit (for a total of $2,400); if he works hard only half the time, he generates on average an additional $800 in profit (for a total of $1,600). The corresponding marginal benefit curve, labeled *MB,* is shown in Figure 21.14(b).

Since the benefit curve rises above the cost curve for some levels of effort, it is efficient for the dealership to employ the salesperson. The efficient outcome entails the salesperson working hard for the number of hours that maximizes the difference between the owner's benefit and the salesperson's cost of effort. As shown in Figure 21.14(b), the solution, 20 hours of hard work each week, equates the marginal benefit and marginal cost of effort.

Unfortunately, the owner of the car dealership can't observe the salesperson's effort, so she can't write an employment contract that either specifies his effort level or that ties his pay to his effort. What can she do? One simple possibility is to pay the salesperson a fixed wage each week. That's obviously a poor alternative because it provides the salesperson with no incentive whatsoever to work hard. With a fixed wage, he'll expend no effort.

If, instead, the salesperson's compensation is tied to the number of cars he sells, he will have an incentive to work hard. Figure 21.15(a) illustrates the salesperson's decision if his compensation consists of $1,200 in base pay plus a bonus of $500 for each car he sells. With this incentive scheme, the owner shares the profit from each car ($1,000) equally with the salesperson. The curve labeled *E* reflects the relation between the salesperson's average weekly earnings and his effort. Each hour of effort has a 4 percent chance of generating a sale, so on average it increases his pay by $20. For now, we'll assume that the salesperson cares only about his average pay, and not about the week-to-week variation in his income. (We'll return to this point shortly.) With this assumption, the curve *E* serves as the salesperson's benefit function. Figure 21.15(a) also includes his personal cost function. Curves depicting his marginal earnings (ME) and marginal costs (MC) are shown in Figure 21.15(b). The salesperson will choose his effort to maximize the difference between his personal benefits (earnings) and costs. Equivalently, he equates his marginal earnings and marginal cost. As shown in the figure, his best choice is to work hard for 10 hours. With this compensation scheme, his effort remains inefficiently low (less than 20 hours per week).

Figure 21.15

Incentive Pay and Effort. The figure shows the salesperson's best choice when he is paid $1,200 per week plus a bonus of $500 for each car he sells (assuming each hour of high effort generates a 4 percent chance of selling a car). The curve labeled *E* in figure (a) shows the salesperson's average earnings for each level of effort. The curve labeled *ME* in figure (b) shows the marginal effect of effort on average earnings. The salesperson's best choice is to work hard for 10 hours.

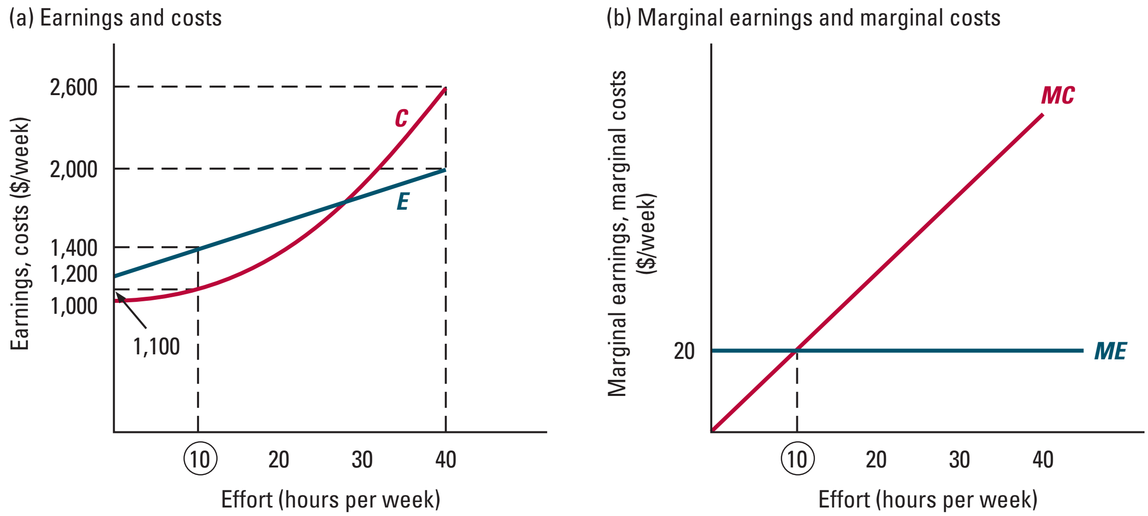
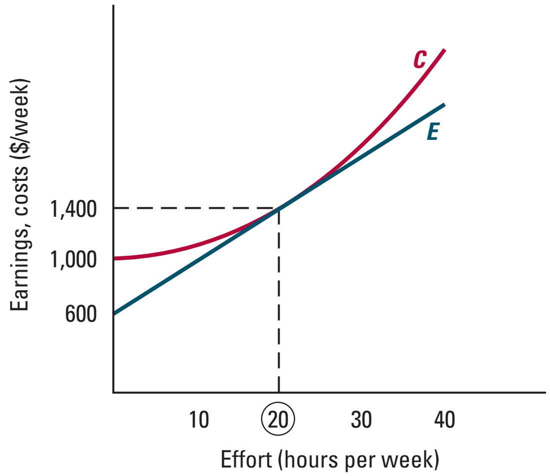


