PART • THREE

Market Structure and Competitive Strategy

Part 3 examines a broad range of markets and explains how the pricing, investment, and output decisions of firms depend on market structure and the behavior of competitors.

Chapters 10 and 11 examine *market power*: the ability to affect price, either by a seller or a buyer. We will see how market power arises, how it differs across firms, how it affects the welfare of consumers and producers, and how it can be limited by government. We will also see how firms can design pricing and advertising strategies to take maximum advantage of their market power.

Chapters 12 and 13 deal with markets in which the number of firms is limited. We will examine a variety of such markets, ranging from *monopolistic competition*, in which many firms sell differentiated products, to a *cartel*, in which a group of firms coordinates decisions and acts as a monopolist. We are particularly concerned with markets in which there are only a few firms. In these cases, each firm must design its pricing, output, and investment strategies, while keeping in mind how competitors are likely to react. We will develop and apply principles from game theory to analyze such strategies.

Chapter 14 shows how markets for factor inputs, such as labor and raw materials, operate. We will examine the firm's input decisions and show how those decisions depend on the structure of the input market. Chapter 15 then focuses on capital investment decisions. We will see how a firm can value the future profits that it expects an investment to yield and then compare this value with the cost of the investment to determine whether the investment is worthwhile. We will also apply this idea to the decisions of individuals to purchase a car or household appliance, or to invest in education.

CHAPTERS

Market Power: Monopoly and Monopsony

10

In a perfectly competitive market, the large number of sellers and buyers of a good ensures that no single seller or buyer can affect its price. The market forces of supply and demand determine price. Individual firms take the market price as a given in deciding how much to produce and sell, and consumers take it as a given in deciding how much to buy.

Monopoly and *monopsony*, the subjects of this chapter, are the polar opposites of perfect competition. A **monopoly** is a market that has only one seller but many buyers. A **monopsony** is just the opposite: a market with many sellers but only one buyer. Monopoly and monopsony are closely related, which is why we cover them in the same chapter.

First we discuss the behavior of a monopolist. Because a monopolist is the sole producer of a product, the demand curve that it faces is the market demand curve. This market demand curve relates the price that the monopolist receives to the quantity it offers for sale. We will see how a monopolist can take advantage of its control over price and how the profit-maximizing price and quantity differ from what would prevail in a competitive market.

In general, the monopolist's quantity will be lower and its price higher than the competitive quantity and price. This imposes a cost on society because fewer consumers buy the product, and those who do pay more for it. This is why antitrust laws exist which forbid firms from monopolizing most markets. When economies of scale make monopoly desirable—for example, with local electric power companies—we will see how the government can increase efficiency by regulating the monopolist's price.

Pure monopoly is rare, but in many markets only a few firms compete with each other. The interactions of firms in such markets can be complicated and often involve aspects of *strategic gaming*, a topic covered in Chapters 12 and 13. In any case, the firms may be able to affect price and may find it profitable to charge a price higher than marginal cost. These firms have *monopoly power*. We will discuss the determinants of monopoly power, its measurement, and its implications for pricing.

Next we will turn to *monopsony*. Unlike a competitive buyer, a monopsonist pays a price that depends on the quantity that it purchases. The monopsonist's problem is to choose the quantity that maximizes its net benefit from the purchase—the value derived from the good less the money paid for it. By showing how the choice is made, we will demonstrate the close parallel between monopsony and monopoly.

CHAPTER OUTLINE

LIST OF EXAMPLES

• **monopoly** Market with only one seller.

• **monopsony** Market with only one buyer.

Although pure monopsony is also unusual, many markets have only a few buyers who can purchase the good for less than they would pay in a competitive market. These buyers have *monopsony power*. Typically, this situation occurs in markets for inputs to production. For example, General Motors, the largest U.S. car manufacturer, has monopsony power in the markets for tires, car batteries, and other parts. We will discuss the determinants of monopsony power, its measurement, and its implications for pricing.

• **market power** Ability of a seller or buyer to affect the price of a good.

Monopoly and monopsony power are two forms of **market power**: the ability—of either a seller or a buyer—to affect the price of a good.[1] Because sellers or buyers often have at least some market power (in most real-world markets), we need to understand how market power works and how it affects producers and consumers.

10.1 MONOPOLY

As the sole producer of a product, a monopolist is in a unique position. If the monopolist decides to raise the price of the product, it need not worry about competitors who, by charging lower prices, would capture a larger share of the market at the monopolist's expense. The monopolist *is* the market and completely controls the amount of output offered for sale.

But this does not mean that the monopolist can charge any price it wants—at least not if its objective is to maximize profit. This textbook is a case in point. Pearson Prentice Hall owns the copyright and is therefore a monopoly producer of this book. So why doesn't it sell the book for $500 a copy? Because few people would buy it, and Prentice Hall would earn a much lower profit.

To maximize profit, the monopolist must first determine its costs and the characteristics of market demand. Knowledge of demand and cost is crucial for a firm's economic decision making. Given this knowledge, the monopolist must then decide how much to produce and sell. The price per unit that the monopolist receives then follows directly from the market demand curve. Equivalently, the monopolist can determine price, and the quantity it will sell at that price follows from the market demand curve.

Average Revenue and Marginal Revenue

• **marginal revenue** Change in revenue resulting from a one-unit increase in output.

In §8.3, we explain that marginal revenue is a measure of how much revenue increases when output increases by one unit.

The monopolist's *average revenue*—the price it receives per unit sold—is precisely the market demand curve. To choose its profit-maximizing output level, the monopolist also needs to know its **marginal revenue**: the change in revenue that results from a unit change in output. To see the relationship among total, average, and marginal revenue, consider a firm facing the following demand curve:

$$P = 6 - Q$$

Table 10.1 shows the behavior of total, average, and marginal revenue for this demand curve. Note that revenue is zero when the price is $6: At that price, nothing is sold. At a price of $5, however, one unit is sold, so total (and marginal) revenue is $5. An increase in quantity sold from 1 to 2 increases revenue from $5 to $8; marginal revenue is thus $3. As quantity sold increases from 2 to 3, marginal revenue falls to $1, and when quantity increases from 3 to 4, marginal revenue

[1] The courts use the term "monopoly power" to mean significant and sustainable market power, sufficient to warrant particular scrutiny under the antitrust laws. In this book, however, for pedagogic reasons we use "monopoly power" differently, to mean market power on the part of sellers, whether substantial or not.

TABLE 10.1 Total, Marginal, and Average Revenue

Price (P)	Quantity (Q)	Total Revenue (R)	Marginal Revenue (MR)	Average Revenue (AR)
$6	0	$0	—	—
5	1	5	$5	$5
4	2	8	3	4
3	3	9	1	3
2	4	8	–1	2
1	5	5	–3	1

becomes negative. When marginal revenue is positive, revenue is increasing with quantity, but when marginal revenue is negative, revenue is decreasing.

When the demand curve is downward sloping, the price (average revenue) is greater than marginal revenue because all units are sold at the same price. If sales are to increase by 1 unit, the price must fall. In that case, all units sold, not just the additional unit, will earn less revenue. Note, for example, what happens in Table 10.1 when output is increased from 1 to 2 units and price is reduced to $4. Marginal revenue is $3: $4 (the revenue from the sale of the additional unit of output) less $1 (the loss of revenue from selling the first unit for $4 instead of $5). Thus, marginal revenue ($3) is less than price ($4).

Figure 10.1 plots average and marginal revenue for the data in Table 10.1. Our demand curve is a straight line and, in this case, the marginal revenue curve has twice the slope of the demand curve (and the same intercept).[2]

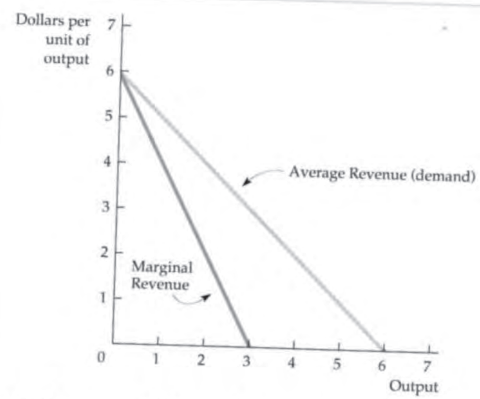

FIGURE 10.1 Average and Marginal Revenue

Average and marginal revenue are shown for the demand curve $P = 6 - Q$.

[2] If the demand curve is written so that price is a function of quantity, $P = a - bQ$, total revenue is given by $PQ = aQ - bQ^2$. Marginal revenue (using calculus) is $d(PQ)/dQ = a - 2bQ$. In this example, demand is $P = 6 - Q$ and marginal revenue is $MR = 6 - 2Q$. (This holds only for small changes in Q and therefore does not exactly match the data in Table 10.1.)

The Monopolist's Output Decision

In §7.1, we explain that marginal cost is the change in variable cost associated with a one-unit increase in output.

What quantity should the monopolist produce? In Chapter 8, we saw that to maximize profit, a firm must set output so that marginal revenue is equal to marginal cost. This is the solution to the monopolist's problem. In Figure 10.2, the market demand curve D is the monopolist's average revenue curve. It specifies the price per unit that the monopolist receives as a function of its output level. Also shown are the corresponding marginal revenue curve MR and the average and marginal cost curves, AC and MC. Marginal revenue and marginal cost are equal at quantity Q^*. Then from the demand curve, we find the price P^* that corresponds to this quantity Q^*.

How can we be sure that Q^* is the profit-maximizing quantity? Suppose the monopolist produces a smaller quantity Q_1 and receives the corresponding higher price P_1. As Figure 10.2 shows, marginal revenue would then exceed marginal cost. In that case, if the monopolist produced a little more than Q_1, it would receive extra profit (MR – MC) and thereby increase its total profit. In fact, the monopolist could keep increasing output, adding more to its total profit until output Q^*, at which point the incremental profit earned from producing one more

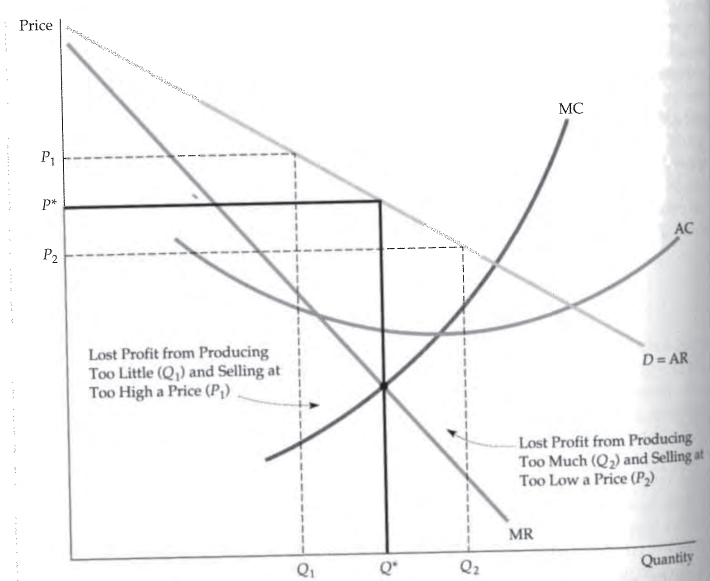

FIGURE 10.2 Profit Is Maximized When Marginal Revenue Equals Marginal Cost

Q^* is the output level at which MR = MC. If the firm produces a smaller output—say, Q_1—it sacrifices some profit because the extra revenue that could be earned from producing and selling the units between Q_1 and Q^* exceeds the cost of producing them. Similarly, expanding output from Q^* to Q_2 would reduce profit because the additional cost would exceed the additional revenue.

unit is zero. So the smaller quantity Q_1 is not profit maximizing, even though it allows the monopolist to charge a higher price. If the monopolist produced Q_1 instead of Q^*, its total profit would be smaller by an amount equal to the shaded area below the MR curve and above the MC curve, between Q_1 and Q^*.

In Figure 10.2, the larger quantity Q_2 is likewise not profit maximizing. At this quantity, marginal cost exceeds marginal revenue. Therefore, if the monopolist produced a little less than Q_2, it would increase its total profit (by MC – MR). It could increase its profit even more by reducing output all the way to Q^*. The increased profit achieved by producing Q^* instead of Q_2 is given by the area below the MC curve and above the MR curve, between Q^* and Q_2.

