# 13

# Mechanism Design

JAMES MIRRLEES WON THE NOBEL PRIZE in economics in 1996 for his pioneering work on optimal nonlinear income taxation and related policy issues. Many noneconomists, and some economists too, found his work difficult to understand. But the *Economist* magazine gave a brilliant characterization of the broad importance and relevance of the work. It said that Mirrlees showed us "how to deal with someone who knows more than you do."[1]

In Chapter 8, we observed some of the ways in which such asymmetric information affects the analysis of games. But the underlying problem for Mirrlees differed slightly from the situations we considered earlier. In his work, one player (the government) needed to devise a set of rules so that the other players' (the taxpayers') incentives were aligned with the first player's goals. Models with this general framework, in which a less-informed player works to create motives for the more-informed player to take actions beneficial to the less informed, now abound and are relevant to a wide range of social and economic interactions. Generally, the less-informed player is called the *principal* while the more-informed is called the *agent*; hence these models are termed *principal–agent* models. And the process that the principal uses to devise the correct set of incentives for the agent is known as **mechanism design**.

In Mirrlees's model, the government seeks a balance between efficiency and equity. It wants the more productive members of society to contribute effort to increase total output; it can then redistribute the proceeds to benefit the poorer

[1] "Economics Focus: Secrets and the Prize," *Economist*, October 12, 1996.

members. If the government knew the exact productive potential of every person and could observe the quantity and quality of his effort, it could simply order everyone to contribute according to their ability, and it could distribute the fruits of their work according to people's needs. But such detailed information is costly or even impossible to obtain, and such redistribution schemes can be equally difficult to enforce. Each person has a good idea of his abilities and needs and chooses his own effort level, but stands to benefit by concealing this information from the government. Pretending to have less ability and more needs will enable him to get away with paying less tax or getting larger checks from the government; the incentive to provide effort is reduced if the government takes part of the yield. The government must calculate its tax policy, or design its fiscal mechanism, taking into account these problems of information and incentives. Mirrlees's contribution was to solve this complex mechanism design problem within the principal–agent framework.

The economist William Vickrey shared the 1996 Nobel Prize in economics with Mirrlees for his own work in mechanism design in the presence of asymmetric information. Vickrey is best known for designing an auction mechanism to elicit truthful bidding, a topic we will study in greater detail in Chapter 16. But his work extended to other mechanisms, such as congestion pricing on highways, and he and Mirrlees laid the groundwork for extensive research in the subfield.

Indeed, in the past thirty years, the general theory of mechanism design has made great advances. The 2007 Nobel Prize in economics was awarded to Leonid Hurwicz, Roger Myerson, and Eric Maskin for their contributions to it. Their work, and that of many others, has taken the theory and applied it to numerous specific contexts, including the design of compensation schemes, insurance policies, and of course tax schedules and auctions. In this chapter, we will develop a few prominent applications, using our usual method of numerical examples followed by exercises.

## 1 PRICE DISCRIMINATION

A firm generally sells to diverse customers with different levels of willingness to pay for its product. Ideally, the firm would like to extract from each customer the maximum that he would be willing to pay. If the firm could do so, charging each customer an individualized price based on willingness to pay, economists would say that it was practicing perfect (or first-degree) **price discrimination**.

Such perfect price discrimination may not be possible for many reasons. The most general underlying reason is that even a customer who is willing to pay a lot prefers to pay less. Therefore, the customer will prefer a lower price, and this firm may have to compete with other firms or resellers who undercut its high price. But even if there are no close competitors, the firm usually does not know how much each individual customer is willing to pay, and the customers will try to get away with pretending to be unwilling to pay a high price so as to secure a lower price. Sometimes even if the firm could detect the willingness to pay, it may be illegal to practice blatant first-degree price discrimination based on the identity of the buyer. In such situations, the firm must devise a product line and prices so that the customers' choices of what they buy (and therefore what they pay) go some way toward the firm's goal of increasing its profit by way of price discrimination.

In our terminology of asymmetric information games developed in Chapter 8, the process by which the firm identifies customer willingness to pay from purchase decisions involves *screening* to achieve *separation of types* (by self-selection). The firm does not know each customer's *type* (willingness to pay), so it tries to acquire this information from their actions. An example that should be familiar to most readers is that of airlines. These firms try to separate business flyers, who are willing to pay more for their tickets, from tourists, who are not willing to pay that much, by offering low prices in return for various restrictions on fares that the business flyers are not willing to accept, such as advance-purchase and minimum-stay requirements.[2] We develop this particular example in more detail to make the ideas more precise and quantifiable.

We consider the pricing decisions of a firm called Pie-In-The-Sky (PITS), an airline running a service from Podunk to South Succotash. It carries some business passengers and some tourists; the former type are willing to pay a higher price than the latter for any particular ticketed seat. To serve the tourists profitably without having to offer the same low price to the business passengers, PITS has to develop a way of creating different versions of the same flight; it then needs to price these options in such a way that each type will choose a different version. As mentioned above, the airline could distinguish between the two types of passengers by offering restricted and unrestricted fares. The practice of offering first-class and economy-class tickets is another way to distinguish between the two groups; we will use that practice as our example.

Suppose that 30% of PITS's customers are businesspeople and 70% are tourists. The table in Figure 13.1 shows the (maximum) willingness to pay for each type of customer for each class of service, along with the costs of providing the two types of service and the potential profits available under each option.

[2] This type of pricing policy, offering different prices to different groups of customers on the basis of some distinguishable characteristic, is generally known as *third-degree* price discrimination. It is thus differentiated from the *first-degree* discrimination described earlier. There is also *second-degree* price discrimination, which occurs when firms charge different unit prices for customers purchasing different quantities of a product.

| Type of service | PITS's cost | Reservation price | | PITS's potential profit | |
|---|---|---|---|---|---|
| | | Tourist | Business | Tourist | Business |
| Economy | 100 | 140 | 225 | 40 | 125 |
| First | 150 | 175 | 300 | 25 | 150 |

**FIGURE 13.1** Airline Price Discrimination

We begin by setting up a ticket-pricing scheme that is ideal from PITS's point of view. Suppose it knows the type of each individual customer; salespeople determine customer type, for example, by observing their style of dress when they come to make their reservations. Also suppose that there are no legal prohibitions on differential pricing and no possibility that lower-priced tickets can be resold to other passengers. (Actual airlines prevent such resale by requiring positive ID for each ticketed passenger.) Then PITS could practice perfect (first-degree) price discrimination.

How much would PITS charge to each type of customer? It could sell a first-class ticket to each businessperson at \$300 for a profit of \$300 − \$150 = \$150 per ticket or sell him an economy ticket at \$225 for a profit of \$225 − \$100 = \$125 per ticket. The former is better for PITS, so it would want to sell \$300 first-class tickets to these business customers. It could sell a first-class ticket to each tourist at \$175 for a profit of \$175 − \$150 = \$25 or an economy ticket at \$140 for a profit of \$140 − \$100 = \$40. Here the latter is better for PITS, so it would want to sell \$140 economy-class tickets to the tourists. Ideally, PITS would like to sell only first-class tickets to business travelers and only economy-class tickets to tourists, in each case at a price equal to the relevant group's maximum willingness to pay. PITS's total profit per 100 customers from this strategy would be

$$(140 - 100) \times 70 + (300 - 150) \times 30 = 40 \times 70 + 150 \times 30 = 2{,}800 + 4{,}500 = 7{,}300.$$

Thus, PITS's best possible outcome earns it a profit of \$7,300 for every 100 customers it serves.

Now turn to the more realistic scenario in which PITS cannot identify the type of each customer or is not allowed to use the information for purposes of overt discrimination. How can it use the different ticket versions to screen its customers?

The first thing PITS should realize is that the pricing scheme devised above will not be the most profitable in the absence of identifying information about each customer. Most important, it cannot charge the business travelers their full \$300 willingness to pay for first-class seats while charging only \$140 for an economy-class seat. Then the businesspeople could buy economy-class seats, for which they are willing to pay \$225, and get an extra benefit, or "consumer surplus" in the jargon of economics, of \$225 − \$140 = \$85. They might use this surplus, for example, for better food or accommodation on their travels. Paying the maximum \$300 that they are willing to pay for a first-class seat would leave them no consumer surplus. Therefore, they would switch to economy class in this situation, and screening would fail. PITS's profit per 100 customers would drop to (140 − 100) × 100 = \$4,000.

The maximum that PITS will be able to charge for first-class tickets must give business travelers at least as much extra benefit as the \$85 they can get if they buy an economy-class ticket. Thus, the price of first-class tickets can be at most \$300 − \$85 = \$215. (Perhaps it should be \$214 to give business travelers a definite positive reason to choose first class, but we will ignore the trivial difference.) PITS can still charge \$140 for an economy-class ticket to extract as much profit as possible from the tourists, so its total profit in this case (from every 100 customers) would be

$$(140 - 100) \times 70 + (215 - 150) \times 30 = 40 \times 70 + 65 \times 30 = 2{,}800 + 1{,}950 = 4{,}750.$$

This profit is more than the \$4,000 that PITS would get if it tried unsuccessfully to implement its perfect discrimination scheme despite its limited information, but less than the \$7,300 it would get if it had full information and successfully practiced perfect price discrimination.

By pricing first-class seats at \$215 and economy-class seats at \$140, PITS can successfully screen and separate the two types of travelers on the basis of their self-selection of the two types of services. But PITS must sacrifice some profit to achieve this indirect discrimination. PITS loses this profit because it must charge the business travelers less than their full willingness to pay. As a result, its profit per 100 passengers drops from the \$7,300 it could achieve if it had full and complete information, to the \$4,750 it achieves from the indirect discrimination based on self-selection. The difference, \$2,550, is precisely 85 × 30, where 85 is the drop in the first-class fare below the business travelers' full willingness to pay for this service, and 30 is the number of these business travelers per 100 passengers served.

Our analysis shows that, in order to achieve separation with its ticket-pricing mechanism, PITS has to keep the first-class fare sufficiently low to give the business travelers enough incentive to choose this service. Those travelers have the option of choosing economy class if it provides more benefit (or surplus) to them; PITS has to ensure that they do not "defect" to making the choice that PITS intends for the tourists. Such a requirement, or constraint, on the screener's strategy arises in all problems of mechanism design and is called an *incentive-compatibility constraint.*

The only way PITS could charge business travelers more than \$215 without inducing their defection would be to increase the economy-class fare. For example, if the first-class fare is \$240 and the economy-class fare is \$165, then business travelers get equal consumer surplus from each class; their surplus is \$300 − \$240 from first class and \$225 − \$165 from economy class, or \$60 from each. At those higher prices, they are still (only just) willing to buy first-class tickets, and PITS could enjoy higher profits from each first-class ticket sale.

But at \$140, the economy-class fare is already at the limit of the tourists' willingness to pay. If PITS raised that fare to \$165, say, it would lose these customers altogether. In order to keep these customers willing to buy, PITS's pricing mechanism must meet an additional requirement, namely the tourists' *participation constraint.*

PITS's pricing strategy is thus squeezed between the participation constraint of the tourists and the incentive-compatibility constraint of the businesspeople. If it charges $X$ for economy and $Y$ for first class, it must keep $X < 140$ to ensure that the tourists still buy tickets, and it must keep $225 - X < 300 - Y$, or $Y < X + 75$, to ensure that the business travelers choose first-class and not economy. Subject to these constraints, PITS wants to charge prices that are as high as possible. Therefore, its profit-maximizing screening strategy is to make $X$ as close to 140 and $Y$ as close to 215 as possible. Ignoring the small differences that are needed to preserve the $<$ signs, let us call the prices 140 and 215. Then charging \$215 for first-class seats and \$140 for economy-class seats is the solution to PITS's mechanism-design problem.