Figure 21.16

An Efficient Incentive Contract that Gives All of the Surplus to the Owner. If the owner gives the salesperson a base pay of $600 plus a bonus of $1,000 for each car he sells, the salesperson will work hard for 20 hours and receive total compensation equal to his personal costs (including opportunity costs). With this incentive scheme, the owner receives all of the surplus.



If the owner of the dealership pays a larger bonus per car sold, the curve *E* will become steeper, and the marginal benefit ME will rise. As a result, the salesperson's best choice, which equates his marginal benefit and marginal cost, will involve greater effort. How can the owner induce the salesperson to choose the efficient effort level? Looking back at Figure 21.14, we see that there is one simple way to accomplish this objective: set the salesperson's compensation equal to the dealership's profits. In that case, his personal benefit curve will be *B,* his marginal benefit will be *MB,* and he will make the efficient choice. In essence, through this incentive scheme, the owner induces the salesperson to behave efficiently by forcing him to absorb all the consequences of varying his effort level.[26]

While offering the salesperson the full profit from all car sales induces him to make an efficient choice, it leaves the owner with no profit: the salesperson receives all the surplus that their relationship creates. To achieve an efficient outcome, however, the owner only needs to provide the salesperson with the right incentives *on the margin.* In Figure 21.16, for example, the owner offers the salesperson a contract that gives him a base pay of $600 and pays him a bonus of $1,000 for each car he sells. The salesperson still chooses the efficient effort level, but now is indifferent between working for the dealership and his best alternative opportunity. With this incentive scheme, the owner retains all of the surplus.

Whether the owner or the salesperson obtains more of the surplus from their relationship will depend on their relative bargaining power. But regardless of how they split the surplus, we've seen that they can enter into a contract that induces efficient effort. (Can you devise a contract that splits the surplus evenly and induces efficient effort?)

Worked-out problem 21.3 shows how to reach these same conclusions using algebra.

[26]In the terminology of Chapter 20, inefficiency arises when the salesperson's effort creates an externality for the owner of the dealership. The incentive scheme described in this paragraph leads to an efficient outcome because it eliminates the externality.

WORKED-OUT PROBLEM 21.3

The Problem A salesperson works for a car dealership for 40 hours per week, but may not choose to work hard all of the time. The dealership's owner cannot observe the salesperson's effort, but can observe the number of cars sold. The salesperson's personal cost of working at the dealership is $C = 1{,}000 + H^2$, where H is the number of hours during which he works hard. The corresponding marginal cost is $MC = 2H$. Without any effort, the salesperson will, on average, generate a profit of \$800. With each hour of high effort, he has a 4 percent chance of selling a car. Each car sale generates a profit of \$1,000. What is the efficient number of hours of high effort? How much surplus does the relationship generate?