We can also see algebraically that Q^* maximizes profit. Profit π is the difference between revenue and cost, both of which depend on Q:

$$\pi(Q) = R(Q) - C(Q)$$

As Q is increased from zero, profit will increase until it reaches a maximum and then begin to decrease. Thus the profit-maximizing Q is such that the incremental profit resulting from a small increase in Q is just zero (i.e., $\Delta\pi/\Delta Q = 0$). Then

$$\Delta\pi/\Delta Q = \Delta R/\Delta Q - \Delta C/\Delta Q = 0$$

But $\Delta R/\Delta Q$ is marginal revenue and $\Delta C/\Delta Q$ is marginal cost. Thus the profit-maximizing condition is that MR – MC = 0, or MR = MC.

An Example

To grasp this result more clearly, let's look at an example. Suppose the cost of production is

$$C(Q) = 50 + Q^2$$

In other words, there is a fixed cost of \$50, and variable cost is Q^2. Suppose demand is given by

$$P(Q) = 40 - Q$$

By setting marginal revenue equal to marginal cost, you can verify that profit is maximized when $Q = 10$, an output level that corresponds to a price of \$30.[3]

Cost, revenue, and profit are plotted in Figure 10.3(a). When the firm produces little or no output, profit is negative because of the fixed cost. Profit increases as Q increases, reaching a maximum of \$150 at $Q^* = 10$, and then decreases as Q is increased further. At the point of maximum profit, the slopes of the revenue and cost curves are the same. (Note that the tangent lines rr' and cc' are parallel.) The slope of the revenue curve is $\Delta R/\Delta Q$, or marginal revenue, and the slope of the cost curve is $\Delta C/\Delta Q$, or marginal cost. Because profit is maximized when marginal revenue equals marginal cost, the slopes are equal.

Figure 10.3(b) shows both the corresponding average and marginal revenue curves and average and marginal cost curves. Marginal revenue and marginal cost intersect at $Q^* = 10$. At this quantity, average cost is \$15 per unit and price is \$30 per unit. Thus average profit is \$30 – \$15 = \$15 per unit. Because 10 units are sold, profit is (10)(\$15) = \$150, the area of the shaded rectangle.

[3] Note that average cost is $C(Q)/Q = 50/Q + Q$ and marginal cost is $\Delta C/\Delta Q = 2Q$. Revenue is $R(Q) = P(Q)Q = 40Q - Q^2$, so marginal revenue is $MR = \Delta R/\Delta Q = 40 - 2Q$. Setting marginal revenue equal to marginal cost gives $40 - 2Q = 2Q$, or $Q = 10$.

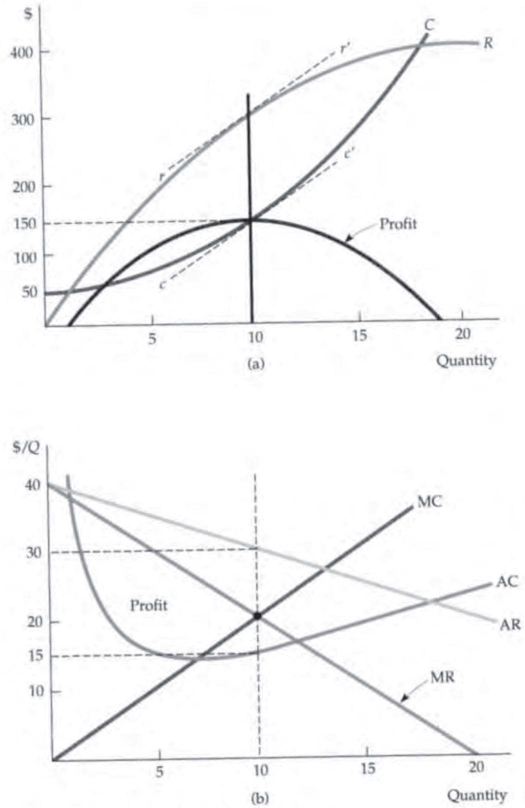

FIGURE 10.3 Example of Profit Maximization

Part **(a)** shows total revenue R, total cost C, and profit, the difference between the two. Part **(b)** shows average and marginal revenue and average and marginal cost. Marginal revenue is the slope of the total revenue curve, and marginal cost is the slope of the total cost curve. The profit-maximizing output is $Q^* = 10$, the point where marginal revenue equals marginal cost. At this output level, the slope of the profit curve is zero, and the slopes of the total revenue and total cost curves are equal. The profit per unit is \$15, the difference between average revenue and average cost. Because 10 units are produced, total profit is \$150.

A Rule of Thumb for Pricing

We know that price and output should be chosen so that marginal revenue equals marginal cost, but how can the manager of a firm find the correct price and output level in practice? Most managers have only limited knowledge of

the average and marginal revenue curves that their firms face. Similarly, they might know the firm's marginal cost only over a limited output range. We therefore want to translate the condition that marginal revenue should equal marginal cost into a rule of thumb that can be more easily applied in practice.

To do this, we first write the expression for marginal revenue:

$$\text{MR} = \frac{\Delta R}{\Delta Q} = \frac{\Delta(PQ)}{\Delta Q}$$

Note that the extra revenue from an incremental unit of quantity, $\Delta(PQ)/\Delta Q$, has two components:

1. Producing one extra unit and selling it at price P brings in revenue $(1)(P) = P$.
2. But because the firm faces a downward-sloping demand curve, producing and selling this extra unit also results in a small drop in price $\Delta P/\Delta Q$, which reduces the revenue from all units sold (i.e., a change in revenue $Q[\Delta P/\Delta Q]$).

Thus,

$$\text{MR} = P + Q\frac{\Delta P}{\Delta Q} = P + P\left(\frac{Q}{P}\right)\left(\frac{\Delta P}{\Delta Q}\right)$$

We obtained the expression on the right by taking the term $Q(\Delta P/\Delta Q)$ and multiplying and dividing it by P. Recall that the elasticity of demand is defined as $E_d = (P/Q)(\Delta Q/\Delta P)$. Thus $(Q/P)(\Delta P/\Delta Q)$ is the reciprocal of the elasticity of demand, $1/E_d$, measured at the profit-maximizing output, and

The elasticity of demand is discussed in §§2.4 and 4.3.

$$\text{MR} = P + P(1/E_d)$$

Now, because the firm's objective is to maximize profit, we can set marginal revenue equal to marginal cost:

$$P + P(1/E_d) = \text{MC}$$

which can be rearranged to give us

$$\frac{P - \text{MC}}{P} = -\frac{1}{E_d} \qquad \textbf{(10.1)}$$

This relationship provides a rule of thumb for pricing. The left-hand side, $(P - \text{MC})/P$, is the markup over marginal cost as a percentage of price. The relationship says that this markup should equal minus the inverse of the elasticity of demand.[4] (This figure will be a *positive* number because the elasticity of demand is *negative*.) Equivalently, we can rearrange this equation to express price directly as a markup over marginal cost:

$$P = \frac{\text{MC}}{1 + (1/E_d)} \qquad \textbf{(10.2)}$$

[4] Remember that this markup equation applies at the point of a profit maximum. If both the elasticity of demand and marginal cost vary considerably over the range of outputs under consideration, you may have to know the entire demand and marginal cost curves to determine the optimum output level. On the other hand, you can use this equation to check whether a particular output level and price are optimal.

For example, if the elasticity of demand is -4 and marginal cost is \$9 per unit, price should be $\$9/(1 - 1/4) = \$9/.75 = \$12$ per unit.

How does the price set by a monopolist compare with the price under competition? In Chapter 8, we saw that in a perfectly competitive market, price equals marginal cost. A monopolist charges a price that exceeds marginal cost, *but by an amount that depends inversely on the elasticity of demand*. As the markup equation (10.1) shows, if demand is extremely elastic, E_d is a large negative number, and price will be very close to marginal cost. In that case, a monopolized market will look much like a competitive one. In fact, when demand is very elastic, there is little benefit to being a monopolist.

In §8.1, we explain that a perfectly competitive firm will choose its output so that marginal cost equals price.

Also note that a monopolist will never produce a quantity of output that is on the inelastic portion of the demand curve—i.e., where the elasticity of demand is less than 1 in absolute value. To see why, suppose that the monopolist is producing at a point on the demand curve where the elasticity is -0.5. In that case, the monopolist could make a greater profit by producing less and selling at a higher price. (A 10-percent reduction in output, for example, would allow for a 20-percent increase in price and thus a 10-percent increase in revenue. If marginal cost were greater than zero, the increase in profit would be even more than 10 percent because the lower output would reduce the firm's costs.) As the monopolist reduces output and raises price, it will move up the demand curve to a point where the elasticity is greater than 1 in absolute value and the markup rule of equation (10.2) will be satisfied.

Suppose, however, that marginal cost is zero. In that case, we cannot use equation (10.2) directly to determine the profit-maximizing price. However, we can see from equation (10.1) that in order to maximize profit, the firm will produce at the point where the elasticity of demand is exactly -1. If marginal cost is zero, maximizing profit is equivalent to maximizing revenue, and revenue is maximized when $E_d = -1$.

In §4.3 and Table 4.3, we explain that when price is increased, expenditure—and thus revenue—increases if demand is inelastic, decreases if demand is elastic, and is unchanged if demand has unit elasticity.

EXAMPLE 10.1 Astra-Merck Prices Prilosec

In 1995, a new drug developed by Astra-Merck became available for the long-term treatment of ulcers. The drug, Prilosec, represented a new generation of antiulcer medication. Other drugs to treat ulcer conditions were already on the market: Tagamet had been introduced in 1977, Zantac in 1983, Pepcid in 1986, and Axid in 1988. As we explained in Example 1.1 (page 10), these four drugs worked in much the same way to reduce the stomach's secretion of acid. Prilosec, however, was based on a very different biochemical mechanism and was much more effective than these earlier drugs. By 1996, it had become the best-selling drug in the world and faced no major competitor.[5]

[5]Prilosec, developed through a joint venture of the Swedish firm Astra and the U.S. firm Merck, was introduced in 1989, but only for the treatment of gastroesophageal reflux disease, and was approved for short-term ulcer treatment in 1991. It was the approval for long-term ulcer treatment in 1995, however, that created a very large market for the drug. In 1998, Astra bought Merck's share of the rights to Prilosec. In 1999, Astra acquired the firm Zeneca and is now called AstraZeneca. In 2001, AstraZeneca earned over \$4.9 billion in sales of Prilosec, which remained the world's best-selling prescription drug. As AstraZeneca's patent on Prilosec neared expiration, the company introduced Nexium, a new (and, according to the company, better) antiulcer drug. In 2006, Nexium was the third-biggest-selling pharmaceutical drug in the world, with sales of about \$5.7 billion.

In 1995, Astra-Merck was pricing Prilosec at about \$3.50 per daily dose. (By contrast, the prices for Tagamet and Zantac were about \$1.50 to \$2.25 per daily dose.) Is this pricing consistent with the markup formula (10.1)? The marginal cost of producing and packaging Prilosec is only about 30 to 40 cents per daily dose. This low marginal cost implies that the price elasticity of demand, E_D, should be in the range of roughly -1.0 to -1.2. Based on statistical studies of pharmaceutical demand, this is indeed a reasonable estimate for the demand elasticity. Thus, setting the price of Prilosec at a markup exceeding 400 percent over marginal cost is consistent with our rule of thumb for pricing.

Shifts in Demand

In a competitive market, there is a clear relationship between price and the quantity supplied. That relationship is the supply curve, which, as we saw in Chapter 8, represents the marginal cost of production for the industry as a whole. The supply curve tells us how much will be produced at every price.

A monopolistic market has no supply curve. In other words, there is no one-to-one relationship between price and the quantity produced. The reason is that the monopolist's output decision depends not only on marginal cost but also on the shape of the demand curve. As a result, shifts in demand do not trace out the series of prices and quantities that correspond to a competitive supply curve. Instead, shifts in demand can lead to changes in price with no change in output, changes in output with no change in price, or changes in both price and output.