This pricing strategy being optimal for PITS depends on the specific numbers in our example. If the proportion of business travelers were much higher, say 50%, PITS would have to revise its optimal ticket prices. With 50% of its customers being businesspeople, the sacrifice of \$85 on each business traveler may be too high to justify keeping the few tourists. PITS may do better not to serve the tourists at all, that is, to violate their participation constraint and to raise the price of first-class service. Indeed, the strategy of discrimination by screening with these percentages of travelers yields PITS a profit, per 100 customers, of

$$(140 - 100) \times 50 + (215 - 150) \times 50 = 40 \times 50 + 65 \times 50 = 2{,}000 + 3{,}250 = 5{,}250.$$

The strategy of serving only business travelers in \$300 first-class seats would yield a profit (per 100 customers) of

$$(300 - 150) \times 50 = 150 \times 50 = 7{,}500,$$

which is higher than with the screening prices. Thus, if there are only relatively few customers with low willingness to pay, the seller might find it better not to serve them at all than to offer sufficiently low prices to the mass of high-paying customers to prevent their switching to the low-priced version.

Precisely what proportion of business travelers constitutes the borderline between the two cases? We leave this as an exercise for you. And we will just point out that an airline's decision to offer low tourist fares may be a profit-maximizing response to the existence of asymmetric information, rather than an indication of some soft spot for vacationers!

## 2 SOME TERMINOLOGY

We have now seen one example of mechanism design in action. There are many others, of course, and we will see additional ones in later sections. We pause briefly here, however, to set out the specifics of the terminology used in most models of this type.

Mechanism-design problems are broadly of two kinds. In the first, which is similar to the price-discrimination example above, one player is better informed (in the example, the customer knows his own willingness to pay), and his information affects the payoff of the other player (in the example, the airline's pricing and therefore its profits). The less-informed player designs a scheme in which the better-informed player must make some choice that will reveal the information, albeit at some cost to the first (in the example, the airline's inability to charge the business flyers their full willingness to pay).

In the second kind of mechanism-design problem, one player takes some action that is not observable to others. For example, an employer cannot observe the quality, or sometimes even the quantity, of the effort an employee exerts, and an insurance company cannot observe all the actions that an insured driver or homeowner takes to reduce the risk of an accident or robbery. In the language of Chapter 8, this problem is one of *moral hazard*. The less-informed player designs a scheme—for example, profit sharing for the employee or deductibles and copayments for insurance—that aligns the other player's incentives to some extent with those of the mechanism designer.

In each case, the less-informed player designs the mechanism; he is called the **principal** in the strategic game. The more-informed player is then called the **agent**; this is most accurate in the case of the employee and less so in the cases of the customer or the insured, but the jargon has become established and we will adopt it. The game is then called a **principal–agent**, or **agency**, problem.

The principal in each case designs the mechanism to maximize his own payoff, subject to two types of constraints. First, the principal knows that the agent will utilize the mechanism to maximize his own (the agent's) payoff. In other words, the principal's mechanism has to be consistent with the agent's incentives. As we saw in Chapter 8, Section 4.B, this is called the *incentive-compatibility constraint*. Second, given that the agent responds to the mechanism in his own best interests, the agency relationship has to give the agent at least as much expected utility as he would get elsewhere, for example by working for someone else, or by driving instead of flying. In Chapter 8, we termed this the *participation constraint*. We saw specific examples of both constraints in the airline price-discrimination story in the previous section; we will meet many other examples and applications later in this chapter.

# 3 COST-PLUS AND FIXED-PRICE CONTRACTS

When writing procurement contracts for the acquisition of certain services, perhaps highway or office-space construction, governments and firms face mechanism-design problems of the kind we have been describing. Two common methods for writing such contracts are "cost-plus" and "fixed-price." In a cost-plus contract, the supplier of the services is paid a sum equal to his cost, plus an allowance for normal profit. In a fixed-price contract, a specific price for the services is agreed on in advance; the supplier keeps any extra profit if his actual cost turns out to be less than anticipated, and he bears the loss if his actual costs are higher.

Each type of contract has its own good and bad points. The cost-plus contract appears not to give the supplier excessive profit; this characteristic is especially important for public-sector procurement contracts, where the citizens are the ones who ultimately pay for the procured services. But the supplier typically has better information about his cost than does the buyer of his services; therefore the supplier can be tempted to overstate the cost or to pad the costs in order to extract some benefit from the wasteful excess. The fixed-price contract, in contrast, gives the supplier every incentive to keep the cost at a minimum and thus to achieve an efficient use of resources. But with this kind of public-sector contract, society has to pay the set price and give away any excess profit (to the supplier). The optimal procurement mechanism should balance these two considerations.

## A. Highway Construction: Full Information

We will consider the example of a state government designing a procurement mechanism for a road-construction project. Specifically, suppose that a major highway is to be built by the state's road contractor and that the government has to decide how many lanes it should have.[3] More lanes yield more social benefit in the form of faster travel and fewer accidents (at least up to a point, beyond

[3] Generally, numerous contractors could be competing for the highway-construction contract. For this example, we restrict ourselves to the case in which there is only one contractor.

which the harm to the countryside will be too great). To be specific, we suppose that the social value $V$ (measured in billions of dollars) from having $N$ lanes on the highway is given by the formula:

$$V = 15N - \frac{N^2}{2}.$$

The cost of construction per lane, including an allowance for normal profit, could be either \$3 billion or \$5 billion per lane, depending on the types of soil and minerals located in the construction zone. For now, we will assume that the government can identify the construction cost as well as the contractor. So it chooses $N$ and writes a contract to maximize the benefit to the state ($V$) net of the fee paid to the contractor (call it $F$); that is, the government's objective is to maximize net benefit, $G$, where $G = V - F$.[4]

Suppose first that the government knows the actual cost is 3 (billion dollars per lane of highway). At this cost level, the government has to pay $3N$ to the contractor for an $N$-lane highway. The government then chooses $N$ to maximize $G$, as above, where the appropriate formula in this situation is:

$$G = V - F = 15N - \frac{N^2}{2} - 3N = 12N - \frac{N^2}{2}.$$

Recall that in the appendix to Chapter 5, we gave a formula for finding the correct value to maximize this type of function. Specifically, the solution to the problem of choosing $X$ to maximize

$$Y = A + BX - CX^2$$

is $X = B/(2C)$. Here $Y$ is $V$, $X$ is $N$, and $A = 0$, $B = 12$, and $C = 1/2$. Applying our solution formula yields the government's optimal choice of $N = 12/(2 \times 1/2) = 12$. The best highway to choose therefore has 12 lanes, and the cost of that 12-lane highway is \$36 billion. So the government offers the contract: "Build a 12-lane highway and we will pay you \$36 billion."[5] This price includes normal profit, so the contractor is happy to take the contract.

Similarly, if the cost is \$5 billion per lane, the optimal $N$ will be 10. The government will offer a \$50 billion contract for the 10-lane highway. And the contractor will accept the contract.

[4] In reality, the cost per lane would not have only two discrete values, but could take any value along a continuous range of possibilities. The probabilities of each value would then correspondingly form a density function on this range. Our methods will not always yield an integer solution, $N$, for each possible cost along this range. But we leave these matters to more advanced treatments and confine ourselves to this simple illustrative example.

[5] In reality, there will be many clauses specifying quality, timing, inspections, and so forth. We leave out these details to keep the exposition of the basic idea of mechanism design simple.

## B. Highway Construction: Asymmetric Information

Now suppose that the contractor knows how to assess the relevant terrain to determine the actual per lane building cost, but the government does not. The government can only estimate what the cost will be. We will assume that it thinks that there is a two-thirds probability of the cost being 3 (billion dollars per lane) and a one-third probability of the cost being 5.

What if the government tries to go ahead with the ideal optimum and offers a pair of contracts: "12-lane highway for \$36 billion" and "10-lane highway for \$50 billion"? If the cost is really only \$3 billion per lane, the contractor will get more profit by taking the latter contract even though that one was designed for the situation in which the cost is \$5 billion per lane. The true cost of the 10-lane highway would be only \$30 billion, and the contractor would earn \$20 billion in excess profit.[6]

This outcome is not very satisfying. The contracts offered do not give the contractor sufficient incentive to choose between them on the basis of cost; he will always take the \$50 billion contract. There must be a better way for the government to design its procurement contract system.

So now we allow the government the freedom to design a more general mechanism to separate the types of projects. Suppose it offers a pair of contracts: "Contract L: Build $N_L$ lanes and get paid $R_L$ dollars" and "Contract H: Build $N_H$ lanes and get paid $R_H$ dollars." If contracts L and H are designed correctly, when cost is low (\$3 billion per lane) the contractor will pick contract L (L stands for "low"), and when cost is high (\$5 billion per lane) he will pick contract H (H stands for "high"). The numbers that the symbols $N_L$, $R_L$, $N_H$, and $R_H$ represent must satisfy certain conditions for this screening mechanism to work.

First, under each contract, the contractor facing the relevant cost (low for contract L and high for contract H) must receive enough to cover his cost (inclusive of normal profit). Otherwise he will not agree to the terms; he will not participate in the contract. Thus, the contract must satisfy two *participation constraints*: $3N_L \leq R_L$ for the contractor when the cost is 3, and $5N_H \leq R_H$ for the contractor when the cost is 5.

Next, the government needs the two contracts to be such that a contractor who knows his cost is low would not benefit by taking contract H and vice versa. That is, the contracts must also satisfy two incentive-compatibility constraints. For example, if the true cost is low, contract L will yield excess profit

[6] If multiple contractors are competing for the job, the ones not selected may spill the beans about the true cost here. But for large highway projects (as for many other large government projects, such as defense contracts), there are often only a few potential contractors, and they do better by colluding among themselves and not revealing the private information. For simplicity, we keep the analysis confined to the case where there is just one contractor.

$R_L - 3N_L$, whereas contract H will yield $R_H - 3N_H$. (Note that in the latter expression, the number of lanes and the payment are as specified in the H contract, but the contractor's cost is still only 3, not 5.) To be incentive compatible for the low-cost case, the contracts must keep the latter expression no larger than the former. Thus, we need $R_L - 3N_L \geq R_H - 3N_H$. Similarly, if the true cost is high, the contractor's excess profit from the L contract must be no larger than his excess profit from the H contract, so to be incentive compatible, we need $R_H - 5N_H \geq R_L - 5N_L$.

The government wants to maximize the net expected social value of the payment and uses the probabilities of the two types as weights to calculate the expectation. Therefore, the government's objective here is to maximize

$$G = \left(\frac{2}{3}\right)\left[15N_L - \frac{(N_L)^2}{2} - R_L\right] + \left(\frac{1}{3}\right)\left[15N_H - \frac{(N_H)^2}{2} - R_H\right].$$

The problem looks formidable, with four choice variables and four inequality constraints. But it simplifies greatly, because two of the constraints are redundant, and the other two must hold as exact equalities, allowing us to solve and substitute for two of the variables.