Suppose the owner gives the salesperson \$1,200 in base pay plus a bonus of \$500 for each car he sells. How hard will the salesperson work? Describe an incentive scheme that leads to the efficient effort level and allows the owner to keep all of the surplus.

The Solution We first derive the dealership's benefit function. Each hour of hard work has a four percent chance of generating \$1,000 in profit, so it produces \$40 in profit on average. The benefit function is therefore $B = 800 + 40H$ and the marginal benefit is $MB = 40$.

The efficient number of hours of high effort equates the marginal benefit and marginal cost (provided this number of hours results in a nonnegative net benefit). Setting $MB = MC$, we have $2H = 40$. The solution is $H = 20$, which generates a net benefit of $[800 + (40)(20)] - (1{,}000 + 20^2) = \$200$.

If the salesperson faces an incentive scheme with a base pay of \$1,200 and a bonus of \$500 for each car he sells, then each hour of high effort yields, on average, \$20 in extra compensation. The salesperson's expected earnings are $E = 1{,}200 + 20H$ and his marginal earnings (marginal benefits) are $ME = 20$. His best choice equates this marginal benefit and marginal cost (provided his net benefit is positive). Setting $ME = MC$, we have $20 = 2H$, which implies $H = 10$. His net benefit is $\$[(1{,}200 + (20)(10)) - (1{,}000 + 10^2)] = \$300$.

For the salesperson to choose the efficient number of hours, he must receive all the benefits and absorb all the costs of his actions on the margin. Thus, his marginal benefit at 20 hours must be \$40. Consider a linear incentive scheme with base pay of K and a bonus of \$1,000 per car (the full profit). In that case, the salesperson's marginal benefit from an hour of hard work will be \$40 on average. To ensure that the owner keeps all of the surplus, we set K so that the salesperson's net benefit is zero (his total earnings equals his total cost):

$$K + (40)(20) = 1{,}000 + 20^2$$

The solution is $K = 600$.

IN-TEXT EXERCISE 21.3 **Repeat worked-out problem 21.3 assuming instead that each car sale generates a profit of \$1,500.**

Application 21.4

Incentive Pay at Safelite Autoglass

Safelite Glass Corporation is the largest installer of automobile windshield glass in the United States. In the mid 1990s, the company implemented a new compensation scheme for its windshield installers. Before 1994, it paid these workers by the hour. Workers' main incentives to work hard arose from the possibility that Safelite might terminate them if their performance was sufficiently poor. Beginning in 1994, however, Safelite switched to a piece-rate incentive scheme, which linked compensation to the number of windshields a worker installed. Safelite actually paid each worker the maximum of the piece rate compensation and $11 per hour. By guaranteeing a minimum level of compensation, the company allowed its installers to continue working as hard as they had been without sacrificing income. However, the piece rate also enabled workers to earn more income by working harder. The plan was phased in across the company's stores during 1994 and 1995.

Economist Edward Lazear studied the effects of this change in compensation on workers' productivity.[27] Overall, the number of windshields installed per day increased by 44 percent. Lazear showed that productivity rose for two separate reasons. First, installers worked harder. Second, the incentive scheme allowed Safelite to hire more productive installers: because high-ability workers could earn more with the piece-rate plan than with a flat hourly wage, they were more willing to seek and accept employment with Safelite after the change. Lazear identified the effects of the new incentive scheme on effort by measuring the increase in the typical worker's productivity when his or her compensation scheme changed. That gain was, on average, 22 percent. The rest of the productivity gain was attributable to Safelite's improved ability to hire highly productive workers.

Ownership and Incentives

We've seen how an incentive contract can induce a salesperson to expend the efficient level of effort. Making him the *owner* of the business would accomplish the same objective. Suppose the dealership's owner sells the business to the salesperson. After the sale, the salesperson will absorb all of the benefits and costs shown in Figure 21.14, so he will make an efficient choice.