This principle is illustrated in Figure 10.4(a) and (b). In both parts of the figure, the demand curve is initially D_1, the corresponding marginal revenue curve is MR_1, and the monopolist's initial price and quantity are P_1 and Q_1. In Figure 10.4(a), the demand curve is shifted down and rotated. The new demand and marginal revenue curves are shown as D_2 and MR_2. Note that MR_2 intersects the marginal cost curve at the same point that MR_1 does. As a result, the quantity produced stays the same. Price, however, falls to P_2.

In Figure 10.4(b), the demand curve is shifted up and rotated. The new marginal revenue curve MR_2 intersects the marginal cost curve at a larger quantity, Q_2 instead of Q_1. But the shift in the demand curve is such that the price charged is exactly the same.

Shifts in demand usually cause changes in both price and quantity. But the special cases shown in Figure 10.4 illustrate an important distinction between monopoly and competitive supply. A competitive industry supplies a specific quantity at every price. No such relationship exists for a monopolist, which, depending on how demand shifts, might supply several different quantities at the same price, or the same quantity at different prices.

The Effect of a Tax

A tax on output can also have a different effect on a monopolist than on a competitive industry. In Chapter 9, we saw that when a specific (i.e., per-unit) tax is imposed on a competitive industry, the market price rises by an amount that is less than the tax, and that the burden of the tax is shared by producers and consumers. Under monopoly, however, price can sometimes rise by *more* than the amount of the tax.

In §9.6, we explain that a specific tax is a tax of a certain amount of money per unit sold, and we show how the tax affects price and quantity.

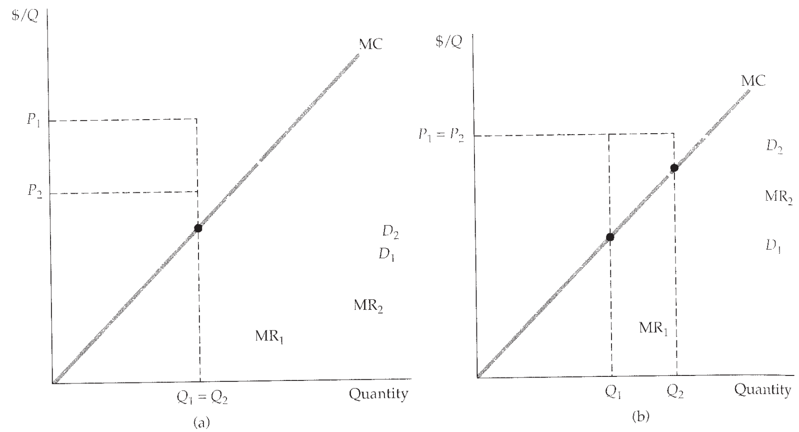

FIGURE 10.4 Shifts in Demand

Shifting the demand curve shows that a monopolistic market has no supply curve—i.e., there is no one-to-one relationship between price and quantity produced. In **(a)**, the demand curve D_1 shifts to new demand curve D_2. But the new marginal revenue curve MR_2 intersects marginal cost at the same point as the old marginal revenue curve MR_1. The profit-maximizing output therefore remains the same, although price falls from P_1 to P_2. In **(b)**, the new marginal revenue curve MR_2 intersects marginal cost at a higher output level Q_2. But because demand is now more elastic, price remains the same.

Analyzing the effect of a tax on a monopolist is straightforward. Suppose a specific tax of t dollars per unit is levied, so that the monopolist must remit t dollars to the government for every unit it sells. Therefore, the firm's marginal (and average) cost is increased by the amount of the tax t. If MC was the firm's original marginal cost, its optimal production decision is now given by

$$MR = MC + t$$

In §8.2, we explain that a firm maximizes its profit by choosing the output at which marginal revenue is equal to marginal cost.

Graphically, we shift the marginal cost curve upward by an amount t, and find the new intersection with marginal revenue. Figure 10.5 shows this. Here Q_0 and P_0 are the quantity and price before the tax is imposed, and Q_1 and P_1 are the quantity and price after the tax.

Shifting the marginal cost curve upward results in a smaller quantity and higher price. Sometimes price increases by less than the tax, but not always—in Figure 10.5, price increases by *more* than the tax. This would be impossible in a competitive market, but it can happen with a monopolist because the relationship between price and marginal cost depends on the elasticity of demand. Suppose, for example, that a monopolist faces a constant elasticity demand curve, with elasticity –2, and has constant marginal cost MC. Equation (10.2) then tells us that price will equal twice marginal cost. With a tax t, marginal cost increases to MC + t, so price increases to 2(MC + t) = 2MC + 2t; that is, it rises by twice the amount of the tax. (However, the monopolist's profit nonetheless falls with the tax.)

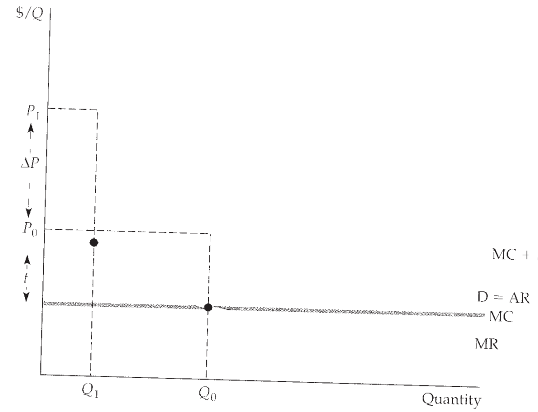

FIGURE 10.5 Effect of Excise Tax on Monopolist

With a tax t per unit, the firm's effective marginal cost is increased by the amount t to MC + t. In this example, the increase in price ΔP is larger than the tax t.

We have seen that a firm maximizes profit by setting output at a level where marginal revenue equals marginal cost. For many firms, production takes place in two or more different plants whose operating costs can differ. However, the logic used in choosing output levels is very similar to that for the single-plant firm.

Suppose a firm has two plants. What should its total output be, and how much of that output should each plant produce? We can find the answer intuitively in two steps.

- **Step 1.** Whatever the total output, it should be divided between the two plants so that *marginal cost is the same in each plant*. Otherwise, the firm could reduce its costs and increase its profit by reallocating production. For example, if marginal cost at Plant 1 were higher than at Plant 2, the firm could produce the same output at a lower total cost by producing less at Plant 1 and more at Plant 2.
- **Step 2.** We know that total output must be such that *marginal revenue equals marginal cost*. Otherwise, the firm could increase its profit by raising or lowering total output. For example, suppose marginal costs were the same at each plant, but marginal revenue exceeded marginal cost. In that case, the firm would do better by producing more at both plants because the revenue earned from the additional units would exceed the cost. Because marginal costs must be the same at each plant, and because marginal revenue must equal marginal cost, we see that profit is maximized when *marginal revenue equals marginal cost at each plant*.

We can also derive this result algebraically. Let Q_1 and C_1 be the output and cost of production for Plant 1, Q_2 and C_2 be the output and cost of production for Plant 2, and $Q_T = Q_1 + Q_2$ be total output. Then profit is

$$\pi = PQ_T - C_1(Q_1) - C_2(Q_2)$$

The firm should increase output from each plant until the incremental profit from the last unit produced is zero. Start by setting incremental profit from output at Plant 1 to zero:

$$\frac{\Delta\pi}{\Delta Q_1} = \frac{\Delta(PQ_T)}{\Delta Q_1} - \frac{\Delta C_1}{\Delta Q_1} = 0$$

Here $\Delta(PQ_T)/\Delta Q_1$ is the revenue from producing and selling one more unit—i.e., *marginal revenue*, MR, for all of the firm's output. The next term, $\Delta C_1/\Delta Q_1$, is *marginal cost* at Plant 1, MC_1. We thus have $MR - MC_1 = 0$, or

$$MR = MC_1$$

Similarly, we can set incremental profit from output at Plant 2 to zero,

$$MR = MC_2$$

Putting these relations together, we see that the firm should produce so that

$$MR = MC_1 = MC_2 \qquad (10.3)$$

Figure 10.6 illustrates this principle for a firm with two plants. MC_1 and MC_2 are the marginal cost curves for the two plants. (Note that Plant 1 has higher marginal costs than Plant 2.) Also shown is a curve labeled MC_T. This is the firm's total marginal cost and is obtained by horizontally summing MC_1 and

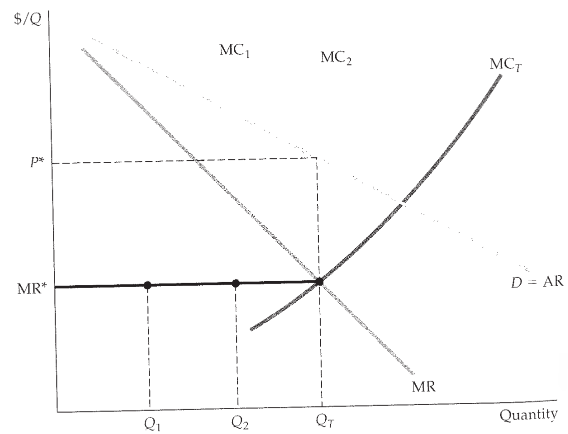

FIGURE 10.6 Production with Two Plants

A firm with two plants maximizes profits by choosing output levels Q_1 and Q_2 so that marginal revenue MR (which depends on *total* output) equals marginal costs for each plant, MC_1 and MC_2.

MC_2. Now we can find the profit-maximizing output levels Q_1, Q_2, and Q_T. First, find the intersection of MC_T with MR; that point determines total output Q_T. Next, draw a horizontal line from that point on the marginal revenue curve to the vertical axis; point MR* determines the firm's marginal revenue. The intersections of the marginal revenue line with MC_1 and MC_2 give the outputs Q_1 and Q_2 for the two plants, as in equation (10.3).

Note that total output Q_T determines the firm's marginal revenue (and hence its price P^*). Q_1 and Q_2, however, determine marginal costs at each of the two plants. Because MC_T was found by horizontally summing MC_1 and MC_2, we know that $Q_1 + Q_2 = Q_T$. Thus these output levels satisfy the condition that $MR = MC_1 = MC_2$.

Note the similarity to the way we obtained a competitive industry's supply curve in §8.5 by horizontally summing the marginal cost curves of the individual firms.

10.2 MONOPOLY POWER

Pure monopoly is rare. Markets in which several firms compete with one another are much more common. We say more about the forms that this competition can take in Chapters 12 and 13. But we should explain here why each firm in a market with several firms is likely to face a downward-sloping demand curve and, as a result, to produce so that price exceeds marginal cost.

Suppose, for example, that four firms produce toothbrushes and have the market demand curve $Q = 50{,}000 - 20{,}000P$, as shown in Figure 10.7(a). Let's assume that these four firms are producing an aggregate of 20,000 toothbrushes per day (5000 each per day) and selling them at \$1.50 each. Note that market demand is relatively inelastic; you can verify that at this \$1.50 price, the elasticity of demand is −1.5.

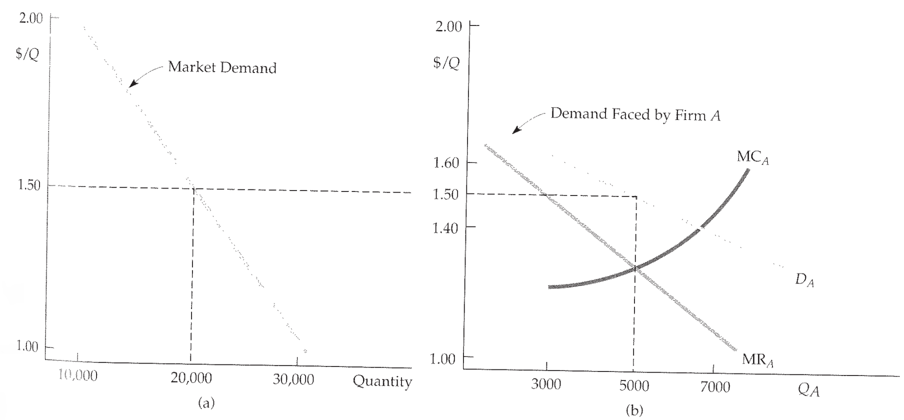

FIGURE 10.7 The Demand for Toothbrushes

Part **(a)** shows the market demand for toothbrushes. Part **(b)** shows the demand for toothbrushes as seen by Firm *A*. At a market price of \$1.50, elasticity of market demand is −1.5. Firm *A*, however, sees a much more elastic demand curve D_A because of competition from other firms. At a price of \$1.50, Firm *A*'s demand elasticity is −6. Still, Firm *A* has some monopoly power: Its profit-maximizing price is \$1.50, which exceeds marginal cost.