Note that if the participation constraint when cost is high, $5N_H \leq R_H$, and the incentive compatibility constraint when cost is low, $R_L - 3N_L \geq R_H - 3N_H$, both hold, then we can get the following string of inequalities (where we have used the fact that $N_H$ will be positive):

$$R_L - 3N_L \geq R_H - 3N_H \geq 5N_H - 3N_H \geq 5N_H \geq 0.$$

The first and last expressions in the inequality string tell us that $R_L - 3N_L \geq 0$. Therefore, we need not consider the participation constraint when cost is low, $3N_L \leq R_L$, separately; it is automatically satisfied when the two other constraints are satisfied.

It is also intuitive that the high-cost firm will not want to pretend to be low cost; it would get compensated for the smaller cost while incurring the larger cost. However, this intuition needs to be verified by the rigorous logic of the analysis. Therefore, we proceed as follows. We will begin by ignoring the second incentive compatibility constraint, $R_H - 5N_H \geq R_L - 5N_L$, and we will solve the problem with just the remaining two constraints. Then we will return and verify that the solution to the two-constraint problem satisfies the ignored third constraint anyway. So our solution must also be the solution to the three-constraint problem. (If something better was available, it would also work better for the less-constrained problem.)

Thus, we have two constraints to consider: $5N_H \leq R_H$ and $R_L - 3N_L \geq R_H - 3N_H$. Write these as $R_H \geq 5N_H$ and $R_L \geq R_H + 3(N_L - N_H)$. Then observe that $R_L$ and $R_H$ each enter negatively in the government's objective; it wants to make them as small as is compatible with the constraints. This result is achieved by satisfying each constraint with equality. So we set $R_H = 5N_H$ and $R_L = R_H + 3(N_L - N_H) = 3N_L + 2N_H$. These expressions for the contract payments can now be substituted into the objective function, G. This substitution yields:

$$G = \left(\frac{2}{3}\right)\left[15N_L - \frac{(N_L)^2}{2} - 3N_L - 2N_H\right] + \left(\frac{1}{3}\right)\left[15N_H - \frac{(N_H)^2}{2} - 5N_H\right].$$

$$= 8N_L - \frac{(N_L)^2}{3} + 2N_H - \frac{(N_H)^2}{6}.$$

The objective function now splits cleanly into two parts; one (the first two terms) involves only $N_L$, and the other (the second two terms) involves only $N_H$. We can apply our maximization formula separately to each part. In the $N_L$ part, the $A = 0$, $B = 8$, and $C = 1/3$, so the optimal $N_L = 8/(2 \times 1/3) = 24/2 = 12$. In the $N_H$ part, the $A = 0$ again, $B = 2$, and $C = 1/6$, so the optimal $N_H = 2/(2 \times 1/6) = 12/2 = 6$.

Now we can use the optimal values for $N_L$ and $N_H$ to derive the optimal payment (R) values, using the formulas for $R_L$ and $R_H$ that we derived just above. Substituting $N_L = 12$ and $N_H = 6$ into those formulas gives us $R_H = 5 \times 6 = 30$ and $R_L = 3 \times 12 + 2 \times 6 = 48$. We thus have optimal values for all of the unknowns in the government's objective function. But remember that we ignored one of the incentive-compatibility constraints, so we need to go back to that now.

We must ensure that the ignored third constraint, $R_H - 5N_H \geq R_L - 5N_L$, holds with our calculated values for the Rs and the Ns. In fact, it does. The left-hand side of the expression equals $30 - 5 \times 6 = 0$. And the right-hand side equals $48 - 5 \times 12 = -12$, so the constraint is indeed satisfied.

Our solution indicates that the government should offer the following two contracts: "contract L: build 12 lanes and get paid 48 (billion dollars)" and "Contract H: build 6 lanes and get paid 30 (billion dollars)." How can we interpret this solution so as best to understand the intuition for it? The intuition is most easily seen when we compare the solution here with the ideal one we found in Section 3.A under full information about costs. Figure 13.2 shows the comparisons in the optimal N and R values.

The optimal mechanism under asymmetric information differs in two important respects from the one we found when information was perfect. First, although the contract intended to be chosen if the contractor's cost is low has the same number of lanes (12) as in the full-information case, its payment to

| | $N_L$ | $R_L$ | $N_H$ | $R_L$ |
|---|---|---|---|---|
| Perfect Information | 12 | 36 | 10 | 50 |
| Asymmetric Information | 12 | 48 | 6 | 30 |

**FIGURE 13.2** Highway-Building Contract Values

the contractor is larger in the asymmetric case (48 instead of 36). Second, the high-cost asymmetric-information contract has a smaller number of lanes (six instead of 10) but pays just the full cost for that number ($30 = 6 \times 5$). Both of these differences separate the types.

Under asymmetric information, the contractor may be tempted to pretend that the cost is high when it is in fact low. The optimal payment mechanism then incorporates both a carrot for truthfully admitting to low cost and a stick for trying to pretend to be high cost. The carrot is the excess profit, $48 - 36 = 12$, that comes from the admission made implicitly by the choice of contract L. The stick is the reduction in excess profit from contract H, achieved by reducing the number of lanes that will be constructed in that case. The ideal high-cost mechanism would have the highway be 10 lanes and would pay \$50 billion; the contractor whose true cost is low would make excess profit of $50 - 3 \times 10 = \$20$ billion. In the information-constrained optimal contract, only six lanes are constructed, and the contractor is paid \$30 billion. If the true cost is low, he makes an excess profit of $30 - 3 \times 6 = \$12$ billion. His benefit from the pretense (implicitly made by the choice of contract H even though his true cost is low) is reduced. In fact it is reduced exactly to the amount that he is guaranteed by the carrot part of the mechanism, thereby exactly offsetting his temptation to pretend high cost.

## 4 EVIDENCE CONCERNING INFORMATION REVELATION MECHANISMS

The mechanisms considered so far have the common feature that the agent has some private information, which we called the player's type in Chapter 8. Further, the principal requires the agent to take some action that is designed to reveal this information. In the terminology of Chapter 8, these mechanisms are examples of screening for the separation of types by self-selection.

We see such mechanisms everywhere. Those for price discrimination are the most ubiquitous. All firms have customers who are diverse in their willingness to pay for the firms' products. As long as a customer is willing to pay more than the firm's incremental cost of supplying the product to him, the firm can turn a profit by dealing with this customer. But this customer's willingness to pay may be relatively low in comparison to that of other potential buyers. If a firm must charge the same price to all of its customers, including those who would have been willing to pay more than this one, charging this customer's willingness to pay means the firm has to sacrifice some profit from its higher-willingness customers. Ideally, the firm would like to discriminate by giving a price break to the less-willing customers without giving the same break to the more-willing ones.

The ability of a firm to practice price discrimination may be limited for reasons other than those of information. It may be illegal to price discriminate.

Competition from other firms may limit this firm's ability to charge high prices to some of its customers. And if the product can be bought by one customer and resold to others, such competition from other buyers may be just as effective a constraint on discriminatory pricing as competition from other firms. But here we focus on the information reasons for price discrimination, keeping the other reasons in the background of the discussion.

Your local coffee shop probably has a "frequent-drinker card"; for every ten cups you buy, you get one free. Why is it in the firm's interest to do this? Frequent drinkers are more likely to be locals, who have the time and incentive to search out the best deals in the neighborhood. To attract those customers away from other competing coffee shops, this one must offer a sufficiently attractive price. In contrast, infrequent customers are more likely to be strangers in the town or in a hurry and have less time and incentive to search for the best deals; when they need a cup of coffee and see a coffee shop, they are willing to pay whatever the price is (within reason). So posting a higher price and giving out frequent-drinker cards enables this coffee shop to give a price break to the price-sensitive regular customers without giving the same price break to the occasional buyers. If you don't have the card, you are revealing yourself as the latter type, willing to pay a higher price.

In a similar manner, many restaurants offer fixed-price three-course menus or blue-plate specials, as well as regular à la carte offerings. This strategy enables them to separate diverse customer types with different tastes for soups, salads, main courses, desserts, and so on.

Book publishers start selling new books in a hardback version and issue a paperback version a year or more later. The price difference between the two versions is generally far greater than the difference in the costs of production of the two kinds of books. The idea behind the pricing scheme is to separate two types of customers, those who need or want to read the book immediately and are willing to pay more for the privilege, and those who are willing to wait until they can get a better price.

We invite you to look for other examples of such screening mechanisms for price discrimination in your own purchases. They appear in myriad ways. You can also read good accounts of such practices. One good source is Tim Harford's *Undercover Economist*.[7]

There is a lot of research literature on procurement mechanisms of the kind we sketched in Section 3.[8] These models pertain to situations where the buyer

[7] Tim Harford, *The Undercover Economist: Exposing Why the Rich Are Rich, the Poor Are Poor—and Why You Can Never Buy a Decent Used Car!* (New York: Oxford University Press, 2005). The first two chapters give examples of pricing mechanisms.

[8] Jean-Jacques Laffont and Jean Tirole, *A Theory of Incentives in Procurement and Regulation* (Cambridge, Mass.: MIT Press, 1993), is the classic of this literature.

confronts just one potential seller whose cost is private information. This type of interaction accurately describes how contracts for major defense weapons systems or very specialized equipment are designed; there is usually only one reliable supplier of such products or services. However, in reality buyers often have the choice of several suppliers, and mechanisms that set the suppliers in competition with each other are beneficial to the buyer. Many such mechanisms take the form of auctions. For example, construction contracts are often awarded by inviting bids and choosing the bidder that offers to do the job for the lowest price (after adjusting for the promised quality of the work and speed of completion or other relevant known attributes of the bid). We will give some examples and discussion of such mechanisms in the chapter on auctions.

## 5 INCENTIVES FOR EFFORT: THE SIMPLEST CASE

We now turn from the first type of mechanism-design problem, those in which the principal's goal is to achieve information revelation, to the second type, in which there is moral hazard. The principal's goal in such situations is to write a contract that will induce the best effort level from the agent, even though that effort level is unobservable by the principal.

### A. Managerial Supervision

Suppose you are the owner of a company that is undertaking a new project. You have to hire a manager to supervise it. The success of the project is uncertain, but good supervision can increase the probability of success. Managers are only human, though; they will try to get away with as little effort as they can! If their effort is observable, you can write a contract that compensates the manager for his trouble sufficiently to bring forth good supervisory effort.[9] But if you cannot observe the effort, you have to try to give him incentives based on success of the project, for example a bonus. Unless good effort absolutely guarantees success, however, such bonuses make the manager's income uncertain. And the manager is likely to be averse to risk, so you have to compensate him for facing such risk. You have to design your compensation policy to maximize your own expected profit, recognizing that the manager's choice of effort depends

[9] Most important, if a dispute arises, you or the manager must be able to prove to a third party, such as an arbitrator or a court, whether the manager made the stipulated effort or shirked. This condition, often called *verifiability*, is more stringent than mere observability by the parties to the contract (you and the manager). We intend such public observability or verifiability when we use the more common term *observability.*

on the nature and amount of the compensation. This is a mechanism-design problem whose solution is intended to cope with the moral-hazard problem of the manager's shirking.