Ownership of an asset entails the right to use the asset in any way the owner wants, and to reap the full benefits from that use (as long has he has not promised the benefits contractually to someone else).[28] Because owners receive all of the benefits resulting from their efforts to generate value from their assets, ownership creates powerful incentives. However, it is not always possible to achieve efficiency solely through the assignment of ownership. For example, most car dealerships have *many* salespeople. Selling the company to one of them would not achieve efficiency, as he or she would still need to employ the others. The dealership must therefore rely on incentive contracts.

---

[27]Edward P. Lazear, "Performance Pay and Productivity," *American Economic Review* 90, December 2000, pp. 1346–1361.

[28]In other words, when an individual owns an asset he controls all of the property rights associated with it (see Section 1.1).

Application 21.5

Owner-Operators in the Trucking Industry

At some point, almost all of us have had the disconcerting experience of driving on an interstate highway surrounded by enormous tractor-trailer trucks. At such anxious moments, few of us—other than economists—wonder who owns those trucks. In some cases, the owner is a large trucking company and the driver is an employee. In other cases, the driver owns the truck and contracts with a company to provide it with trucking services.

What difference does the truck's ownership make? One important issue relates to care of the truck. By driving at a steady speed, a truck's driver can minimize wear and tear. But an employee-driver may well prefer to drive fast and take longer breaks (while still arriving at his destination on time). This is particularly true on long-haul routes, where there is more opportunity to make up time.

Traditionally, it has been difficult for trucking companies to monitor drivers' behavior, even by inspecting trucks for signs of misuse. The relationship between owners and drivers has therefore traditionally involved moral hazard. Ownership by drivers potentially solves this problem. Because drivers then bear the full costs of maintenance and repair, they are likely to drive more responsibly.

Economists George Baker and Thomas Hubbard have examined the roles of moral hazard and incentives in determining ownership patterns within the trucking industry.[29] Two empirical findings led them to conclude that these roles are quite important. First, driver ownership was much more prevalent for trucks used on long-haul routes, where the incentive problem is more severe, than for those used on short-haul routes. In 1987, for example, roughly 21 percent of tractor-trailer trucks used on routes longer than 200 miles were owner-operated, compared with only 8 percent of trucks used on routes shorter than 50 miles.[30]

Second, between 1990 and 1994, the use of on-board computers, a new innovation, diffused through the trucking industry. These devices recorded how the truck was operated, allowing trucking companies to closely monitor driver behavior. If, prior to this development, moral hazard had been an important motivation for driver ownership, the adoption of on-board computers should have reduced the prevalence of owner-operators. Baker and Hubbard found just this pattern: on-board computer adoption was associated with a significant reduction in driver ownership. Indeed, between 1990 and 1994, the fraction of tractor-trailer trucks that were owner-operated fell from 14 to 10 percent. Moreover, these reductions were concentrated in trucks operated on long-haul routes, where driver ownership fell from 21 to 14 percent over four years.

The Costs of Incentives

Unfortunately, it is not always possible to achieve efficiency through incentive schemes. Incentive pay can create costs that we have not yet considered. One problem concerns risk and uncertainty. Take the case of a salesperson at an automobile dealership. Because the number of cars sold depends on factors other than the salesperson's effort, incentive pay makes his income uncertain. For example, depending on the number and types of customers who show up at the dealership, a salesperson who works hard for 20 hours may sell

[29] George P. Baker and Thomas N. Hubbard, "Contractibility and Asset Ownership: On-Board Computers and Governance in U.S. Trucking," *Quarterly Journal of Economics* 119, November 2004, pp. 1443–1479.