Now suppose that Firm A is deciding whether to lower its price to increase sales. To make this decision, it needs to know how its sales would respond to a change in its price. In other words, it needs some idea of the demand curve it faces, as opposed to the *market* demand curve. A reasonable possibility is shown in Figure 10.7(b), where the firm's demand curve D_A is much more elastic than the market demand curve. (At the \$1.50 price the elasticity is −6.0.) The firm might predict that by raising the price from \$1.50 to \$1.60, its sales will drop—say, from 5000 units to 3000—as consumers buy more toothbrushes from other firms. (If *all* firms raised their prices to \$1.60, sales for Firm A would fall only to 4500.) For several reasons, sales won't drop to zero as they would in a perfectly competitive market. First, if Firm A's toothbrushes are a little different from those of its competitors, some consumers will pay a bit more for them. Second, other firms might also raise their prices. Similarly, Firm A might anticipate that by lowering its price from \$1.50 to \$1.40, it can sell more toothbrushes—perhaps 7000 instead of 5000. But it will not capture the entire market: Some consumers might still prefer the competitors' toothbrushes, and competitors might also lower their prices.

Thus, Firm A's demand curve depends both on how much its product differs from its competitors' products and on how the four firms compete with one another. We will discuss product differentiation and interfirm competition in Chapters 12 and 13. But one important point should be clear: *Firm A is likely to face a demand curve which is more elastic than the market demand curve, but which is not infinitely elastic like the demand curve facing a perfectly competitive firm.*

Given knowledge of its demand curve, how much should Firm A produce? The same principle applies: The profit-maximizing quantity equates marginal revenue and marginal cost. In Figure 10.7(b), that quantity is 5000 units. The corresponding price is \$1.50, which exceeds marginal cost. Thus, although Firm A is not a pure monopolist, *it does have monopoly power*—it can profitably charge a price greater than marginal cost. Of course, its monopoly power is less than it would be if it had driven away the competition and monopolized the market, but it might still be substantial.

This raises two questions.

1. How can we *measure* monopoly power in order to compare one firm with another? (So far we have been talking about monopoly power only in *qualitative* terms.)
2. What are the *sources* of monopoly power, and why do some firms have more monopoly power than others?

We address both these questions below, although a more complete answer to the second question will be provided in Chapters 12 and 13.

Remember the important distinction between a perfectly competitive firm and a firm with monopoly power: *For the competitive firm, price equals marginal cost; for the firm with monopoly power, price exceeds marginal cost.* Therefore, a natural way to measure monopoly power is to examine the extent to which the profit-maximizing price exceeds marginal cost. In particular, we can use the markup ratio of price minus marginal cost to price that we introduced earlier as part of a rule of thumb for pricing. This measure of monopoly power, introduced by economist Abba Lerner in 1934, is called the **Lerner Index of Monopoly Power**. It is the difference between price and marginal cost, divided by price. Mathematically:

Lerner Index of Monopoly Power Measure of monopoly power calculated as excess of price over marginal cost as a fraction of price.

$$L = (P - \text{MC})/P$$

The Lerner index always has a value between zero and one. For a perfectly competitive firm, $P = \text{MC}$, so that $L = 0$. The larger is L, the greater is the degree of monopoly power.

This index of monopoly power can also be expressed in terms of the elasticity of demand facing the firm. Using equation (10.1), we know that

$$L = (P - \text{MC})/P = -1/E_d \qquad \textbf{(10.4)}$$

Remember, however, that E_d is now the elasticity of the *firm's* demand curve, not the market demand curve. In the toothbrush example discussed previously, the elasticity of demand for Firm A is − 6.0, and the degree of monopoly power is $1/6 = 0.167$.[6]

Note that considerable monopoly power does not necessarily imply high profits. Profit depends on *average* cost relative to price. Firm A might have more monopoly power than Firm B but earn a lower profit because of higher average costs.

In the previous section, we used equation (10.2) to compute price as a simple markup over marginal cost:

$$P = \frac{\text{MC}}{1 + (1/E_d)}$$

This relationship provides a rule of thumb for *any* firm with monopoly power. We must remember, however, that E_d is the elasticity of demand for the *firm*, not the elasticity of *market* demand.

It is harder to determine the elasticity of demand for the firm than for the market because the firm must consider how its competitors will react to price changes. Essentially, the manager must estimate the percentage change in the firm's unit sales that is likely to result from a 1-percent change in the firm's price. This estimate might be based on a formal model or on the manager's intuition and experience.

Given an estimate of the firm's elasticity of demand, the manager can calculate the proper markup. If the firm's elasticity of demand is large, this markup will be small (and we can say that the firm has very little monopoly power). If the firm's elasticity of demand is small, this markup will be large (and the firm will have considerable monopoly power). Figures 10.8(a) and 10.8(b) illustrate these two extremes.

[6] There are three problems with applying the Lerner index to the analysis of public policy toward firms. First, because marginal cost is difficult to measure, average variable cost is often used in Lerner index calculations. Second, if the firm prices below its optimal price (possibly to avoid legal scrutiny), its potential monopoly power will not be noted by the index. Third, the index ignores dynamic aspects of pricing such as effects of the learning curve and shifts in demand. See Robert S. Pindyck, "The Measurement of Monopoly Power in Dynamic Markets," *Journal of Law and Economics* 28 (April 1985): 193–222.

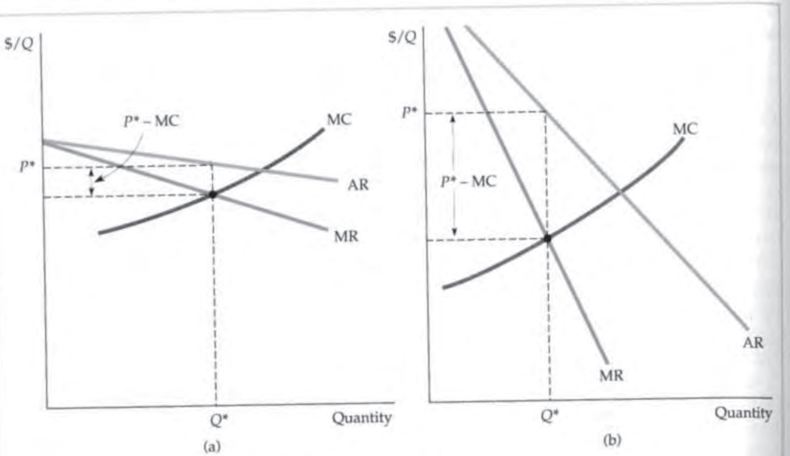

FIGURE 10.8 Elasticity of Demand and Price Markup

The markup $(P - \text{MC})/P$ is equal to minus the inverse of the elasticity of demand facing the firm. If the firm's demand is elastic, as in **(a)**, the markup is small and the firm has little monopoly power. The opposite is true if demand is relatively inelastic, as in **(b)**.

EXAMPLE 10.2 Markup Pricing: Supermarkets to Designer Jeans

Three examples should help clarify the use of markup pricing. Consider a supermarket chain. Although the elasticity of market demand for food is small (about –1), several supermarkets usually serve most areas. Thus no single supermarket can raise its prices very much without losing customers to other stores. As a result, the elasticity of demand for any one supermarket is often as large as –10. Substituting this number for E_d in equation (10.2), we find $P = \text{MC}/(1 - 0.1) = \text{MC}/(0.9) = (1.11)\text{MC}$. In other words, the manager of a typical supermarket should set prices about 11 percent above marginal cost. For a reasonably wide range of output levels (over which the size of the store and the number of its employees will remain fixed), marginal cost includes the cost of purchasing the food at wholesale, plus the costs of storing the food, arranging it on the shelves, etc. For most supermarkets, the markup is indeed about 10 or 11 percent.

Small convenience stores, which are often open 7 days a week and even 24 hours a day, typically charge higher prices than supermarkets. Why? Because a convenience store faces a less elastic demand curve. Its customers are generally less price sensitive. They might need a quart of milk or a loaf of bread late at night or may find it inconvenient to drive to the supermarket. Because the elasticity of demand for a convenience store is about –5, the markup equation implies that its prices should be about 25 percent above marginal cost, as indeed they typically are.

The Lerner index, $(P - \text{MC})/P$, tells us that the convenience store has more monopoly power, but does it make larger profits? No. Because its volume is far smaller and its average fixed costs are larger, it usually earns a much smaller profit than a large supermarket despite its higher markup.

Finally, consider a producer of designer jeans. Many companies produce jeans, but some consumers will pay much more for jeans with a designer label. Just how much more they will pay—or more exactly, how much sales will drop in response to higher prices—is a question that the producer must carefully consider because it is critical in determining the price at which the clothing will be sold (at wholesale to retail stores, which then mark up the price further). With designer jeans, demand elasticities in the range of –2 to –3 are typical for the major labels. This means that price should be 50 to 100 percent higher than marginal cost. Marginal cost is typically \$15 to \$20 per pair, and depending on the brand, the wholesale price is in the \$20 to \$40 range. In contrast, "mass-market" jeans will typically wholesale for \$18 to \$25 per pair. Why? Because without the designer label, they are far more price elastic.

EXAMPLE 10.3 The Pricing of Videos

During the mid-1980s, the number of households owning videocassette recorders (VCRs) grew rapidly, as did the markets for rentals and sales of prerecorded cassettes. Although at that time many more videocassettes were rented through small retail outlets than sold outright, the market for sales was large and growing. Producers, however, found it difficult to decide what price to charge for cassettes. As a result, in 1985 popular movies were selling for vastly different prices, as you can see from the data in Table 10.2.

Note that while *The Empire Strikes Back* was selling for nearly \$80, *Star Trek*, a film that appealed to the same audience and was about as popular, sold for only about \$25. These price differences reflected uncertainty and a wide divergence of

TABLE 10.2 Retail Prices of VHS and DVDs

1985		2007	
Title	Retail Price VHS	Title	Retail Price DVD
Purple Rain	\$29.98	*Pirates of the Caribbean*	\$19.99
Raiders of the Lost Ark	\$24.95	*The Da Vinci Code*	\$19.99
Jane Fonda Workout	\$59.95	*Mission: Impossible III*	\$17.99
The Empire Strikes Back	\$79.98	*King Kong*	\$19.98
An Officer and a Gentleman	\$24.95	*Harry Potter and the Goblet of Fire*	\$17.49
Star Trek: The Motion Picture	\$24.95	*Ice Age*	\$19.99
Star Wars	\$39.98	*The Devil Wears Prada*	\$17.99

Source (2007): Based on http://www.amazon.com. Suggested retail price.

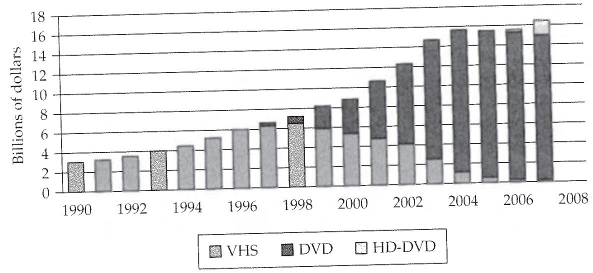

FIGURE 10.9 Video Sales

Between 1990 and 1998, lower prices induced consumers to buy many more videos. By 2001, sales of DVDs overtook sales of VHS videocassettes. High-definition DVDs were introduced in 2006, and are expected to displace sales of conventional DVDs.

views on pricing by producers. The issue was whether lower prices would induce consumers to buy videocassettes rather than rent them. Because producers do not share in the retailers' revenues from rentals, they should charge a low price for cassettes only if that will induce enough consumers to buy them. Because the market was young, producers had no good estimates of the elasticity of demand, so they based prices on hunches or trial and error.[7]

As the market matured, however, sales data and market research studies put pricing decisions on firmer ground. Those studies strongly indicated that demand was price elastic and that the profit-maximizing price was in the range of $15 to $30. By the 1990s, most producers had lowered prices across the board. When DVDs were first introduced in 1997, the prices of top-selling DVDs were much more uniform. Since that time, prices of popular DVDs have remained fairly uniform and continued to fall. As Table 10.2 shows, by 2007, prices were typically $20 or less. As a result, video sales have steadily increased, as shown in Figure 10.9. With the introduction of high-definition (HD) DVDs in 2006, sales of conventional DVDs are expected to fall as consumers gradually switch to the new format.