Let us consider a numerical example. Suppose that if the project succeeds, it will earn the company a profit of \$1 million over material and wage costs. If it fails, the profit will be zero. With good supervision, the probability of success is one-half, but if supervision is poor, the probability of success is only one-quarter.

As mentioned above, the manager is risk averse. We saw in the appendix to Chapter 8 how risk aversion can be captured by a concave utility function. So let us take a simple case, where the manager's utility $u$ from income $y$ (measured in millions of dollars) is the square-root function: $u = \sqrt{y}$. Suppose also that the manager gets disutility 0.1 from the extra effort that is needed for good supervision. Finally, suppose that if the manager does not work for you, he can get another job that does not require any extra effort and that pays \$90,000, or \$0.09 million, yielding utility $\sqrt{0.09} = 0.3$. Thus, if you want to hire the manager without requiring good supervision, you have to pay at least \$90,000. If you want good supervision, you have to guarantee the manager at least as much utility as he could get from taking the other job; you must pay the $y$ that ensures $\sqrt{y} - 0.1$ is at least 0.3, or $\sqrt{y} \geq 0.4$, or $y \geq 0.16$, or \$160,000.

If effort is observable, you can write one of two contracts: (1) I pay you \$90,000, and I don't care if you shirk; or (2) I pay you \$160,000, and you have to make a good supervisory effort. This second contract can be enforced by a court, so if the manager accepts it, he will in fact make good effort. Your expected profit from each contract depends on the probability that the project succeeds with the specified level of effort. So expected profit from the first is $(1/4) \times 1 - 0.09 = 0.160$, or \$160,000, and that from the second is $(1/2) \times 1 - 0.16 = 0.340 =$ \$340,000. Therefore, you are better off paying the manager to provide good effort. In an ideal world of full information, you will use the second contract.

Now consider the more realistic scenario in which the manager's effort is not observable. This situation presents no extra problems if you would like the manager to exert low effort, and the first contract above applies. But if you would like good supervisory effort, you must use an incentive mechanism based on the only observable, namely success or failure of the project. So suppose you offer a contract that pays the manager $x$ if the project fails and $y$ if it succeeds. (Note that $x$ may be zero, but if that is optimal, it should emerge from the solution. In fact it will not be zero, because of the manager's risk aversion.)

To induce the manager to choose high effort, you must ensure that his expected utility from doing so is higher than his expected utility from shirking. With high effort, he can guarantee a one-half chance that the project succeeds, and he therefore faces a one-half chance that it fails. With ordinary effort, he can guarantee only a one-quarter chance of success (a three-quarter chance of failure). So your contract must ensure the following:

$$(1/2)\sqrt{y} + (1/2)\sqrt{x} - 0.1 > (1/4)\sqrt{y} + (3/4)\sqrt{x}, \quad \text{or}$$

$$(1/4)(\sqrt{y} - \sqrt{x}) \geq 0.1, \quad \text{or} \quad \sqrt{y} - \sqrt{x} \geq 0.4.$$

This expression is the *incentive-compatibility constraint* in this problem.

Next, you have to ensure that the manager gets enough expected utility to be willing to work for you *in the way you want* (exerting high supervisory effort) rather than taking his other possible offer. So his expected utility from accepting your job and exerting high effort must exceed his utility from the alternate job; your contract must then satisfy the following:

$$(1/2)\sqrt{y} + (1/2)\sqrt{x} - 0.1 \geq 0.3, \quad \text{or} \quad \sqrt{y} + \sqrt{x} \geq 0.8.$$

This expression is the *participation constraint* for your contract intended to elicit high supervisory effort.

Subject to these constraints, you want to maximize your expected profit, Π. You calculate that expected profit under the assumption that by meeting the constraints above, you are eliciting high supervisory effort. Thus, you assume that your project succeeds with probability one-half and your expected profit expression is:

$$\Pi = (1/2)(1 - y) + (1/2)(0 - x) = (1 - y - x)/2.$$

The mathematics in this problem becomes much easier if we work with the square roots of $x$ and $y$ instead of $x$ and $y$ themselves (that is, we work with the utilities of income instead of the incomes). Write these utilities as $X = \sqrt{x}$ and $Y = \sqrt{y}$, so $x = X^2$ and $y = Y^2$. Then you want to maximize

$$\Pi = (1 - Y^2 - X^2)/2$$

subject to the participation constraint

$$Y + X \geq 0.8$$

and the incentive-compatibility constraint

$$Y - X \geq 0.4.$$

Both $X$ and $Y$ enter negatively into the expression for your expected profit, so you want to make both as small as is compatible with the constraints. The participation constraint eventually holds with equality when both $X$ and $Y$ are made small. What about the incentive-compatibility constraint? If it does not also eventually hold with equality, then it does not constrain the choices and can be ignored. Let us suppose that is the case. Then we can substitute $X = 0.8 - Y$ from the participation constraint into your profit expression and write

$$\Pi = (1 - Y^2 - X^2)/2 = [1 - Y^2 - (0.8 - Y)^2]/2$$
$$= (1 - Y^2 - 0.64 + 1.6Y - Y^2)/2$$
$$= (0.36 + 1.6Y - 2Y^2)/2 = 0.18 + 0.8Y - Y^2.$$

To maximize this profit expression, we again use the formula from the appendix to Chapter 5; we have $B = 0.8$ and $C = 1$. This yields the optimal $Y = 0.8/(2 \times 1) = 0.4$. Then $X = 0.8 - 0.4 = 0.4$ also.

This solution implies that if the incentive-compatibility constraint is ignored, the optimal mechanism requires equal payment to the manager whether the project succeeds or fails. This payment is just enough to give the manager a utility of $0.4 = 0.3 + 0.1$ (his utility from easy work elsewhere plus compensation for the disutility of the extra effort for high supervision) to meet the participation constraint. This result is intuitive and in keeping with our discussion of optimal risk bearing in Chapter 8, Section 1. The manager is risk averse and you are risk neutral (concerned with expected profit alone), so it is efficient for you to bear all of the risk and to keep the manager's income nonrandom.[10]

But if the manager gets the same income whether the project succeeds or fails, he has no incentive to make the unobservable effort. So the ignored incentive-compatibility constraint is not going to be fulfilled automatically, and we must make sure that $X$ and $Y$ do satisfy it. We therefore need both of the constraints to hold with equality: $Y + X = 0.8$ and $Y - X = 0.4$. Adding the two constraints together, we get $2Y = 1.2$ or $Y = 0.6$; this result immediately yields $X = 0.2$. Translating from utilities into dollar amounts, we have $x = X^2 = 0.04$ and $y = Y^2 = 0.36$. Thus, the manager should be paid \$40,000 if the project fails and \$360,000 if it succeeds. The payment for failure is less than the \$90,000 he would be paid for the low-effort contract 1 in the full-information case, and the payment for success is more than the \$160,000 for the high-effort contract 2 in the full-information case. Thus, the manager faces a combination of a stick (low pay if the project fails) and a carrot (high pay if it succeeds), just as does the contractor in the highway construction example of Section 3.

With this scheme, you (the owner) make an expected profit of:

$$\Pi = (1 - 0.36 - 0.04)/2 = 0.30,$$

or \$300,000. This amount is less than the \$340,000 you would make in the full-information ideal, when you could write an enforceable contract stipulating high effort. The \$40,000 difference is an unavoidable cost of the information asymmetry.

The manager's compensation scheme can be described as a base salary of \$40,000 and a success bonus of \$320,000, or equivalently, a \$40,000 salary and a 32% share in the operating profit of \$1 million. It would not be desirable for you

[10] The case in which the owner is also risk averse can be treated by similar methods.

to rely on profit sharing alone, offering the manager no base salary. Why not? If the salary component were zero, then in the event of the project's success you would have to pay the manager an amount $y$ defined by $(1/2)\sqrt{y} - 0.1 = 0.3$, or $y = 0.64$, or $640,000 to ensure his participation. Your expected profit would be

$$\Pi = (1 - 0.64 - 0)/2 = 0.180, \quad \text{or } \$180{,}000.$$

Thus, your profit in this case would be $120,000 lower than when you offered a $40,000 base salary with bonus (and a full $160,000 below what you could earn in the full-information case). The reduction in profit is due to the fact that the manager is risk averse. A pure-bonus scheme makes his income very risky, so to ensure his participation you have to make the bonus so large that it cuts into your profit. The optimal asymmetric information payment scheme balances the stick and the carrot optimally to provide enough incentive for the manager to make high supervisory effort, but without imposing too much risk on his income.

## B. Insurance Provision

Moral hazard can arise in other relationships beyond those in the labor market described above. Insurance markets in particular are subject to problems of moral hazard. And insurance companies must determine whether and how to offer appropriate insurance contracts that encourage their clients to take appropriate actions to reduce their likelihood of needing to file a claim with the company. For example, insurers would like those to whom they sell health insurance to continue regular wellness visits to their physicians and those to whom they sell car insurance to continue to practice defensive-driving techniques.[11] Because the insurance company cannot usually observe the clients' actions, however, creating the appropriate insurance policy will require an understanding of the theory of mechanism design in the face of asymmetric information.

Here we return to our example of a farmer facing the risk of crop failure due to some bad-weather outcome, such as a drought. We met this farmer originally in Chapter 8, Section 1. There we supposed that the farmer's income would be $160,000 if the weather proved favorable and $40,000 if not. When the two possibilities are equally likely, probability 0.5 each, the farmer's expected income is $0.5 \times \$160{,}000 + 0.5 \times \$40{,}000 = \$100{,}000$. The farmer faces considerable risk around this average value, however, and if he is risk averse, he will care about the expected utility of the outcomes rather than just about his expected income.

Suppose then that the farmer is indeed risk averse. His utility function is $u = \sqrt{I}$, where $I$ represents his income. The farmer therefore gets utility of $400 = \sqrt{160{,}000}$ if the weather is good (wet) and utility of $200 = \sqrt{40{,}000}$ if the weather is bad (dry). His expected utility is then $0.5 \times 400 + 0.5 \times 200 = 300$.

[11] Indeed, insurance companies regard the policyholders' failure to take such risk-reducing precautions as immoral behavior; this is the origin of the term *moral hazard*.

What would happen if this farmer could avoid the risk associated with a year of drought? Specifically, what would his situation be if he could ensure that his income was always \$100,000 (the expected value here) rather than \$160,000 half of the time and \$40,000 the other half of the time? Ignoring for a moment how he could make this happen, we note that the farmer gets utility of about $316 \approx \sqrt{100{,}000}$ every year under this outcome. The farmer therefore would enjoy higher expected utility ($316 > 300$) if he could find a way to smooth his income (and his utility) across the good- and bad-weather years.

One possible way for the farmer to achieve income smoothing is by way of insurance. A risk-neutral insurance company could offer the farmer a contract where the farmer pays the company \$60,000 in good-weather years and the insurer pays the farmer \$60,000 in bad-weather years. Because the probability of each outcome is 50%, the company's expected profit from this contract is exactly zero, making it just willing to offer the contract to the farmer. The farmer is strictly better off accepting the contract, however; his expected utility rises. So an insurance contract that is full (completely covers the cost of a bad outcome) and fair (priced just to offset the cost of the farmer's claims) would be acceptable to both parties.