[30] If driver ownership solves the moral hazard problem associated with the care of a truck, why aren't all trucks owner-operated? One of the likely reasons is that the relationship between the driver and the trucking company involves other incentive issues. For example, if the driver owns the truck, the company may need to negotiate with him to pick up a load when it identifies someone who needs trucking services. This can take time, and also may give the driver the ability to extract more surplus when other trucks are unavailable. Company ownership gives the firm greater ability to control the deployment of trucks, and therefore provide trucking services cheaply and quickly.

10 cars, 5 cars, or none at all. If he doesn't like uncertainty—that is, if he is risk averse (a term introduced in Chapter 11)—a contract featuring strong incentives will create large costs associated with risk-bearing. In that case, the most efficient incentive contract will balance the benefits of greater effort against the costs of exposure to risk. As a result, an efficient incentive scheme typically does not make an employee absorb all of the marginal costs and benefits of his actions. Add-On 21B discusses the trade-off between incentives and risk-bearing in greater detail.

A second problem with incentive pay is that it may induce employees to concentrate on measured aspects of performance and ignore other aspects that are not easily measured. For example, an important part of a salesperson's job is to help customers with service issues related to prior sales. This service is important because dissatisfied customers are unlikely to return and may tell others about their experience, harming the dealership's reputation and depressing its future sales. Unfortunately, while the dealership can measure the salesperson's current sales quite accurately, it cannot easily measure contributions to overall customer satisfaction. If it offers strong incentives on current sales—the aspect of performance it can measure—the salesperson may devote most of his attention to customers who are considering purchases, and largely ignore those who are seeking after-sale service. Thus, in getting more of what it pays for (current sales effort), the dealership may get less of what it does not specifically pay for (customer service). Finding the right level of incentive pay then involves balancing its positive effect on current sales with its negative effect on after-sales customer service. Sometimes, the negative effect can be so great that it is best not to use an incentive scheme at all.

An important example of this second problem arises in the context of compensation for teachers in elementary and secondary schools. Recent education reforms have stressed teacher accountability by tying teachers' pay to the performance of their students on standardized tests. These reforms have been controversial. Critics complain that they lead teachers to "teach to the test." While they create strong incentives for teachers to improve their students' test performance, they may also cause teachers to ignore other important (and less measurable) aspects of good education, such as fostering creativity. Application 21.6 discusses another example of "getting (exactly) what you pay for."

Application 21.6

Health-Care Report Cards

In the United States, over 1 million people experience heart attacks each year. About 38 percent of those attacks result in death. Indeed, coronary artery disease accounts for roughly 20 percent of all deaths in the United States, more than any other cause.

In light of these statistics, the care and prevention of heart attacks is a critical priority for public policy. In the early 1990s the states of New York and Pennsylvania decided to publish "health-care report cards" that publicly reported patient death rates from coronary artery bypass graft surgery (commonly known as bypass surgery) for each hospital and physician who performed that procedure. The purpose was twofold: first, allow consumers to make more informed choices among hospitals and physicians; second, create incentives for hospitals and physicians to improve their surgical procedures.

Unfortunately, the health-care report cards led to some serious unintended consequences. Hospital administrators and cardiac surgeons soon discovered that they could greatly improve their measured performance by carefully selecting their patients. By operating only on relatively healthy patients and diverting more serious cases to other health care providers, a hospital or doctor could dramatically reduce measured fatalities from the procedure. Surveys and anecdotal evidence suggested that following the introduction of report cards, finding health care providers willing to serve severely ill cardiac patients became considerably more difficult.

To evaluate these claims, economists David Dranove, Daniel Kessler, Mark McClellan, and Mark Satterthwaite compared cardiac care outcomes in New York and Pennsylvania with outcomes in states that did not adopt health-care report cards.[31] To measure a patient's health status prior to surgery (and hence the likely severity of his illness), they calculated the patient's total medical expenditures in the year prior to his or her hospital admission (we'll call these *prior year expenditures*). They found that between 1990 and 1994 (a period that spans the adoption of health-care report cards in New York and Pennsylvania), the prior year medical expenditures of patients receiving bypass surgery fell by 7 to 9 percent in New York and Pennsylvania, but were virtually unchanged in other states. This pattern confirms the suspicion that doctors and hospitals in New York and Pennsylvania had indeed become more selective in accepting patients for bypass surgery. The authors also found that health outcomes for severely ill heart attack victims in New York and Pennsylvania deteriorated following the introduction of health-care report cards.