10.3 SOURCES OF MONOPOLY POWER

Why do some firms have considerable monopoly power while other firms have little or none? Remember that monopoly power is the ability to set price above marginal cost and that the amount by which price exceeds marginal cost depends inversely on the elasticity of demand facing the firm. As equation (10.4) shows, *the less elastic its demand curve, the more monopoly power a firm has.* The ultimate determinant of monopoly power is therefore the firm's elasticity of demand. Thus we should rephrase our question: Why do some firms (e.g., a supermarket chain) face demand curves that are more elastic than those faced by others (e.g., a producer of designer clothing)?

Three factors determine a firm's elasticity of demand.

1. *The elasticity of market demand.* Because the firm's own demand will be at least as elastic as market demand, the elasticity of market demand limits the potential for monopoly power.
2. *The number of firms in the market.* If there are many firms, it is unlikely that any one firm will be able to affect price significantly.
3. *The interaction among firms.* Even if only two or three firms are in the market, each firm will be unable to profitably raise price very much if the rivalry among them is aggressive, with each firm trying to capture as much of the market as it can.

Let's examine each of these three determinants of monopoly power.

If there is only one firm—a pure monopolist—its demand curve is the market demand curve. In this case, the firm's degree of monopoly power depends completely on the elasticity of market demand. More often, however, several firms compete with one another; then the elasticity of market demand sets a lower limit on the magnitude of the elasticity of demand for each firm. Recall our example of the toothbrush producers illustrated in Figure 10.7 (page 361). The market demand for toothbrushes might not be very elastic, but each firm's demand will be more elastic. (In Figure 10.7, the elasticity of market demand is −1.5, and the elasticity of demand for each firm is −6.) A particular firm's elasticity depends on how the firms compete with one another. But no matter how they compete, the elasticity of demand for each firm could never become smaller in magnitude than −1.5.

Because the demand for oil is fairly inelastic (at least in the short run), OPEC could raise oil prices far above marginal production cost during the 1970s and early 1980s. Because the demands for such commodities as coffee, cocoa, tin, and copper are much more elastic, attempts by producers to cartelize these markets and raise prices have largely failed. In each case, the elasticity of market demand limits the potential monopoly power of individual producers.

The second determinant of a firm's demand curve—and thus of its monopoly power—is the number of firms in its market. Other things being equal, the monopoly power of each firm will fall as the number of firms increases: As more and more firms compete, each firm will find it harder to raise prices and avoid losing sales to other firms.

What matters, of course, is not just the total number of firms, but the number of "major players"—firms with significant market share. For example, if only two large firms account for 90 percent of sales in a market, with another 20 firms accounting for the remaining 10 percent, the two large firms might have considerable monopoly power. When only a few firms account for most of the sales in a market, we say that the market is highly *concentrated*.[8]

[7] "Video Producers Debate the Value of Price Cuts," *New York Times*, February 19, 1985. For a study of videocassette pricing, see Carl E. Enomoto and Soumendra N. Ghosh, "Pricing in the Home-Video Market" (working paper, New Mexico State University, 1992).

[8] A statistic called the *concentration ratio*, which measures the percentage of sales accounted for by, say, the four largest firms, is often used to describe the concentration of a market. Concentration is one, but not the only, determinant of market power.

It is sometimes said (not always jokingly) that the greatest fear of American business is competition. That may or may not be true. But we would certainly expect that when only a few firms are in a market, their managers will prefer that no new firms enter. An increase in the number of firms can only reduce the monopoly power of each incumbent firm. An important aspect of competitive strategy (discussed in detail in Chapter 13) is finding ways to create **barriers to entry**—conditions that deter entry by new competitors.

barrier to entry Condition that impedes entry by new competitors.

Sometimes there are natural barriers to entry. For example, one firm may have a *patent* on the technology needed to produce a particular product. This makes it impossible for other firms to enter the market, at least until the patent expires. Other legally created rights work in the same way—a *copyright* can limit the sale of a book, music, or a computer software program to a single company, and the need for a government *license* can prevent new firms from entering the markets for telephone service, television broadcasting, or interstate trucking. Finally, *economies of scale* may make it too costly for more than a few firms to supply the entire market. In some cases, economies of scale may be so large that it is most efficient for a single firm—*a natural monopoly*—to supply the entire market. We will discuss scale economies and natural monopoly in more detail shortly.

In §7.4, we explain that a firm enjoys economies of scale when it can double its output with less than a doubling of cost.

The Interaction Among Firms

The ways in which competing firms interact is also an important—and sometimes the most important—determinant of monopoly power. Suppose there are four firms in a market. They might compete aggressively, undercutting one another's prices to capture more market share. This could drive prices down to nearly competitive levels. Each firm will fear that if it raises its price it will be undercut and lose market share. As a result, it will have little monopoly power.

On the other hand, the firms might not compete much. They might even collude (in violation of the antitrust laws), agreeing to limit output and raise prices. Because raising prices in concert rather than individually is more likely to be profitable, collusion can generate substantial monopoly power.

We will discuss the interaction among firms in detail in Chapters 12 and 13. Now we simply want to point out that, other things being equal, monopoly power is smaller when firms compete aggressively and is larger when they cooperate.

Remember that a firm's monopoly power often changes over time, as its operating conditions (market demand and cost), its behavior, and the behavior of its competitors change. Monopoly power must therefore be thought of in a dynamic context. For example, the market demand curve might be very inelastic in the short run but much more elastic in the long run. (Because this is the case with oil, the OPEC cartel enjoyed considerable short-run but much less long-run monopoly power.) Furthermore, real or potential monopoly power in the short run can make an industry more competitive in the long run: Large short-run profits can induce new firms to enter an industry, thereby reducing monopoly power over the longer term.

10.4 THE SOCIAL COSTS OF MONOPOLY POWER

In a competitive market, price equals marginal cost. Monopoly power, on the other hand, implies that price exceeds marginal cost. Because monopoly power results in higher prices and lower quantities produced, we would expect it to make consumers worse off and the firm better off. But suppose we value the

welfare of consumers the same as that of producers. In the aggregate, does monopoly power make consumers and producers better or worse off?

We can answer this question by comparing the consumer and producer surplus that results when a competitive industry produces a good with the surplus that results when a monopolist supplies the entire market.[9] (We assume that the competitive market and the monopolist have the same cost curves.) Figure 10.10 shows the average and marginal revenue curves and marginal cost curve for the monopolist. To maximize profit, the firm produces at the point where marginal revenue equals marginal cost, so that the price and quantity are P_m and Q_m. In a competitive market, price must equal marginal cost, so the competitive price and quantity, P_c and Q_c are found at the intersection of the average revenue (demand) curve and the marginal cost curve. Now let's examine how surplus changes if we move from the competitive price and quantity, P_c and Q_c to the monopoly price and quantity, P_m and Q_m.

In §9.1, we explain that consumer surplus is the total benefit or value that consumers receive beyond what they pay for a good; producer surplus is the analogous measure for producers.

Under monopoly, the price is higher and consumers buy less. Because of the higher price, those consumers who buy the good lose surplus of an amount given by rectangle *A*. Those consumers who do not buy the good at price P_m but who would buy at price P_c also lose surplus—namely, an amount given by triangle *B*. The total loss of consumer surplus is therefore $A + B$. The producer, however, gains rectangle *A* by selling at the higher price but loses triangle *C*, the additional profit it would have earned by selling $Q_c - Q_m$ at price P_c. The total gain in producer surplus is therefore $A - C$. Subtracting the loss of consumer surplus from the gain in producer surplus, we see a net loss of surplus given by

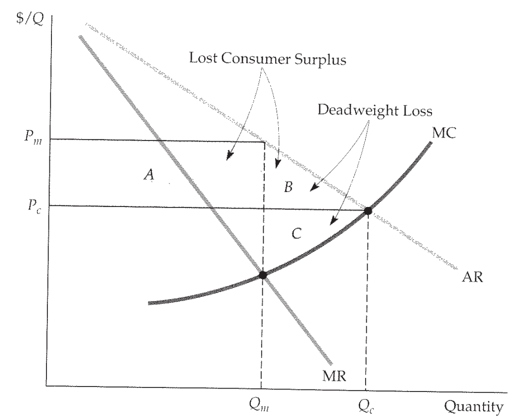

FIGURE 10.10 Deadweight Loss from Monopoly Power

The shaded rectangle and triangles show changes in consumer and producer surplus when moving from competitive price and quantity, P_c and Q_c, to a monopolist's price and quantity, P_m and Q_m. Because of the higher price, consumers lose $A + B$ and producer gains $A - C$. The deadweight loss is $B + C$.

[9] If there were two or more firms, each with some monopoly power, the analysis would be more complex. However, the basic results would be the same.

B + C. This is the *deadweight loss from monopoly power*. Even if the monopolist's profits were taxed away and redistributed to the consumers of its products, there would be an inefficiency because output would be lower than under conditions of competition. The deadweight loss is the social cost of this inefficiency.

rent seeking Spending money in socially unproductive efforts to acquire, maintain, or exercise monopoly.

In practice, the social cost of monopoly power is likely to exceed the deadweight loss in triangles B and C of Figure 10.10. The reason is that the firm may engage in **rent seeking**: spending large amounts of money in socially unproductive efforts to acquire, maintain, or exercise its monopoly power. Rent seeking might involve lobbying activities (and perhaps campaign contributions) to obtain government regulations that make entry by potential competitors more difficult. Rent-seeking activity could also involve advertising and legal efforts to avoid antitrust scrutiny. It might also mean installing but not utilizing extra production capacity to convince potential competitors that they cannot sell enough to make entry worthwhile. We would expect the economic incentive to incur rent-seeking costs to bear a direct relation to the gains from monopoly power (i.e., rectangle A minus triangle C.) Therefore, the larger the transfer from consumers to the firm (rectangle A), the larger the social cost of monopoly.[10]

Here's an example. In 1996, the Archer Daniels Midland Company (ADM) successfully lobbied the Clinton administration for regulations requiring that the ethanol (ethyl alcohol) used in motor vehicle fuel be produced from corn. (The government had already planned to add ethanol to gasoline in order to reduce the country's dependence on imported oil.) Ethanol is chemically the same whether it is produced from corn, potatoes, grain, or anything else. Then why require that it be produced only from corn? Because ADM had a near monopoly on corn-based ethanol production, so the regulation would increase its gains from monopoly power.

Because of its social cost, antitrust laws prevent firms from accumulating excessive amounts of monopoly power. We will say more about such laws at the end of the chapter. Here, we examine another means by which government can limit monopoly power—price regulation.

We saw in Chapter 9 that in a competitive market, price regulation always results in a deadweight loss. This need not be the case, however, when a firm has monopoly power. On the contrary, price regulation can eliminate the deadweight loss that results from monopoly power.

Figure 10.11 illustrates price regulation. P_m and Q_m are the price and quantity that result without regulation—i.e., at the point where marginal revenue equals marginal cost. Now suppose the price is regulated to be no higher than P_1. To find the firm's profit-maximizing output, we must determine how its average and marginal revenue curves are affected by the regulation.

Because the firm can charge no more than P_1 for output levels up to Q_1, its new average revenue curve is a horizontal line at P_1. For output levels greater than Q_1, the new average revenue curve is identical to the old average revenue

[10]The concept of rent seeking was first developed by Gordon Tullock. For more detailed discussions see Gordon Tullock, *Rent Seeking* (Brookfield, VT: Edward Elgar, 1993), or Robert D. Tollison and Roger D. Congleton, *The Economic Analysis of Rent Seeking* (Brookfield, VT: Edward Elgar, 1995).