So far, this example has no information problem. But the farmer could take various actions to reduce the probability of the low income level associated with drought. He may be able to construct some water-catchment basins, for example, that would allow him to water his crops in all but the driest of years. However, the construction and maintenance of the basins will be at some cost to the farmer. If the basins are of good quality and well maintained, they will help protect the farmer from the risks associated with a drought. If the basins are shoddy ones that leak and are not well cared for, they will fail to do their job and so do not reduce the risk of crop failure from drought. If the farmer is well insured and the quality of the basins and their level of maintenance is not observable by simple inspection, he may be tempted to shirk the task to save himself the costs; this potential for shirking is the source of moral hazard in our example.

Suppose that the farmer's disutility of making the extra effort to construct and maintain high-quality water basins is 25,[12] and that with them in existence the farmer reduces the probability of the bad outcome to 25%. Then the farmer's expected income with the basins is $0.75 \times \$160{,}000 + 0.25 \times \$40{,}000 = \$130{,}000$ and his expected utility (in the absence of insurance) is $0.75 \times \sqrt{160{,}000} + 0.25 \times \sqrt{40{,}000} - 25 = 0.75 \times 400 + 0.25 \times 200 - 25 = 350 - 25 = 325$. The farmer's expected utility is higher with the basins than without them ($325 > 300$), so if no

[12] Formally, the farmer's utility function is now $u = \sqrt{I} - E$, where $I$ is again income, and $E$ is the disutility of effort, 25 if the basins are of good quality and 0 if they are shoddy.

insurance is available, the farmer will definitely want to make the risk-reducing effort to construct the water basins.

The farmer could still benefit from insurance in this case. An income-smoothing policy that guaranteed him \$130,000 every year would ensure an expected utility of $360 (\approx \sqrt{130{,}000}) - 25 = 335$, even when he builds and maintains high-quality basins. This utility is higher than the 325 he receives when he builds the basins but has no insurance, so the farmer would definitely prefer the insurance.

Suppose that a full and fair insurance contract could be written that stipulated that the farmer exert the effort necessary to reduce the probability of a bad outcome to 25%. Suppose further that the insurance company could verify the farmer's effort by sending an insurance agent to the farm to check on the water basins. Then the contract that guaranteed the farmer \$130,000 income each year would entail the farmer's paying the insurance company \$30,000 in a good-weather year and the insurer paying the farmer \$90,000 in a bad-weather year. As before, the insurance company reaps an expected profit of exactly zero with this contract ($0.75 \times 30{,}000 - 0.25 \times 90{,}000 = 0$) but the farmer's expected utility increases (to 335), so both parties will agree to the contract.

If the insurer cannot verify the farmer's effort, then the situation changes. The farmer could cheat and accept the "pay \$30,000 in a good year, get \$90,000 in a bad year" insurance contract but not make the stipulated effort (build shoddy basins and provide no maintenance). Then the probability of having a bad year reverts to 50%, but the farmer's income is \$130,00 every year. His expected utility from accepting such a contract but making no effort is 360 ($\approx \sqrt{130{,}000}$), which is better than all of the other possibilities we have so far considered. Of course, the insurance company does badly in this case. Its expected profit is $0.5 \times \$30{,}000 - 0.5 \times \$90{,}000 = -\$30{,}000$. The insurer cannot survive this contract, given the moral-hazard problem, and so will not offer it to the farmer.

Does this mean that the farmer cannot get insurance at all when he has the option to build and maintain water basins, but his insurer can't verify their quality and maintenance? No. But it does mean that he cannot get full insurance. There is still the option of a *partial* insurance contract in which the insurance company takes on a part, but not all, of the risk associated with a bad outcome.

Recall that when the farmer can build and maintain high-quality basins, full insurance entails his paying the insurer \$30,000 in a good year and receiving \$90,000 in a bad year. This contract gave the farmer no incentive actually to build or maintain the basins and left the insurer with a negative expected profit. To design the optimal insurance scheme here, the insurance company needs to determine the right $X$ to require as payment from the farmer in a good year (leaving the farmer $\$160{,}000 - X$) and the right $Y$ to pay out to the farmer in a bad year (boosting the farmer's income to $40,000 + Y$). Then the optimal mechanism must maximize the insurance company's expected profit, given $X$, $Y$, and the probabilities of the different outcomes while ensuring both that the farmer retains the incentive to build the catch basins and that he is willing to accept the insurance contract.

Because calculating the optimal values for $X$ and $Y$ here is quite complex, we will instead consider a specific pair of numbers that offers the farmer some insurance, gives him enough incentive to make the effort to reduce his risk, and breaks even for the company. Suppose the insurance company offers a contract that goes one-third of the way toward full insurance. Such a contract would stipulate a payment from the farmer of $10,000 in a good year (leaving him with $150,000) and a payment to the farmer of $30,000 in a bad year (bringing him to $70,000). If the farmer does build and maintain the high-quality basins, then this contract leaves the insurance company with an expected profit of $0.75 \times \$10{,}000 - 0.25 \times \$30{,}000 = \$7{,}500 - \$7{,}500 = 0$, so the company is just willing to offer insurance at this level.

But will the farmer make the stipulated effort? In other words, is the contract incentive compatible? It is if the farmer's expected utility with the insurance and the effort exceed his expected utility of accepting the insurance but not putting out the effort. That is, the contract must satisfy the following inequality[13]:

$$0.75 \times \sqrt{150{,}000} + 0.25 \times \sqrt{70{,}000} - 25 > 0.50 \times \sqrt{150{,}000} + 0.50 \times \sqrt{70{,}000}.$$

Calculating out the values of the two expressions yields (approximately) $331 > 326$, which is true. So the partial insurance contract is incentive compatible; it will induce the appropriate bad-outcome-reducing effort on the part of the farmer.

And does the contract satisfy the participation constraint as well? Yes. It must provide the farmer with an expected utility at least as large as he could achieve in the absence of insurance. That level, which we calculated earlier, is 325; here he gets 331. The farmer is better off with this partial insurance contract than with no insurance at all, and both parties will agree to this contract.

Evidence supporting this theory of insurance and moral hazard can be found in any of your insurance contracts. Most policies come with various requirements of deductibles and copayments that leave some of the policyholder's risk uninsured in order to reduce moral hazard.

[13] Compare the first term on the left-hand side with the first term on the right-hand side of the incentive-compatibility constraint. Exerting the effort to construct high-quality basins raises the coefficient multiplying the high utility, $\sqrt{150{,}000}$, from 0.50 to 0.75. Similarly, comparing the second terms on either side, you will see that failure to make the effort raises the coefficient multiplying the low utility, $\sqrt{70{,}000}$, from 0.25 to 0.50. These differences are analogous to the carrot and stick aspects of the incentive scheme in Section 5.A above.

# 6 INCENTIVES FOR EFFORT: EVIDENCE AND EXTENSIONS

The theme of the managerial-effort-incentive scheme of Section 5.A was the trade-off between giving the manager a more powerful incentive to provide the optimal effort level and requiring him to bear more of the risk in the firm's profit. This trade-off is an important consideration in practice, but it must be considered in combination with other features of the relationship between the firm and its employee. Most of these other features have to do with multiple dimensions of the activities that go on within the firm. The quality and quantity of effort are not just a matter of good or bad, and outcomes are not just a matter of success or failure; each can range over many possibilities, and such entities as hours and profits can vary continuously. The firm has many employees, and the overall outcome for the firm depends on some combination of their actions. Most firms have multiple outputs, and each employee performs multiple tasks. And the firm and its employees interact over a long period of time, not just for one project or over a short duration. All of these features correspondingly require more complex incentive schemes. In this section, we outline a few of these and refer you to a rich body of literature for further details.[14] The mathematics of these schemes gets correspondingly complex, so we will merely give you the intuition behind them and leave formal rigorous analyses to more advanced courses.

## A. Nonlinear Incentive Schemes

Can the optimal managerial effort scheme always be characterized by a base salary with a profit-share component? No. If there are three possible outcomes—failure, modest success, and huge success—then the percentage bonus for going from failure to modest success may not equal that for going from modest to huge success. So the optimal scheme may be nonlinear.

Suppose we alter the managerial supervision example in Section 5.A to allow for three possible outcomes: profit over material and wage cost of 0, $500,000, or $1 million. Suppose also that good supervisory effort yields probabilities of success

[14] Canice Prendergast, "The Provision of Incentives in Firms," *Journal of Economic Literature*, vol. 37, no. 1 (March 1999), pp. 7–63, is an excellent survey of the theory and practice of incentive mechanisms. Prendergast gives references to the original research literature from which many findings and anecdotes are mentioned in this section, so we will not repeat the specific citations. James N. Baron and David M. Kreps, *Strategic Human Resources: Frameworks for General Managers* (New York: Wiley, 1999), is a wider-ranging book on personnel management, combining perspectives from economics, sociology, and social psychology; chapters 8, 11, and 16 and appendixes C and D are closest to the concerns of this chapter and this book.

of one-sixth, one-third, and one-half for the three possible outcomes in the same order. Poor supervision reverses the probabilities of success to one-half, one-third, and one-sixth, respectively. Then a somewhat harder calculation along the same lines as above, which we relegate to an optional exercise, shows that the optimal payments are $30,625 for failure, $160,000 for the modest success, and $225,625 for the top outcome. If we interpret this payment scheme as a $30,625 base salary with a bonus for success, then the bonus is $129,375 for achieving the $500,000 profit and $195,000 for achieving the $1 million profit. The bonus represents a 26% share of profits for the first level of success but only a 13% share for the second level.

Special forms of nonlinear schemes are often used in practice. The most common of such schemes incorporates a stipulated, fixed bonus that is paid if a certain performance standard or quota is achieved. When might such a scheme be desirable?

A quota-bonus scheme constitutes a powerful incentive if it can be set at such a level that an increase in the worker's effort substantially increases the probability of meeting the quota. To illustrate such a case, consider a firm that wants each salesman to produce $1 million in sales, and it is willing to pay up to $100,000 for this level of performance. If it pays a flat 10% commission, the salesman's incremental effort in pushing sales from $900,000 to $1 million will bring him $10,000. But if the firm offers a wage of $60,000 and a bonus of $40,000 for meeting the quota of $1 million, then this last bit of effort pushes the salesman up to his quota and earns him an extra $40,000. Thus, the quota gives the salesman a much stronger incentive to make the incremental effort.

But the quota-bonus scheme is not without its drawbacks. The level at which the quota is set must be judged quite precisely. Suppose the firm misjudges and sets the quota at $1.2 million, and the salesman knows that the probability of reaching that level of sales, even with superhuman effort, is quite small. The salesman may then give up, make very little effort, and settle for earning just the base salary. The salesman's resulting sales may fall far short of even $1 million. Conversely, the pure quota-bonus scheme gives him no incentive to go beyond the $1 million level. Finally, the quota must be applied over a specific period, usually the calendar year. This requirement produces even more perverse incentives. A salesman who has bad luck in the first few months of a year will realize that he has no chance of making his quota that year, so he will take things easy for the rest of the year. If in contrast he has very good luck and meets the quota by July, again he has no incentive to exert himself for the rest of the year. And he may be able to manipulate the scheme by conspiring with his customers to shift sales from one year to another to improve his chances of making the quota in both years. A linear scheme like the one with profit sharing described above is less open to such manipulation.

Therefore, firms usually combine a quota scheme with a more graduated piecewise linear-payment scheme. For example, the salesman may get a base salary, a low rate of commission for sales between $500,000 and $1 million, a higher rate of commission for sales between $1 million and $2 million, and so on.