Other Sources of Incentives

So far, we've discussed two ways to provide people with incentives: performance-based compensation (or other contractual payments) and asset ownership. In this section, we mention some other commonly observed forms of incentives.

Sometimes, bonuses for good performance do not directly involve money. For example, the prospect of a promotion provides incentives not only because promotions are usually associated with increased compensation, but also because the worker's job becomes more interesting or entails greater status.

In ongoing relationships, incentives are often informal, rather than formally specified in a contract. Promotions and raises, for example, are usually based on an informal understanding of the criteria used to evaluate performance, rather than the achievement of measurable benchmarks. Likewise, a firm and one of its input suppliers may have an informal understanding about the input's quality. Supplying a lower quality version of the input may not constitute breach of contract, but it may nevertheless induce the firm to terminate the relationship and find another supplier.

The desire to establish or maintain a reputation can also provide incentives. A poorly performing supplier may lose not only the affected customer's business, but also the business of other customers who hear about the problem. (The health-care report cards discussed in Application 21.6 represented an attempt to harness the incentives associated with reputation.) In a similar vein, workers are often motivated by the desire to create reputations for reliability and hard work, particularly early in their careers. A worker who builds a favorable reputation by performing exceptionally is likely to find herself with attractive opportunities in the future, including tempting offers from other employers.

[31]David Dranove, Daniel Kessler, Mark McClellan, and Mark Satterthwaite, "Is More Information Better? The Effects of 'Report Cards' on Health Care Providers," *Journal of Political Economy* 111, June 2003, pp. 555–588.

The efficient provision of incentives often involves a blend of the various alternatives that we've discussed, including performance-based pay, ownership rights (such as stock or stock options), informal understandings, and reputational considerations. When designing formal incentive schemes, it is always important to bear in mind that these schemes do not operate in a vacuum but rather supplement other forms of incentives.

CHAPTER SUMMARY

1. **Adverse selection**
   - **a.** Often one party to a transaction has more information than another.
   - **b.** Adverse selection is present if an informed individual is more willing to trade when trading is less advantageous to an uninformed trading partner.
   - **c.** The presence of adverse selection leads to deadweight losses and can even drive attractive trading partners out of the market entirely, a phenomenon known as market unraveling.
   - **d.** To reduce the losses due to adverse selection, governments sometimes mandate minimum quality standards or impose product liability laws. In addition, profit-seeking firms also collect information and provide it to uninformed individuals, and the uninformed sometimes share information or engage in costly information gathering.
2. **Signaling**
   - **a.** Signaling occurs when an informed individual undertakes a costly activity to convince others of particular facts. In many situations, signaling offers a partial solution to problems that arise from adverse selection.
   - **b.** If education is less costly for people with high ability than for people with low ability, then education may serve as a signal of ability.
   - **c.** In a separating equilibrium, workers with high ability will obtain more education than workers with low ability. Employers will infer a worker's ability from the amount of schooling received. As a result, additional education will lead to higher pay, even if it has no effect on productivity. That outcome is Pareto inefficient.
   - **d.** There are many separating equilibria, distinguished by the amount of education that high-ability workers obtain. According to one theory, only the most efficient separating equilibrium is likely to persist.
   - **e.** In a pooling equilibrium, workers of all abilities will obtain the same level of education. Employers infer that a worker must have lower-than-average ability if she obtains less education.
   - **f.** There are many pooling equilibria, distinguished by the amount of education that all workers obtain. According to one theory, pooling equilibria are unlikely to persist, even though they may be more efficient than separating equilibria.
   - **g.** Pooling equilibria may be particularly inefficient if high-ability workers drop out of the labor force in disproportionate numbers as the wage rate declines. In that case, a separating equilibrium can prevent the market from unraveling partially or completely.
   - **h.** Government intervention is potentially justified only if there are beneficial public policies that require no more information than the government is likely to possess.
   - **i.** Because signaling is potentially wasteful, taxes on signals can be relatively efficient sources of revenue.
3. **Screening**
   - **a.** Uninformed parties can attempt to overcome asymmetric information by testing either the informed parties or the goods those parties seek to trade. Screening occurs when an uninformed party establishes a test that induces informed parties to self-select, thereby revealing what they know.
   - **b.** If hard work is less costly for people with high ability than for people with low ability, then employers may attempt to screen workers. In a separating equilibrium, they offer demanding jobs with high pay and easy jobs with low pay. High-ability workers self-select into the demanding jobs, and low-ability workers self-select into the easy jobs.
   - **c.** Competition between employers can lead to a separating equilibrium only if low-ability workers are sufficiently numerous. If employers cannot observe each others' offers, competition cannot lead to a pooling equilibrium.
   - **d.** The market failures associated with adverse selection and competitive screening are most frequently cited