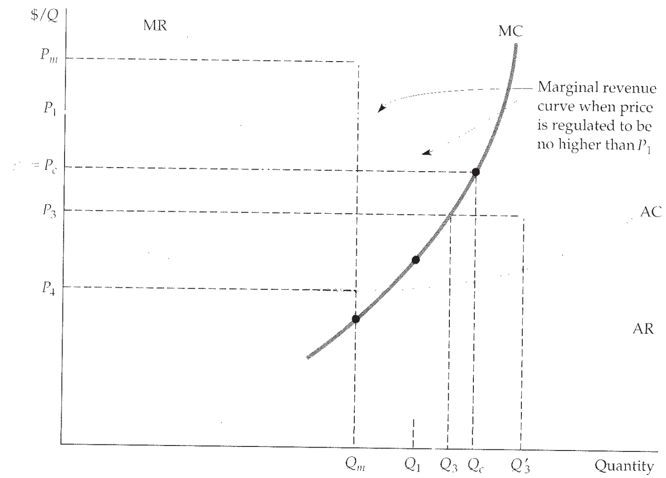

FIGURE 10.11 Price Regulation

If left alone, a monopolist produces Q_m and charges P_m. When the government imposes a price ceiling of P_1 the firm's average and marginal revenue are constant and equal to P_1 for output levels up to Q_1. For larger output levels, the original average and marginal revenue curves apply. The new marginal revenue curve is, therefore, the dark purple line, which intersects the marginal cost curve at Q_1. When price is lowered to P_c, at the point where marginal cost intersects average revenue, output increases to its maximum Q_c. This is the output that would be produced by a competitive industry. Lowering price further, to P_3 reduces output to Q_3 and causes a shortage, $Q'_3 - Q_3$.

curve. At these output levels, the firm will charge less than P_1 and so will be unaffected by the regulation.

The firm's new marginal revenue curve corresponds to its new average revenue curve and is shown by the purple line in Figure 10.11. For output levels up to Q_1, marginal revenue equals average revenue. (Recall that, as with a competitive firm, if average revenue is constant, average revenue and marginal revenue are equal.) For output levels greater than Q_1, the new marginal revenue curve is identical to the original curve. Thus the complete marginal revenue curve now has three pieces: (1) the horizontal line at P_1 for quantities up to Q_1; (2) a vertical line at the quantity Q_1 connecting the original average and marginal revenue curves; and (3) the original marginal revenue curve for quantities greater than Q_1.

To maximize its profit, the firm should produce the quantity Q_1 because that is the point at which its marginal revenue curve intersects its marginal cost curve. You can verify that at price P_1 and quantity Q_1, the deadweight loss from monopoly power is reduced.

As the price is lowered further, the quantity produced continues to increase and the deadweight loss to decline. At price P_c where average revenue and marginal cost intersect, the quantity produced has increased to the competitive level; the deadweight loss from monopoly power has been eliminated. Reducing the price even more—say, to P_3—results in a *reduction* in quantity. This reduction is equivalent to imposing a price ceiling on a competitive industry. A shortage develops, $(Q_3' - Q_3)$, in addition to the deadweight loss from regulation. As the price is lowered further, the quantity produced continues to fall and the shortage grows. Finally, if the price is lowered below P_4, the minimum average cost, the firm loses money and goes out of business.

natural monopoly Firm that can produce the entire output of the market at a cost lower than what it would be if there were several firms.

Price regulation is most often used for *natural monopolies*, such as local utility companies. A **natural monopoly** is a firm that can produce the entire output of the market at a cost that is lower than what it would be if there were several firms. If a firm is a natural monopoly, it is more efficient to let it serve the entire market rather than have several firms compete.

A natural monopoly usually arises when there are strong economies of scale, as illustrated in Figure 10.12. If the firm represented by the figure was broken up into two competing firms, each supplying half the market, the average cost for each would be higher than the cost incurred by the original monopoly.

Note in Figure 10.12 that because average cost is declining everywhere, marginal cost is always below average cost. If the firm were unregulated, it would produce Q_m and sell at the price P_m. Ideally, the regulatory agency would like to push the firm's price down to the competitive level P_c. At that level, however, price would not cover average cost and the firm would go out of business. The best alternative is therefore to set the price at P_r, where average cost and average revenue intersect. In that case, the firm earns no monopoly profit, while output remains as large as possible without driving the firm out of business.

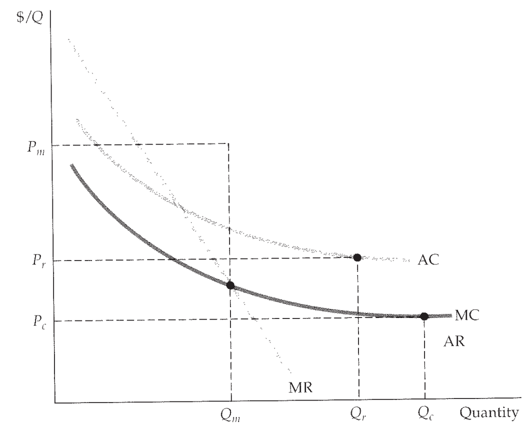

FIGURE 10.12 Regulating the Price of a Natural Monopoly

A firm is a natural monopoly because it has economies of scale (declining average and marginal costs) over its entire output range. If price were regulated to be P_c the firm would lose money and go out of business. Setting the price at P_r yields the largest possible output consistent with the firm's remaining in business; excess profit is zero.

Recall that the competitive price (P_c in Figure 10.11) is found at the point at which the firm's marginal cost and average revenue (demand) curves intersect. Likewise for a natural monopoly: The minimum feasible price (P_r in Figure 10.12) is found at the point at which average cost and demand intersect. Unfortunately, it is often difficult to determine these prices accurately in practice because the firm's demand and cost curves may shift as market conditions evolve.

As a result, the regulation of a monopoly is sometimes based on the rate of return that it earns on its capital. The regulatory agency determines an allowed price, so that this rate of return is in some sense "competitive" or "fair." This practice is called **rate-of-return regulation**: The maximum price allowed is based on the (expected) rate of return that the firm will earn.[11]

rate-of-return regulation Maximum price allowed by a regulatory agency is based on the (expected) rate of return that a firm will earn.

Unfortunately, difficult problems arise when implementing rate-of-return regulation. First, although it is a key element in determining the firm's rate of return, a firm's capital stock is difficult to value. Second, while a "fair" rate of return must be based on the firm's actual cost of capital, that cost depends in turn on the behavior of the regulatory agency (and on investors' perceptions of what allowed rates of return will be in the future).

The difficulty of agreeing on a set of numbers to be used in rate-of-return calculations often leads to delays in the regulatory response to changes in cost and other market conditions (not to mention long and expensive regulatory hearings). The major beneficiaries are usually lawyers, accountants, and, occasionally, economic consultants. The net result is *regulatory lag*—the delays of a year or more usually entailed in changing regulated prices.

Another approach to regulation is setting price caps based on the firm's variable costs, past prices, and possibly inflation and productivity growth. A price cap can allow for more flexibility than rate-of-return regulation. Under price cap regulation, for example, a firm would typically be allowed to raise its prices each year (without having to get approval from the regulatory agency) by an amount equal to the actual rate of inflation, minus expected productivity growth. Price cap regulation of this sort has been used to control prices of long distance and local telephone service.

By the 1990s, the regulatory environment in the United States had changed dramatically. Many parts of the telecommunications industry had been deregulated, as had electric utilities in many states. Because scale economies had been largely exhausted, there was no reason to regard these firms as natural monopolies. In addition, technological change made entry by new firms relatively easy.

10.5 MONOPSONY

So far, our discussion of market power has focused entirely on the seller side of the market. Now we turn to the *buyer* side. We will see that if there are not too many buyers, they can also have market power and use it profitably to affect the price they pay for a product.

[11] Regulatory agencies often use a formula like the following to determine price:

$$P = AVC + (D + T + sK)/Q$$

where AVC is average variable cost, Q is output, s is the allowed "fair" rate of return, D is depreciation, T is taxes, and K is the firm's current capital stock.

First, a few terms.

- **Monopsony** refers to a market in which there is a single buyer.
- An **oligopsony** is a market with only a few buyers.
- With one or only a few buyers, some buyers may have **monopsony power**: a buyer's ability to affect the price of a good. Monopsony power enables the buyer to purchase a good for less than the price that would prevail in a competitive market.

oligopsony Market with only a few buyers.

monopsony power Buyer's ability to affect the price of a good.

Suppose you are trying to decide how much of a good to purchase. You could apply the basic marginal principle—keep purchasing units of the good until the last unit purchased gives additional value, or utility, just equal to the cost of that last unit. In other words, on the margin, additional benefit should just be offset by additional cost.

Let's look at this additional benefit and additional cost in more detail. We use the term **marginal value** to refer to the additional benefit from purchasing one more unit of a good. How do we determine marginal value? Recall from Chapter 4 that an individual demand curve determines marginal value, or marginal utility, as a function of the quantity purchased. Therefore, your *marginal value schedule* is your *demand* curve for the good. An individual's demand curve slopes downward because the marginal value obtained from buying one more unit of a good declines as the total quantity purchased increases.

marginal value Additional benefit derived from purchasing one more unit of a good.

In §4.1, we explain that as we move down along a demand curve, the value the consumer places on an additional unit of the good falls.

The additional cost of buying one more unit of a good is called the **marginal expenditure**. What that marginal expenditure is depends on whether you are a competitive buyer or a buyer with monopsony power. Suppose you are a competitive buyer—in other words, you have no influence over the price of the good. In that case, the cost of each unit you buy is the same no matter how many units you purchase; it is the market price of the good. Figure 10.13(a) illustrates this principle. The price you pay per unit is your **average expenditure** per unit, and it is the same for all units. But what is your *marginal expenditure* per unit? As a competitive buyer, your marginal expenditure is equal to your average expenditure, which in turn is equal to the market price of the good.

marginal expenditure Additional cost of buying one more unit of a good.

average expenditure Price paid per unit of a good.

Figure 10.13(a) also shows your marginal value schedule (i.e., your demand curve). How much of the good should you buy? You should buy until the marginal value of the last unit is just equal to the marginal expenditure on that unit. Thus you should purchase quantity Q^* at the intersection of the marginal expenditure and demand curves.

We introduced the concepts of marginal and average expenditure because they will make it easier to understand what happens when buyers have monopsony power. But before considering that situation, let's look at the analogy between competitive buyer conditions and competitive seller conditions. Figure 10.13(b) shows how a perfectly competitive seller decides how much to produce and sell. Because the seller takes the market price as given, both average and marginal revenue are equal to the price. The profit-maximizing quantity is at the intersection of the marginal revenue and marginal cost curves.

Now suppose that you are the *only* buyer of the good. Again you face a market supply curve, which tells you how much producers are willing to sell as a function of the price you pay. Should the quantity you purchase be at the point where your marginal value curve intersects the market supply curve? No. If you want to maximize your net benefit from purchasing the good, you should purchase a smaller quantity, which you will obtain at a lower price.

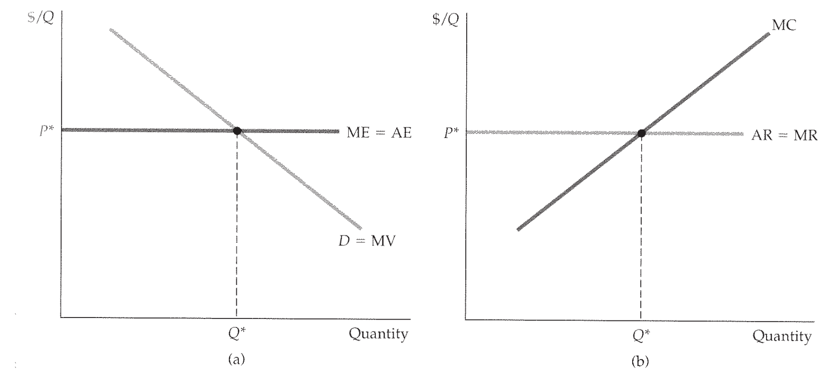

FIGURE 10.13 Competitive Buyer Compared to Competitive Seller

In (a), the competitive buyer takes market price P^* as given. Therefore, marginal expenditure and average expenditure are constant and equal; quantity purchased is found by equating price to marginal value (demand). In **(b)**, the competitive seller also takes price as given. Marginal revenue and average revenue are constant and equal; quantity sold is found by equating price to marginal cost.