Managers of mutual funds, for example, are rewarded for good performance over a calendar year. These rewards come from their firm in the form of bonuses but also from the public when they invest more in those specific funds. If these reward schemes are nonlinear, the managers respond by changing the risk profile of their funds' portfolios. We saw in the appendix to Chapter 8 that a person with a concave utility function is risk averse and one with a convex utility function is a risk lover. In the same way that a risk-loving individual prefers risky situations to safe ones, a manager facing a convex reward scheme will take excessive risk with his fund's portfolio.

## B. Incentives in Teams

Rarely do the employees of a firm act as individuals on separate tasks. Salesmen working in distinct assigned regions come closest to being so separate, although even in that case the performance of an individual salesman is affected by the support of others in the office. Usually people work in teams, and the outcome for the team and for each member depends on the efforts of all. A firm's profit as a whole, for example, depends on the performance of all of its workers and managers. This interaction creates special problems for the design of incentives.

When one worker's earnings depend on the profit of the firm as whole, each worker will see only a weak link between his effort and the aggregate profit, and each will have only a small fractional share in that aggregate profit. This share is a very weak incentive for other workers to exert effort. Even in a smaller team, each member will be tempted to shirk and become a free rider on the efforts of the others. This outcome mirrors the prisoners' dilemma of collective action we saw in the street-garden example of Chapters 3 and 4, and throughout Chapter 10. If the team is small and stable over a sufficiently long time, we can expect its members to resolve the dilemma by devising internal and perhaps nonmonetary schemes of rewards and punishments like the ones we saw in Chapter 10, Section 3.

In another context, the existence of many workers on a team can sharpen incentives. Suppose a firm has many workers performing similar tasks, perhaps selling different components from the firm's product line. If there is a common (positively correlated) random component to each worker's sales, perhaps based on the strength of the underlying economy, then the sales of one worker relative to those of another worker are a good indicator of their relative effort levels. For example, the efforts of workers 1 and 2, denoted by $x_1$ and $x_2$, might be related to their sales, $y_1$ and $y_2$, according to the formulas $y_1 = x_1 + r$ and $y_2 = x_2 + r$, where $r$ represents the common random error in sales (or the common "luck factor," to use the terminology of Chapter 8, Section 1.C). In this case, it follows that $y_2 - y_1 = x_2 - x_1$ with no randomness; that is, the difference in observed sales will exactly equal the difference in exerted effort across workers 1 and 2.

The firm employing these workers can then reward them according to their relative outcomes. This payment scheme entails no risk for the workers. The trade-off we considered in Section 5, between providing optimal effort and sharing in the profits of the firm, vanishes. Now, if the first worker has a poor sales record and tries to blame it on bad luck, the firm can respond, "Then how come this other worker achieved so much more? Luck was common to the two of you, so you must have made less effort." Of course, if the two workers can collude, they can defeat the firm's purpose, but otherwise the firm can implement a powerful incentive scheme by setting workers in competition with each other. An extreme example of such a scheme is a tournament in which the best performer gets a prize.

Tournaments also help mitigate another potential moral-hazard problem. In reality, the criteria of success are themselves not easily or publicly observable. Then the owner of the firm may be tempted to claim that no one has performed well enough and that no one should be paid a bonus. A tournament with a prize that must be awarded to someone or a given aggregate bonus pool that must be distributed among the workers eliminates this moral hazard on the part of the principal.

## C. Multiple Tasks and Outcomes

Employees usually perform several tasks for their employers. These various tasks lead to several measurable outcomes of employee effort. Incentives for providing effort to the different tasks then interact. And this interaction makes mechanism design more complex for the firm.

The outcome of each of an agent's tasks depends partly on the agent's effort and partly on chance. That is why an outcome-based incentive scheme generally inflicts some risk on the agent's payoff. If the chance element is small, then the risk to the agent is small and the incentive to exert effort can be made more powerful. Of course, the outcomes of different tasks are likely to be affected by chance to different extents. So if the principal considers the tasks one at a time, he will use powerful incentives for effort on the tasks that have smaller elements of chance and weaker incentives for effort on the tasks where outcomes are more uncertain indicators of the agent's effort. But the powerful incentive on one task will divert the agent's effort away from the other task, further weakening the performance on that task. To avoid this substitution of effort toward the task with the powerful incentive, the principal has to weaken the incentive on that task, too.

An example of this can be found in our own lives. Professors are supposed to do research as well as teaching. There are many accurate indicators of good research: publications in and appointments to editorial positions for prestigious journals, elections to scientific academies, and so on. By contrast, good teaching can only be observed less accurately and with long lags. Students often need years of experience to recognize the value of what they learned in college; in the short term, they may be more impressed by showmanship than by scholarship. If these two tasks required of faculty members were considered in isolation, university administrators would attach powerful incentives to research and weaker incentives to teaching. But if they did so, professors would divert their efforts away from teaching and toward research (even more than they already do in some institutions). Therefore, the imprecise observation of teaching outcomes forces deans and presidents to offer only weak incentives for research as well.

The most cited example of a situation with multiple tasks and outcomes occurs in school teaching. Some outcomes of teaching, such as test scores, are precisely observable, whereas other valuable aspects of education, such as ability to work in teams or speak in public, are less accurately measurable. If teachers are rewarded on the basis of their students' test scores, they will "teach to the test," and the other dimensions of their students' education will get ignored. Such "gaming" of an incentive scheme also extends to sports. If a baseball hitter is rewarded for hitting home runs, he will neglect other aspects of batting (taking pitches, sacrifice bunts, etc.) that can sometimes be better for his team's chances of winning a game. Similarly, salesmen may sacrifice long-term customer relationships in favor of driving home a sale to meet a short-term sales goal.

If this problem of dysfunctional effects of some incentives on other tasks is too severe, other systems of rewarding tasks may be needed. A more holistic but subjective measure of performance, for example the boss's overall evaluation, may be used. This alternative is not without its own problems; workers may then divert their effort into activities that find favor with the boss!

## D. Incentives over Time

Many employment relationships last for a long time, and that opens up opportunities for the firm to devise incentive schemes where performance at one time is rewarded at a later time. Firms regularly use promotions, seniority-based salaries, and other forms of deferred compensation. In effect, workers are underpaid relative to their performance in the earlier stages of their careers with the firm and overpaid in later years. The prospect of future rewards motivates younger workers to exert good effort and also induces them to stay with the firm, thus reducing job turnover. Of course, the firm may be tempted to renege on its implicit promise of overpayment in later years; therefore such schemes must be credible if they are to be effective. They are more likely to be used effectively in firms that have a long record of stability and a reputation for treating their senior workers well.

A different way that the prospect of future compensation can keep workers motivated is through the use of an "efficiency wage." The firm pays a worker more than the going wage, and the excess is a surplus, or economic rent, for the worker. So long as the worker makes good effort, he will go on earning this surplus. But if he shirks, he may be detected, at which point he will be fired and will have to go back to the general labor market, where he can earn only the going wage.

The firm faces a problem in mechanism design when it tries to determine the appropriate efficiency wage level. Suppose the going wage is $w_0$, and the firm's efficiency wage is $w > w_0$. Let the monetary equivalent of the worker's subjective cost of making good effort be $e$. Each pay period the worker has the choice of whether to make this effort. If the worker shirks, he saves $e$. But with probability $p$, the shirking will be detected. If it is discovered that he has been shirking, the worker will lose the surplus $(w - w_0)$, starting in the next pay period and continuing indefinitely. Let $r$ be the rate of interest from one period to the next. Then if the worker shirks today, the expected discounted present value of the worker's loss in the next pay period is $p\,(w - w_0)/(1 + r)$. And the worker loses $w - w_0$ with probability $p$ in all future pay periods. A calculation similar to the ones we performed for repeated games in Chapter 10 and its appendix shows that the total expected discounted present value of the future loss to the worker is

$$p\left[\frac{w - w_0}{1 + r} + \frac{w - w_0}{(1 + r)^2} + \cdots\right] = p(w - w_0)\,\frac{1/(1 + r)}{1 - 1/(1 + r)} = \frac{p(w - w_0)}{r}.$$

To deter shirking, the firm needs to make sure that this expected loss is at least as high as the worker's immediate gain from shirking, $e$. Therefore, the firm must pay an efficiency wage that satisfies:

$$\frac{p(w - w_0)}{r} \geq e \quad \text{or} \quad w - w_0 \geq \frac{er}{p} \quad \text{or} \quad w \geq w_0 + \frac{er}{p}.$$

The smallest efficiency wage is the one that makes this expression hold with equality. And the more accurately the firm can detect shirking (that is, the higher is $p$), the smaller its excess over the going wage needs to be.

A repeated relationship may also enable the firm to design a sharper incentive scheme in another way. In any one period, as we explained above, the worker's observed outcome is a combination of the worker's effort and an element of chance. But if the outcome is poor year after year, the worker cannot credibly blame bad luck year after year. Therefore, the average outcome over a long period can, by the law of large numbers, be used as a more accurate measure of the worker's average effort, and the worker can be rewarded or punished accordingly.

## SUMMARY

The study of *mechanism design* can be summed up as learning "how to deal with someone who knows more than you do." Such situations occur in numerous contexts, usually in interactions involving a more-informed player, called the *agent*, and a less-informed player, called the *principal*, who wants to design a mechanism to give the agent the correct incentives to help the principal attain his goal.

Mechanism-design problems are of two types. The first type involves information revelation, in which the principal creates a scheme to screen information from the agent. The second type involves moral hazard, in which the principal creates a scheme to elicit the optimal level of an observable action by the agent. In all cases, the principal attempts to maximize its own objective function subject to the incentive compatibility and participation constraints of the agent.

Firms use information-revelation schemes in creating pricing structures that separate customers by their willingness to pay for the firm's product. Procurement contracts are also often designed to separate projects, or contractors, according to various levels of cost. Evidence of both price discrimination and screening with procurement contracts can be seen in actual markets.

When facing moral hazard, employers must devise contracts that encourage their employees to provide optimal effort. Similarly, insurance companies must write policies that give their clients the right incentives to protect against the insured bad outcome's occurring. In some simple situations, optimal contracts will be linear schemes, but in the presence of more complex relationships, nonlinear schemes may be more beneficial. Incentive systems designed for workers in teams, or when relationships continue over time, are correspondingly more complex than those written for simpler situations.

## KEY TERMS

**agent** (521)
**mechanism design** (515)
**price discrimination** (516)
**principal** (521)
**principal-agent (agency) problem** (521)

## SOLVED EXERCISES

**S1.** Firms that provide insurance to clients to protect them from the costs associated with theft or accident must necessarily be interested in the behavior of their policyholders. Sketch some ideas for the creation of an incentive scheme that such a firm might use to deter and detect fraud or lack of care on the part of its policyholders.

**S2.** Some firms sell goods and services either singly or in bundles in order to increase their own profit by separating consumers with different preferences.

(a) List three examples of quantity discounts offered by firms.

(b) How do quantity discounts allow firms to screen consumers by their preferences?