as justifications for government intervention in the context of insurance markets. Governments typically provide various forms of social insurance.

e. The government has one critical advantage over private insurers: the power to compel participation. It can use that power to mandate a break-even outcome with pooling, or to induce one type of policyholder to cross-subsidize another. In principle, governments can thereby improve upon the fairness, efficiency, and stability of a free market.

4. **Incentives and moral hazard**

a. In some circumstances, one party to a transaction takes actions that a trading partner cannot observe, and that affect the benefits the partner receives from the trade.

b. A solution to moral hazard, albeit often an imperfect one, is for the uninformed party to provide the trading partner with incentives to take desirable actions by means of an incentive contract, which ties rewards or punishments to observable measures of performance.

c. Efficiency can be achieved (in the absence of risk aversion) by having the party taking unobservable actions absorb the full consequences of his actions.

d. Ownership is one way to confront an individual with the full benefits and costs of their actions.

e. Incentive pay can also create costs and lead to inefficiencies. One such cost involves risk bearing. Another is that individuals may be led to take actions that improve measurable aspects of performance at the expense of aspects that are less easily measured.

f. Incentive pay is but one source of incentives. Others include the nonmonetary benefits of promotion, and the consequences violating an informal understanding or tarnishing a reputation.

ADDITIONAL EXERCISES

Exercise 21.1: Repeat worked-out problem 21.1 (page 21-6), but assume that the supply function of low-ability workers is $Q_L^s = 0.20(W - 2{,}000)$.

Exercise 21.2: It is often said that as soon as the buyer of a new car drives it off of a dealer's lot, its value falls significantly below the purchase price. Why might this be true? Suppose this pattern held in 1985, but that it is no longer true today. What do you think might account for that change?

Exercise 21.3: Each entry-level software programmer in Palo Alto, California, has either high or low ability. All potential employers value a high-ability worker at $12,000 per month and a low-ability worker at $6,000. The supply of high-ability workers is $Q_H^s = 0.1(W - 7{,}000)$ and the supply of low-ability workers is $Q_L^s = 0.1(W - 2{,}000)$, where W is the monthly wage. [These are the functions that lead to the supply curves in Figure 21.3(a).] If workers' abilities are observable to employers, what are the equilibrium wages? How many workers of each type will employers hire? If workers' abilities are not observed by employers, what is the equilibrium wage? How many workers of each type will employers hire? What is the deadweight loss due to asymmetric information?

Exercise 21.4: Each entry-level software programmer in Palo Alto, California, has either high or low ability. All potential employers value a high-ability worker at $44,000 per month and a low-ability worker at $6,000. The supply of high-ability workers is $Q_H^s = 0.05(W - 2{,}000)$ and the supply of low-ability workers is $Q_L^s = 0.1(W - 2{,}000)$, where W is the monthly wage. If workers' abilities are observable to employers, what are the equilibrium wages? How many workers of each type will employers hire? If workers' abilities are not observed by employers, what is the equilibrium wage? How many workers of each type will employers hire? What is the deadweight loss due to asymmetric information?