To determine how much to buy, set the marginal value from the last unit purchased equal to the marginal expenditure on that unit.[12] Note, however, that the market supply curve is not the marginal expenditure curve. The market supply curve shows how much you must pay *per unit*, as a function of the total number of units you buy. In other words, the supply curve is the *average expenditure* curve. And because this average expenditure curve is upward sloping, the marginal expenditure curve must lie above it. The decision to buy an extra unit raises the price that must be paid for *all* units, not just the extra one.[13]

Figure 10.14 illustrates this principle. The optimal quantity for the monopsonist to buy, Q_m^*, is found at the intersection of the demand and marginal expenditure curves. The price that the monopsonist pays is found from the supply curve: It is the price P_m^* that brings forth the supply Q_m^*. Finally, note that this quantity Q_m^* is less, and the price P_m^* is lower, than the quantity and price that would prevail in a competitive market, Q_c and P_c.

[12]Mathematically, we can write the net benefit NB from the purchase as NB = $V - E$, where V is the value to the buyer of the purchase and E is the expenditure. Net benefit is maximized when $\Delta NB/\Delta Q = 0$. Then

$$\Delta NB/\Delta Q = \Delta V/\Delta Q - \Delta E/\Delta Q = MV - ME = 0$$

so that MV = ME.

[13]To obtain the marginal expenditure curve algebraically, write the supply curve with price on the left-hand side: $P = P(Q)$. Then total expenditure E is price times quantity, or $E = P(Q)Q$, and marginal expenditure is

$$ME = \Delta E/\Delta Q = P(Q) + Q(\Delta P/\Delta Q)$$

Because the supply curve is upward sloping, $\Delta P/\Delta Q$ is positive, and marginal expenditure is greater than average expenditure.

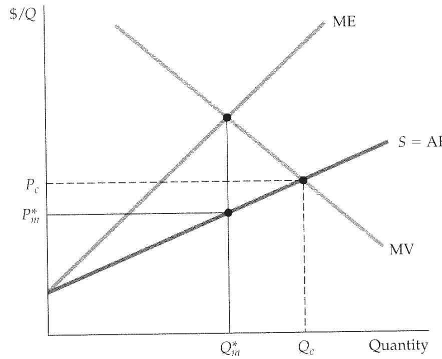

FIGURE 10.14 Monopsonist Buyer

The market supply curve is monopsonist's average expenditure curve AE. Because average expenditure is rising, marginal expenditure lies above it. The monopsonist purchases quantity Q^*_m, where marginal expenditure and marginal value (demand) intersect. The price paid per unit P^*_m is then found from the average expenditure (supply) curve. In a competitive market, price and quantity, P_c and Q_c, are both higher. They are found at the point where average expenditure (supply) and marginal value (demand) intersect.

Monopsony and Monopoly Compared

Monopsony is easier to understand if you compare it with monopoly. Figures 10.15(a) and 10.15(b) illustrate this comparison. Recall that a monopolist can charge a price above marginal cost because it faces a downward-sloping demand, or average revenue curve, so that marginal revenue is less than average revenue. Equating marginal cost with marginal revenue leads to a quantity Q^* that is less than what would be produced in a competitive market, and to a price P^* that is higher than the competitive price P_c.

The monopsony situation is exactly analogous. As Figure 10.15(b) illustrates, the monopsonist can purchase a good *at a price below its marginal value* because it faces an upward-sloping supply, or average expenditure, curve. Thus for a monopsonist, marginal expenditure is greater than average expenditure. Equating marginal value with marginal expenditure leads to a quantity Q^* that is less than what would be bought in a competitive market, and to a price P^* that is lower than the competitive price P_c.

10.6 MONOPSONY POWER

Much more common than pure monopsony are markets with only a few firms competing among themselves as buyers, so that each firm has some monopsony power. For example, the major U.S. automobile manufacturers compete with one another as buyers of tires. Because each of them accounts for a large share of the tire market, each has some monopsony power in that market. General

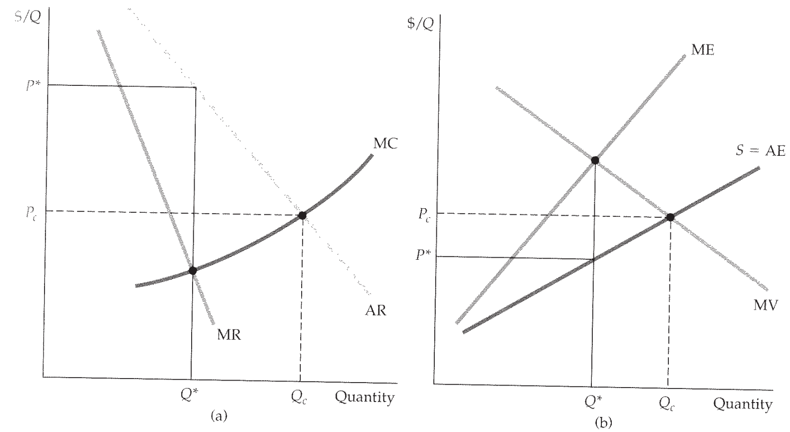

FIGURE 10.15 Monopoly and Monopsony

These diagrams show the close analogy between monopoly and monopsony. **(a)** The monopolist produces where marginal revenue intersects marginal cost. Average revenue exceeds marginal revenue, so that price exceeds marginal cost. **(b)** The monopsonist purchases up to the point where marginal expenditure intersects marginal value. Marginal expenditure exceeds average expenditure, so that marginal value exceeds price.

Motors, the largest, might be able to exert considerable monopsony power when contracting for supplies of tires (and other automotive parts).

In a competitive market, price and marginal value are equal. A buyer with monopsony power, however, can purchase a good at a price below marginal value. The extent to which price is marked down below marginal value depends on the elasticity of supply facing the buyer.[14] If supply is very elastic (E_S is large), the markdown will be small and the buyer will have little monopsony power. Conversely, if supply is very inelastic, the markdown will be large and the buyer will have considerable monopsony power. Figures 10.16(a) and 10.16(b) illustrate these two cases.

Sources of Monopsony Power

What determines the degree of monopsony power in a market? Again, we can draw analogies with monopoly and monopoly power. We saw that monopoly power depends on three things: the elasticity of market demand, the number of sellers in the market, and the way those sellers interact. Monopsony power depends on three similar things: The elasticity of market supply, the number of buyers in the market, and the way those buyers interact.

Elasticity of Market Supply A monopsonist benefits because it faces an upward-sloping supply curve, so that marginal expenditure exceeds average

[14] The exact relationship (analogous to equation (10.1)) is given by $(MV - P)/P = 1/E_S$. This equation follows because MV = ME and $ME = \Delta(PQ)/\Delta Q = P + Q(\Delta P/\Delta Q)$.

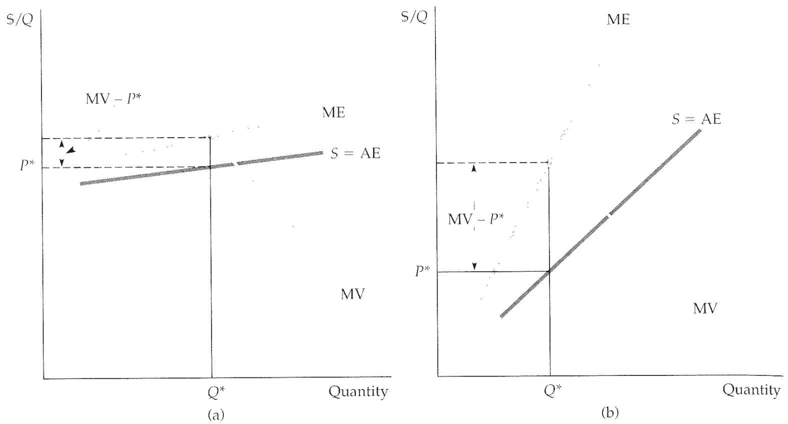

FIGURE 10.16 Monopsony Power: Elastic versus Inelastic Supply

Monopsony power depends on the elasticity of supply. When supply is elastic, as in **(a)**, marginal expenditure and average expenditure do not differ by much, so price is close to what it would be in a competitive market. The opposite is true when supply is inelastic, as in **(b)**.

expenditure. The less elastic the supply curve, the greater the difference between marginal expenditure and average expenditure and the more monopsony power the buyer enjoys. If only one buyer is in the market—a pure monopsonist—its monopsony power is completely determined by the elasticity of market supply. If supply is highly elastic, monopsony power is small and there is little gain in being the only buyer.

Number of Buyers Most markets have more than one buyer, and the number of buyers is an important determinant of monopsony power. When the number of buyers is very large, no single buyer can have much influence over price. Thus each buyer faces an extremely elastic supply curve, so that the market is almost completely competitive. The potential for monopsony power arises when the number of buyers is limited.

Interaction Among Buyers Finally, suppose three or four buyers are in the market. If those buyers compete aggressively, they will bid up the price close to their marginal value of the product, and will thus have little monopsony power. On the other hand, if those buyers compete less aggressively, or even collude, prices will not be bid up very much, and the buyers' degree of monopsony power might be nearly as high as if there were only one buyer.

So, as with monopoly power, there is no simple way to predict how much monopsony power buyers will have in a market. We can count the number of buyers, and we can often estimate the elasticity of supply, but that is not enough. Monopsony power also depends on the interaction among buyers, which can be more difficult to ascertain.

Because monopsony power results in lower prices and lower quantities purchased, we would expect it to make the buyer better off and sellers worse off. But suppose we value the welfare of buyers and sellers equally. How is aggregate welfare affected by monopsony power?

We can find out by comparing the buyer and seller surplus that results from a competitive market to the surplus that results when a monopsonist is the sole buyer. Figure 10.17 shows the average and marginal expenditure curves and marginal value curve for the monopsonist. The monopsonist's net benefit is maximized by purchasing a quantity Q_m at a price P_m such that marginal value equals marginal expenditure. In a competitive market, price equals marginal value. Thus the competitive price and quantity, P_c and Q_c, are found where the average expenditure and marginal value curves intersect. Now let's see how surplus changes if we move from the competitive price and quantity, P_c and Q_c, to the monopsony price and quantity, P_m and Q_m.

With monopsony, the price is lower and less is sold. Because of the lower price, sellers lose an amount of surplus given by rectangle A. In addition, sellers lose the surplus given by triangle C because of the reduced sales. The total loss of producer (seller) surplus is therefore $A + C$. By buying at a lower price, the buyer gains the surplus given by rectangle A. However, the buyer buys less, Q_m instead of Q_c, and so loses the surplus given by triangle B. The total gain in surplus to the buyer is therefore $A - B$. Altogether, there is a net loss of surplus given by $B + C$. This is the *deadweight loss from monopsony power*. Even if the monopsonist's gains were taxed away and redistributed to the producers, there would be an inefficiency because output would be lower than under competition. The deadweight loss is the social cost of this inefficiency.

Note the similarity with the deadweight loss from monopoly power discussed in §10.4.

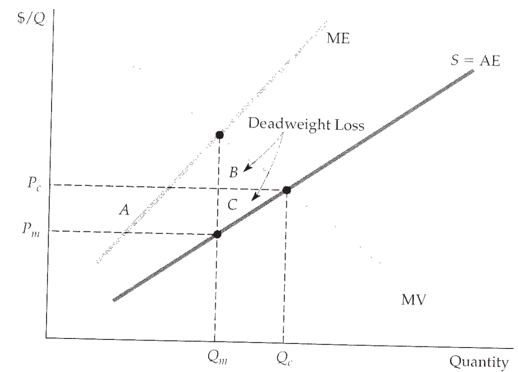

FIGURE 10.17 Deadweight Loss from Monopsony Power

The shaded rectangle and triangles show changes in buyer and seller surplus when moving from competitive price and quantity, P_c and Q_c, to the monopsonist's price and quantity, P_m and Q_m. Because both price and quantity are lower, there is an increase in buyer (consumer) surplus given by $A - B$. Producer surplus falls by $A + C$, so there is a deadweight loss given by triangles B and C.

Bilateral Monopoly

• **bilateral monopoly** Market with only one seller and one buyer.