**S3.** Omniscient Wireless Limited (OWL) is planning to roll out a new nationwide, broadband, wireless telephone service next month. The firm has conducted market research indicating that its 10 million potential consumers are in two segments, which they call the Light segment and the Regular segment. Light users have less demand for wireless-phone service and in particular they seem unlikely to have any value for more than 300 minutes of calls per month. Regular users have more demand for mobile-phone service generally and have high value for more than 300 minutes per month. OWL analysts have determined that the best plans to offer to consumers entail 300 minutes per month and 600 minutes per month, respectively. They estimate that 50% of users are Light and 50% are Regular, and that each type has the following willingness to pay for each type of service:

| | 300 minutes | 600 minutes |
|---|---|---|
| Light user (50%) | \$20 | \$30 |
| Regular user (50%) | \$25 | \$70 |

OWL's cost per additional minute of wireless service is negligible, so the subscription cost to the company is \$10 per user, no matter which plan the user chooses.

Each potential customer calculates the net payoff (benefit *minus* price) that she would get from each of the usage plans and buys the plan that would give the higher net payoff, so long as this payoff is not negative. If both plans give equal, nonnegative net payoffs for a buyer, she goes for 600 minutes; if both plans have negative net payoffs for a buyer, she does not purchase. OWL wants to maximize its expected profit per potential customer.

(a) Suppose the firm were to offer only the 300-minute plan, but not the 600-minute plan. What would be the optimal price to charge, and what would be the average profit per potential customer?

(b) Suppose instead that the firm were to offer only the 600-minute plan. What would be the optimal price, and what would be the average profit per potential customer?

(c) Suppose the firm wanted to offer both plans. Suppose further that it wanted the Light users to purchase the 300-minute plan and the Regular users to purchase the 600-minute plan. Write down the incentive-compatibility constraint for the Light user.

(d) Similarly, write down the incentive-compatibility constraint for the Regular user.

(e) Use the results from parts (c) and (d) to calculate the optimal pair of prices to charge for the 300-minute and 600-minute services, so that each user type will purchase its intended service plan. What would be the average profit per potential customer?

(f) Consider the outcomes described in parts (a), (b), and (e). For each of the three situations, describe whether it is a separating outcome, a pooling outcome, or a semiseparating outcome.

**S4.** Mictel Corporation has a world monopoly on the production of personal computers. It can make two kinds of computers: low end and high end. One-fifth of the potential buyers are casual users, and the rest are intensive users.

The costs of production of the two kinds of machines as well as the benefits gained from the two, by the two types of prospective buyers, are given in the following table (all figures are in thousands of dollars):

| | | COST | BENEFIT FOR USER TYPE | |
|---|---|---|---|---|
| | | | Casual | Intensive |
| PC TYPE | Low-end | 1 | 4 | 5 |
| | High-end | 3 | 5 | 8 |

Each type of buyer calculates the net payoff (benefit *minus* price) that he would get from each type of machine and buys the type that would give the higher net payoff, provided that this payoff is nonnegative. If both types give equal, nonnegative net payoffs for a buyer, he goes for the high end; if both types have negative net payoff for a buyer, he does not purchase.

Mictel wants to maximize its expected profit.

(a) If Mictel were omniscient, then, when a prospective customer came along, knowing his type, the company could offer to sell him just one type of machine at a stated price, on a take-it-or-leave-it basis. What machine would Mictel offer, and at what price, to what buyer?

In fact, Mictel does not know the type of any particular buyer. It just makes its catalog available for all buyers to choose from.

(b) First, suppose the company produces just the low-end machines and sells them for price $x$. What value of $x$ will maximize its profit? Why?

(c) Next, suppose Mictel produces just the high-end machines and sells them for price $y$. What value of $y$ will maximize its profit? Why?

(d) Finally, suppose the company produces both types of machines, selling the low-end ones for price $x$ and the high-end ones for price $y$. What incentive-compatibility constraints on $x$ and $y$ must the company satisfy if it wants the casual users to buy the low-end machines and the intensive users to buy the high-end machines?

(e) What participation constraints must $x$ and $y$ satisfy for the casual users to be willing to buy the low-end machines and for the intensive users to be willing to buy the high-end machines?

(f) Given the constraints in parts (d) and (e), what values of $x$ and $y$ will maximize the expected profit when the company sells both types of machines? What is the company's expected profit from this policy?

(g) Putting it all together, decide what production and pricing policy the company should pursue.

**S5.** Redo Exercise S4, assuming that one-half of Mictel's customers are casual users.

**S6.** Using the insights gained in Exercises S4 and S5, solve Exercise S4 for the general case in which the proportion of casual users is $c$ and the proportion of intensive users is $(1 - c)$. The answers to some parts will depend on the value of $c$. In these instances, list all relevant cases and how they depend on $c$.

**S7.** Sticky Shoe, the discount movie theater, sells popcorn and soda at its concession counter. Cameron, Jessie, and Sean are regular patrons of Sticky Shoe, and the valuations of each for popcorn and soda are as follows:

| | Popcorn | Soda |
|---|---|---|
| Cameron | \$3.50 | \$3.00 |
| Jessica | \$4.00 | \$2.50 |
| Sean | \$1.50 | \$3.50 |

There are 2,997 other residents of Harkinsville who see movies at Sticky Shoe. One-third of them have valuations identical to Cameron, one-third to Jessica, and one-third to Sean. If a customer is indifferent between buying and not, she buys. It costs Sticky Shoe essentially nothing to produce each additional order of popcorn or soda.

**(a)** If Sticky Shoe sets separate prices for popcorn and soda, what price should it set for each concession to maximize its profit? How much profit does Sticky Shoe make selling concessions separately?

**(b)** What does each type of customer (Cameron, Jessica, Sean) buy when Sticky Shoe sets separate profit-maximizing prices for popcorn and soda?

**(c)** Instead of selling the concessions separately, Sticky Shoe decides always to sell the popcorn and soda together in a combo, charging a single price for both. What single combo price would maximize its profit? How much profit does Sticky Shoe make selling only combos?

**(d)** What does each type of customer buy when Sticky Shoe sets a single profit-maximizing price for a popcorn and soda combo? How does this compare with the answer in part (b)?

**(e)** Which pricing scheme does each customer type prefer? Why?

**(f)** If Sticky Shoe sold the concessions both as a combo and separately, which products (popcorn, soda, or the combo) does it want to sell to each customer type? How can Sticky Shoe make sure that each customer type purchases exactly the product that it intends for him or her to purchase?

**(g)** What prices—for the popcorn, soda, and combo—would Sticky Shoe set to maximize its profit? How much profit does Sticky Shoe make selling the concessions at these three prices?

**(h)** How do the answers to parts (a), (c), and (g) differ? Explain why.

**S8.** Section 5.A of this chapter discusses the principal–agent problem in the context of a company deciding whether and how to induce a manager to put in high effort to increase the chances that the project succeeds. The value of a successful project is $1 million; the probability of success given high effort is 0.5; the probability of success given low effort is 0.25. The manager's utility is the square root of compensation (measured in millions of dollars), and his disutility from exerting high effort is 0.1. However, the reservation wage of the manager is now $160,000.

**(a)** What contract does the company offer if it wants only low effort from the manager?

**(b)** What is the expected profit to the company when it induces low managerial effort?

(c) What contract pair $(y, x)$—where $y$ is the salary given for a successful project and $x$ is the salary given for a failed project—should the company offer the manager to induce high effort?

(d) What is the company's expected profit when it induces high effort?

(e) Which level of effort does the company want to induce from its manager? Why?

**S9.** A company has purchased fire insurance for its main factory. The probability of a fire in the factory without a fire-prevention program is 0.01. The probability of a fire in a factory with a fire-protection program is 0.001. If a fire occurred, the value of the loss would be \$300,000. A fire-prevention program would cost \$80 to run, but the insurance company cannot costlessly observe whether or not the prevention program has been implemented.

(a) Why does moral hazard arise in this situation? What is its source?

(b) Can the insurance company eliminate the moral hazard problem? If so, how? If not, explain why not.

**S10.** Mozart moved from Salzburg to Vienna in 1781, hoping for a position at the Habsburg court. Instead of applying for a position, he waited for the emperor to call him, because "if one makes any move oneself, one receives less pay." Discuss this situation using the theory of games with asymmetric information, including theories of signaling and screening.

**S11.** **(Optional, requires calculus)** You are Oceania's Minister for Peace, and it is your job to purchase war materials for your country. The net benefit, measured in Oceanic dollars, from quantity $Q$ of these materials is $2Q^{1/2} - M$, where $M$ is the amount of money paid for the materials.

There is just one supplier—Baron Myerson's Armaments (BMA). You do not know BMA's cost of production. Everyone knows that BMA's cost per unit of output is constant, and that it is equal to 0.10 with probability $p = 0.4$ and equal to 0.16 with probability $1 - p$. Call BMA "low cost" if its unit cost is 0.10 and "high cost" if it is 0.16. Only BMA knows its true cost type with certainty.

In the past, your ministry has used two kinds of purchase contracts: cost plus and fixed price. But cost-plus contracts create an incentive for BMA to overstate its costs, and fixed-price contracts may compensate the firm more than is necessary. You decide to offer a menu of two possibilities:

Contract 1: Supply us quantity $Q_1$, and we will pay you money $M_1$.
Contract 2: Supply us quantity $Q_2$, and we will pay you money $M_2$.

The idea is to set $Q_1$, $M_1$, $Q_2$, and $M_2$ such that a low-cost BMA will find contract 1 more profitable, and a high-cost BMA will find contract 2 more profitable. If another contract is exactly as profitable, a low-cost BMA will choose contract 1, and a high-cost BMA will choose contract 2. Further, regardless of its cost, BMA will need to receive at least zero economic profit in any contract it accepts.

**(a)** Write expressions for the profit of a low-cost BMA and a high-cost BMA when it supplies quantity $Q$ and is paid $M$.

**(b)** Write the incentive-compatibility constraints to induce a low-cost BMA to select contract 1 and a high-cost BMA to select contract 2.

**(c)** Give the participation constraints for each type of BMA.

**(d)** Assuming that each of the BMA types chooses the contract designed for it, write the expression for Oceania's expected net benefit.

Now your problem is to choose $Q_1$, $M_1$, $Q_2$, and $M_2$ to maximize the expected net benefit found in part (d) subject to the incentive-compatibility (IC) and participation constraints (PC).

**(e)** Assume that $Q_1 > Q_2$, and further assume that constraints $IC_1$ and $PC_2$ bind—that is, they will hold with equalities instead of weak inequalities. Use these constraints to derive lower bounds on your feasible choices of $M_1$ and $M_2$ in terms of $Q_1$ and $Q_2$.

**(f)** Show that when $IC_1$ and $PC_2$ bind, $IC_2$ and $PC_1$ are automatically satisfied.

**(g)** Substitute out for $M_1$ and $M_2$, using the expressions found in part (e) to express your objective function in terms of $Q_1$ and $Q_2$.

**(h)** Write the first-order conditions for the maximization, and solve them for $Q_1$ and $Q_2$.

**(i)** Solve for $M_1$ and $M_2$.

**(j)** What is Oceania's expected net benefit from offering this menu of contracts?

**(k)** What general principles of screening are illustrated in the menu of contracts you found?

**S12.** **(Optional)** Revisit Oceania's problem in Exercise S11 to see how the optimal menu found in that problem compares with some alternative contracts.

**(a)** If you decided to offer a single fixed-price contract that was intended to attract only the low-cost BMA, what would it be? That is, what single $(Q, M)$ pair would be optimal if you knew BMA was low cost?