Exercise 21.5: Groucho Marx once famously quipped, "I would never belong to a club that would have me as a member." Interpret Groucho's observation as an example of adverse selection.

Exercise 21.6: Give five examples of signaling phenomena, other than those discussed in the text.

Exercise 21.7: Let's make the following change to the model of educational signaling discussed in Section 21.2: Assume that education is more costly for high-ability workers than for low-ability workers (perhaps because those with high ability have more to lose by staying in school). Draw a new figure like Figure 21.4, showing the relationship between the indifference curves of high-ability and low-ability workers. Are there any separating equilibria? Are there any pooling equilibria? In each case, explain your answer using graphs.

Exercise 21.8: Consider once again the problem described in worked-out problem 21.2 (page 21-19). Suppose the government taxes wages based on the amount of education a worker has received. For a worker receiving E years of education, the tax is $\$5(E - 10)$ per hour worked. Thus, the utility functions become $U_H(E, W) = W - 5E - 5(E - 10)$ for a high-ability worker, and $U_L(E, W) = W - 10E - 5(E - 10)$ for a low-ability worker.

a. Find the most efficient separating equilibrium and compare it to the most efficient separating equilibrium with no tax. Are the high-ability workers better off or worse off? Do they obtain more or less education? What about low-ability workers?

b. Now suppose that high-ability and low-ability workers are equally numerous. The government distributes all revenue back to workers through lump-sum payments (so that no worker thinks her own decisions affect the amount received); every worker receives exactly the same amount. Are high-ability workers better off or worse off? What about low-ability workers? How do your answers change if high-ability workers outnumber low-ability workers? What if low-ability workers outnumber high-ability workers?

Exercise 21.9: Give five examples of screening, other than those discussed in the text.

Exercise 21.10: Let's make the following change to the model of workplace responsibilities discussed in Section 21.3: Assume that more able workers find tasks more tiring (perhaps because they have less tolerance for tedious make-work). Draw a new figure like Figure 21.8, showing the relationship between the indifference curves of high-ability and low-ability workers. Are there any separating equilibria? Are there any pooling equilibria? In each case, explain your answer using graphs.

Exercise 21.11: Consider once again the model of educational attainment discussed in Section 21.2. Suppose that employers can announce in advance the amount of education that will be required for any given job. In other words, treat this as a screening problem rather than as a signaling problem. Are there any separating equilibria? Pooling equilibria? If so, what are they? Upon what do your answers depend? Explain using graphs.

Exercise 21.12: After reading Sections 21.2 and 21.3, a student complains: "The world doesn't work this way. Once an employee spends some time at a firm, her employer will learn her true ability and treat her accordingly." Does this criticism undermine the analysis and conclusions presented in this chapter? Why or why not?

Exercise 21.13: After learning about market failures, George states: "Any time the market outcome is Pareto inefficient, government intervention can make everyone better off, at least in principle." Do you agree? Why, or why not? Would you agree with a qualified version of George's claim? If so, what qualifications would you add?

Exercise 21.14: Repeat in-text exercise 21.3 (page 21-37) assuming instead that each car sale generates a profit of $2,000.

Exercise 21.15: Repeat in-text exercise 21.3 (page 21-37) assuming instead that the salesperson's personal cost is $C = 1{,}000 + 2H^2$, and his marginal cost is $MC = 4H$.

Exercise 21.16: Why might it be a good idea to compensate those teaching classes in bookkeeping based on their students' performance on standardized tests, but not a good idea to compensate those teaching Ph.D. students in math based on their students' perfomance on standardized tests?

Exercise 21.17: In Chile, bus drivers have at times been paid based on the number of riders they pick up. Discuss the likely differences in driver behavior under this plan compared to a plan in which drivers are paid by the day.