**(b)** Would a high-cost BMA want to accept the contract offered in part (a)? Why or why not?

**(c)** Given the probability that BMA is low cost, what would the expected net benefit to Oceania be from offering the contract in part (a)? How does this compare with the expected net benefit from offering a menu of contracts, as found in part (j) of Exercise S11?

**(d)** What single fixed-price contract would you offer to a high-cost BMA?

**(e)** Would a low-cost BMA want to accept the contract found in part (d)? What would its profit be if it did?

**(f)** Given your answer in part (e), what would be the expected net benefit to Oceania from offering the contract in part (d)? How does this compare with the expected net benefit from offering a menu of contracts, found in part (j) of Exercise S11?

**(g)** Consider the case in which an industrial spy within BMA has promised to divulge the true per-unit cost, so that Oceania could offer the optimal single, fixed-price contract geared toward BMA's true type. What would Oceania's expected net benefit be if it knew that it was going to learn BMA's true type? How does this compare with parts (c) and (f) of this exercise and with part (j) of Exercise S11?

## UNSOLVED EXERCISES

**U1.** What problems of moral hazard and/or adverse selection arise in your dealings with each of the following? In each case, outline some appropriate incentive schemes and/or signaling and screening strategies to cope with these problems. No mathematical analysis is expected, but you should state clearly the economic reasoning of why and how your suggested methods work.

**(a)** Your financial adviser tells you what stocks to buy or sell.

**(b)** You consult a realtor when you are selling your house.

**(c)** You visit your doctor, whether for routine check-ups or treatments.

**U2.** MicroStuff is a software company that sells two popular applications, WordStuff and ExcelStuff. It doesn't cost anything for MicroStuff to make each additional copy of its applications. MicroStuff has three types of potential customers, represented by Ingrid, Javiera, and Kathy. There are 100 million potential customers of each type, whose valuations for each application are as follows:

| | WordStuff | ExcelStuff |
|---|---|---|
| Ingrid | 100 | 20 |
| Javiera | 30 | 100 |
| Kathy | 80 | 0 |

(a) If MicroStuff sets separate prices for WordStuff and ExcelStuff, what price should it set for each application to maximize its profit? How much profit does MicroStuff earn with these prices?

(b) What does each type of customer (Ingrid, Javiera, Kathy) buy when MicroStuff sets profit-maximizing, separate prices for WordStuff and ExcelStuff?

(c) Instead of selling the applications separately, MicroStuff decides always to sell WordStuff and ExcelStuff together in a bundle, charging a single price for both. What single price for the bundle would maximize its profit? How much profit does MicroStuff make selling its software only in bundles?

(d) What does each type of customer buy when MicroStuff sets a single, profit-maximizing price for a bundle of WordStuff and ExcelStuff? How does this compare with the answer in part (b)?

(e) Which pricing scheme does each customer type prefer? Why?

(f) If MicroStuff sold the applications both as a bundle and separately, which products (WordStuff, ExcelStuff, or the bundle) would it want to sell to each customer type? How can MicroStuff make sure that each customer type purchases exactly the product that it intends for them to purchase?

(g) What prices—for WordStuff, ExcelStuff, and the bundle—would MicroStuff set to maximize its profit? How much profit does MicroStuff make selling the products with these three prices?

(h) How do the answers to parts (a), (c), and (g) differ? Explain why.

**U3.** Consider a managerial effort example similar to the one in Section 5. The value of a successful project is \$420,000; the probabilities of success are 1/2 with good supervision and 1/4 without. The manager is risk neutral, not risk averse as in the text, so his expected utility equals his expected income minus his disutility of effort. He can get other jobs paying \$90,000, and his disutility for exerting the extra effort for good supervision on your project is \$100,000.

(a) Show that inducing high effort would require the firm to offer a compensation scheme with a negative base salary; that is, if the project fails, the manager pays the firm an amount stipulated in the scheme.

(b) How might a negative base salary be implemented in reality?

(c) Show that if a negative base salary is not feasible, then the firm does better to settle for the low-pay, low-effort situation.

**U4.** Cheapskates is a very minor-league professional hockey team. Its facilities are large enough to accommodate all of the 1,000 fans who might want to watch its home games. It can provide two types of seats—ordinary and luxury. There are also two sorts of fans: 60% of the fans are blue-collar fans, and the rest are white-collar fans. The costs of providing each type of seat and the fans' willingness to pay for each type of seat are given in the following table (measured in dollars):

| | | COST | Willingness to Pay | |
|---|---|---|---|---|
| | | | Blue-Collar | White-Collar |
| SEAT TYPE | Ordinary | 4 | 12 | 14 |
| | Luxury | 8 | 15 | 22 |

Each fan will buy at most one seat, depending on the consumer surplus he would get (maximum willingness to pay minus the actual price paid) from the two kinds. If the surplus for both is negative, then he won't buy any. If at least one kind gives him nonnegative surplus, then he will buy the kind that gives him the larger surplus. If the two kinds give him equal, nonnegative surplus, then the blue-collar fan will buy the ordinary kind of seat, and the white-collar fan will buy the luxury kind.

The team owners provide and price their seating to maximize profit, measured in thousands of dollars per game. They set prices for each kind of seat, sell as many tickets as are demanded at these prices, and then provide the numbers and types of seats of each kind for which the tickets have sold.

(a) First, suppose the team owners can identify the type of each individual fan who arrives at the ticket window (presumably by the color of his collar) and can offer him just one type of seat at a stated price, on a take-it-or-leave-it basis. What is the owners' maximum profit, $\pi^*$, under this system?

(b) Now, suppose that the owners cannot identify any individual fan, but they still know the proportion of blue-collar fans. Let the price of an ordinary seat be X and the price of a luxury seat be Y. What are the incentive-compatibility constraints that will ensure that the blue-collar fans buy the ordinary seats and the white-collar fans buy the luxury seats? Graph these constraints on an X-Y coordinate plane.

(c) What are the participation constraints for the fans' decisions on whether to buy tickets at all? Add these constraints to the graph in part (b).

(d) Given the constraints in parts (b) and (c), what prices X and Y maximize the owners' profit, $\pi_2$, under this price system? What is $\pi_2$?

(e) The owners are considering whether to set prices so that only the white-collar fans will buy tickets. What is their profit, $\pi_w$, if they decide to cater to only the white-collar fans?

(f) Comparing $\pi_2$ and $\pi_w$, determine the pricing policy that the owners will set. How does their profit achieved from this policy compare with the case of full information, where they earn $\pi^*$?

(g) What is the "cost of coping with the information asymmetry" in part (f)? Who bears this cost? Why?

**U5.** Redo Exercise U4 above, assuming that 10% of the fans are blue collar.

**U6.** Using the insights you gained in Exercises U4 and U5, solve Exercise U4 for the general case where a fraction $B$ of the fans is blue collar and fraction $(1 - B)$ is white collar. The answers to some parts will depend on the value of $B$. In these instances, list all relevant cases and how they depend on $B$.

**U7.** In many situations, agents exert effort in order to get promoted to a better-paid position, where the reward for that position is fixed and where agents compete among themselves for those positions. Tournament theory considers a group of agents competing for a fixed set of prizes. In this case, all that matters for winning is one's positions relative to others, rather than one's absolute level of performance.

(a) Discuss the reasons that a firm might wish to employ the tournament scheme described above. Consider the effects on the incentives of both the firm and the workers.

(b) Discuss the reasons that a firm might *not* wish to employ the tournament scheme described above.

(c) State one specific prediction of tournament theory and provide an example of empirical evidence in support of that prediction.

**U8.** Repeat Exercise S8 with the following adjustments: Due to the departure of some of their brightest engineers, the probability of success given a high managerial effort is only 0.4, and the probability of success given a low managerial effort is reduced to 0.24.

**U9.** **(Optional)** A teacher wants to find out how confident the students are about their own abilities. He proposes the following scheme: "After you answer this question, state your estimate of the probability that you are right. I will then check your answer to the question. Suppose you have given the probability estimate $x$. If your answer is actually correct, your grade will be $\log(x)$. If incorrect, it will be $\log(1 - x)$." Show that this scheme will elicit the students' own truthful estimates—that is, if the truth is $p$, show that a student's stated estimate $x = p$.

**U10.** **(Optional)** Redo Exercise S11, but assume that the probability that BMA is low cost is 0.6.

**U11.** **(Optional)** Repeat Exercise S11, but assume that a low-cost BMA has a per-unit cost of 0.2, and a high-cost BMA has a per-unit cost of 0.38. Let the probability that BMA is low cost be 0.4.

**U12.** **(Optional)** Revisit the situation in which Oceania is procuring arms from BMA. (See Exercise S11.) Now consider the case in which BMA has three possible cost types: $c_1$, $c_2$, and $c_3$, where $c_3 > c_2 > c_1$. BMA has cost $c_1$ with probability $p_1$, cost $c_2$ with probability $p_2$, and cost $c_3$ with probability $p_3$, where $p_1 + p_2 + p_3 = 1$. In what follows, we will say that BMA is of type *i* if its cost is $c_i$, for $i = 1, 2, 3$.

You offer a menu of three possibilities: "Supply us quantity $Q_i$, and we will pay you $M_i$," for $i = 1, 2$, and 3. Assume that more than one contract is equally profitable, so that a BMA of type *i* will choose contract *i*. To meet the participation constraint, contract *i* should give BMA of type *i* nonnegative profit.

(a) Write an expression for the profit of type-*i* BMA when it supplies quantity *Q* and is paid *M*.

(b) Give the participation constraints for each BMA type.

(c) Write the six incentive-compatibility constraints. That is, for each type *i* give separate expressions that state that the profit that BMA receives under contract *i* is greater than or equal to the profit it would receive under the other two contracts.

(d) Write down the expression for Oceania's expected net benefit, *B*. This is the objective function (what you want to maximize).

Now your problem is to choose the three $Q_i$ and the three $M_i$ to maximize expected net benefit, subject to the incentive-compatibility (IC) and participation constraints (PC).

(e) Begin with just three constraints: the IC constraint for type 2 to prefer contract 2 over contract 3, the IC constraint for type 1 to prefer contract 1 over contract 2, and the participation constraint for type 3. Assume that $Q_1 > Q_2 > Q_3$. Use these constraints to derive lower bounds on your feasible choices of $M_1$, $M_2$, $M_3$ in terms of $c_1$, $c_2$, and $c_3$ and $Q_1$, $Q_2$, and $Q_3$. (Note that two or more of the *c*s and *Q*s may appear in the expression for the lower bound for each of the *M*s.)

(f) Prove that these three constraints—the two ICs and one PC in part (e)—will be binding at the optimum.

(g) Now prove that when the three constraints in part (e) are binding, the other six constraints (the remaining four ICs and two PCs) are automatically satisfied.

**(h)** Substitute out for the $M_i$ to express your objective function in terms of the three $Q_i$ only.

**(i)** Write the first-order conditions for the maximization, and solve for each of the $Q_i$. That is, take the three partial derivatives $\partial Q_i / \partial B$, set them equal to zero, and solve for $Q_i$.

**(j)** Show that the assumption made above, $Q_1 > Q_2 > Q_3$, will be true at the optimum if:

$$\frac{c_3 - c_2}{c_2 - c_1} > \frac{p_1 p_3}{p_2}.$$