What happens when a monopolist meets a monopsonist? It's hard to say. We call a market with only one seller and only one buyer a **bilateral monopoly**. If you think about such a market, you'll see why it is difficult to predict the price and quantity. Both the buyer and the seller are in a bargaining situation. Unfortunately, no simple rule determines which, if either, will get the better part of the bargain. One party might have more time and patience, or might be able to convince the other party that it will walk away if the price is too low or too high.

Bilateral monopoly is rare. Markets in which a few producers have some monopoly power and sell to a few buyers who have some monopsony power are more common. Although bargaining may still be involved, we can apply a rough principle here: *Monopsony power and monopoly power will tend to counteract each other.* In other words, the monopsony power of buyers will reduce the effective monopoly power of sellers, and vice versa. This tendency does not mean that the market will end up looking perfectly competitive; if, for example, monopoly power is large and monopsony power small, the residual monopoly power would still be significant. But in general, monopsony power will push price closer to marginal cost, and monopoly power will push price closer to marginal value.

EXAMPLE 10.4 Monopsony Power in U.S. Manufacturing

Monopoly power, as measured by the price-cost margin $(P - MC)/P$, varies considerably across manufacturing industries in the United States. Some industries have price-cost margins close to zero, while in others margins are as high as 0.4 or 0.5. These variations are due in part to differences in the determinants of monopoly power: In some industries, market demand is more elastic than in others; some industries have more sellers than others; and in some industries, sellers compete more aggressively than in others. But something else can help explain these variations in monopoly power—differences in monopsony power among the firms' customers.

The role of monopsony power was investigated in a statistical study of 327 U.S. manufacturing industries.[15] The study sought to determine the extent to which variations in price—cost margins could be attributed to variations in monopsony power by buyers in each industry. Although the degree of buyers' monopsony power could not be measured directly, data were available for variables that help determine monopsony power, such as buyer concentration (the fraction of total sales going to the three or four largest firms) and the average annual size of buyers' orders.

The study found that buyers' monopsony power had an important effect on the price—cost margins of sellers and could significantly reduce any monopoly power that sellers might otherwise have. Take, for example, the concentration of buyers, an important determinant of monopsony power. In industries where only four or five buyers account for all or nearly all sales, the price-cost margins of sellers would on average be as much as 10 percentage points lower than in comparable industries with hundreds of buyers accounting for sales.

A good example of monopsony power in manufacturing is the market for automobile parts and components, such as brakes and radiators. Each major car producer in the United States typically buys an individual part from at least three, and often as many as a dozen, suppliers. In addition, for a standardized product, such as brakes, each automobile company usually produces part of its needs itself, so that it is not totally reliant on outside firms. This puts companies like General Motors and Ford in an excellent bargaining position with respect to their suppliers. Each supplier must compete for sales against five or 10 other suppliers, but each can sell to only a few buyers. For a specialized part, a single auto company may be the *only* buyer. As a result, the automobile companies have considerable monopsony power.

This monopsony power becomes evident from the conditions under which suppliers must operate. To obtain a sales contract, a supplier must have a track record of reliability, in terms of both product quality and ability to meet tight delivery schedules. Suppliers are also often required to respond to changes in volume as auto sales and production levels fluctuate. Finally, pricing negotiations are notoriously difficult; a potential supplier will sometimes lose a contract because its bid is a penny per item higher than those of its competitors. Not surprisingly, producers of parts and components usually have little or no monopoly power.

[15]The study was by Steven H. Lustgarten, "The Impact of Buyer Concentration in Manufacturing Industries," *Review of Economics and Statistics* 57 (May 1975): 125—32.

10.7 LIMITING MARKET POWER: THE ANTITRUST LAWS

We have seen that market power—whether wielded by sellers or buyers—harms potential purchasers who could have bought at competitive prices. In addition, market power reduces output, which leads to a deadweight loss. Excessive market power also raises problems of equity and fairness: If a firm has significant monopoly power, it will profit at the expense of consumers. In theory, a firm's excess profits could be taxed away and redistributed to the buyers of its products, but such a redistribution is often impractical. It is difficult to determine what portion of a firm's profit is attributable to monopoly power, and it is even more difficult to locate all the buyers and reimburse them in proportion to their purchases.

How, then, can society limit market power and prevent it from being used anticompetitively? For a natural monopoly, such as an electric utility company, direct price regulation is the answer. But more generally, the answer is to prevent firms from acquiring excessive market power in the first place, and to limit the use of that power if it is acquired. In the United States, this is done via the **antitrust laws:** a set of rules and regulations designed to promote a competitive economy by prohibiting actions that restrain, or are likely to restrain, competition, and by restricting the forms of allowable market structure.

• **antitrust laws** Rules and regulations prohibiting actions that restrain, or are likely to restrain, competition.

Monopoly power can arise in a number of ways, each of which is covered by the antitrust laws. Section 1 of the Sherman Act (which was passed in 1890) prohibits contracts, combinations, or conspiracies in restraint of trade. One obvious example of an illegal combination is an explicit agreement among producers to restrict their outputs and/or "fix" price above the competitive level. There have been numerous instances of such illegal combinations. For example:

- In 1996, Archer Daniels Midland Company (ADM) and two other major producers of lysine (an animal feed additive) pleaded guilty to criminal charges of price fixing. In 1999, three ADM executives were sentenced to prison terms ranging from two to three years for their roles in the price-fixing scheme.[16]
- In 1999, four of the world's largest drug and chemical companies—Roche A.G. of Switzerland, BASF A.G. of Germany, Rhone-Poulenc of France, and Takeda Chemical Industries of Japan—were charged by the U.S. Department of Justice with taking part in a global conspiracy to fix the prices of vitamins sold in the United States. The companies pleaded guilty to price fixing and agreed to pay fines totaling more than $1 billion.[17]
- In 2002, the U.S. Department of Justice began an investigation of price fixing by DRAM (dynamic access random memory) producers. By 2006, five manufacturers—Hynix, Infineon, Micron Technology, Samsung, and Elpida—had pled guilty for participating in an international price-fixing scheme. As part of these pleas, the companies agreed to pay fines totaling close to $1 billion to the DOJ, and several executives received prison sentences.

• **parallel conduct** Form of implicit collusion in which one firm consistently follows actions of another.

Two firms need not meet or talk on the telephone to violate Section 1 of the Sherman Act; *implicit* collusion in the form of **parallel conduct** can also be construed as violating the law. For example, if Firm *B* consistently follows Firm *A*'s pricing (parallel pricing), and if the firm's conduct is contrary to what one would expect companies to do in the absence of collusion (such as raising prices in the face of decreased demand and over-supply), an implicit understanding may be inferred.[18]

Section 2 of the Sherman Act makes it illegal to monopolize or to attempt to monopolize a market and prohibits conspiracies that result in monopolization. The Clayton Act (1914) did much to pinpoint the kinds of practices that are likely to be anticompetitive. For example, the act makes it unlawful for a firm with a large market share to require the buyer or lessor of a good not to buy from a competitor. It also makes it illegal to engage in **predatory pricing**—pricing designed to drive current competitors out of business and to discourage new entrants (so that the predatory firm can enjoy higher prices in the future).

• **predatory pricing** Practice of pricing to drive current competitors out of business and to discourage new entrants in a market so that a firm can enjoy higher future profits.

Monopoly power can also be achieved by a merger of firms into a larger and more dominant firm, or by one firm acquiring or taking control of another firm by purchasing its stock. The Clayton Act prohibits mergers and acquisitions if they "substantially lessen competition" or "tend to create a monopoly."

[16]In the lysine case, proof of the conspiracy came in part from tapes of meetings at which prices were set and market shares divided up. At one meeting with executives from Ajinimoto Company of Japan, another lysine producer, James Randall, then the president of ADM, said, "We have a saying at this company. Our competitors are our friends and our customers are our enemies." See "Video Tapes Take Star Role at Archer Daniels Trial," *New York Times*, August 4, 1998; "Three Sentenced in Archer Daniels Midland Case," *New York Times*, July 10, 1999. In 1993, ADM and three other firms were also charged with fixing carbon dioxide prices.

[17]"Tearing Down the Facades of 'Vitamins Inc.'," *New York Times*, October 10, 1999.

[18]The Sherman Act applies to all firms that do business in the United States (to the extent that a conspiracy to restrain trade could affect U.S. markets). However, foreign governments (or firms operating under their government's control) are not subject to the act, so OPEC need not fear the wrath of the Justice Department. Also, firms *can* collude with respect to *exports*. The Webb-Pomerene Act (1918) allows price fixing and related collusion with respect to export markets, *as long as domestic markets are unaffected by such collusion*. Firms operating in this manner must form a "Webb-Pomerene Association" and register it with the government.

The antitrust laws also limit possible anticompetitive conduct by firms in other ways. For example, the Clayton Act, as amended by the Robinson-Patman Act (1936), makes it illegal to discriminate by charging buyers of essentially the same product different prices if those price differences are likely to injure competition. Even then, firms are not liable if they can show that the price differences were necessary to meet competition. (As we will see in the next chapter, price discrimination is a common practice. It becomes the target of antitrust action only when buyers suffer economic damages and competition is reduced.)

Another important component of the antitrust laws is the Federal Trade Commission Act (1914, amended in 1938, 1973, 1975), which created the Federal Trade Commission (FTC). This act supplements the Sherman and Clayton acts by fostering competition through a whole set of prohibitions against unfair and anticompetitive practices, such as deceptive advertising and labeling, agreements with retailers to exclude competing brands, and so on. Because these prohibitions are interpreted and enforced in administrative proceedings before the FTC, the act provides broad powers that reach further than those of other antitrust laws.

The antitrust laws are actually phrased vaguely in terms of what is and what is not allowed. They are intended to provide a general statutory framework to give the Justice Department, the FTC, and the courts wide discretion in interpreting and applying them. This approach is important because it is difficult to know in advance what might be an impediment to competition. Such ambiguity creates a need for common law (i.e., the practice whereby courts interpret statutes) and supplemental provisions and rulings (e.g., by the FTC or the Justice Department).

Enforcement of the Antitrust Laws

The antitrust laws are enforced in three ways:

1. *Through the Antitrust Division of the Department of Justice.* As an arm of the executive branch, its enforcement policies closely reflect the view of the administration in power. Responding to an external complaint or an internal study, the department can institute a criminal proceeding, bring a civil suit, or both. The result of a criminal action can be fines for the corporation and fines or jail sentences for individuals. For example, individuals who conspire to fix prices or rig bids can be charged with a felony and, if found guilty, may be sentenced to jail—something to remember if you are planning to parlay your knowledge of microeconomics into a successful business career! Losing a civil action forces a corporation to cease its anticompetitive practices and often to pay damages.
2. *Through the administrative procedures of the Federal Trade Commission.* Again, action can result from an external complaint or from the FTC's own initiative. Should the FTC decide that action is required, it can either request a voluntary understanding to comply with the law or seek a formal commission order requiring compliance.
3. *Through private proceedings.* Individuals or companies can sue for *treble (three-fold) damages* inflicted on their businesses or property. The prospect of treble damages can be a strong deterrent to would-be violators. Individuals or companies can also ask the courts for injunctions to force wrongdoers to cease anticompetitive actions.

U.S. antitrust laws are more stringent and far-reaching than those of most other countries. In fact, some people have argued that they have prevented American industry from competing effectively in international markets. The laws certainly constrain American business and may at times have put American firms at a disadvantage in world markets. But this criticism must be weighed against their benefits: Antitrust laws have been crucial for maintaining competition, and competition is essential for economic efficiency, innovation, and growth.

As the European Union has grown, its methods of antitrust enforcement have evolved. The responsibility for the enforcement of antitrust concerns that involve two or more member states resides in a single entity, the Competition Directorate, located in Brussels. Separate and distinct antitrust authorities within individual member states are responsible for those issues whose effects are felt largely or entirely within particular countries.

At first glance, the antitrust laws of the European Union are quite similar to those of the United States. Article 81 of the Treaty of the European Community concerns restraints of trade, much like Section 1 of the Sherman Act. Article 82, which focuses on abuses of market power by *dominant* firms, is similar in many ways to Section 2 of the Sherman Act. Finally, with respect to mergers, the European Merger Control Act is similar in spirit to Section 7 of the Clayton Act.

Nevertheless, there remain a number of procedural and substantive differences between antitrust laws in Europe and the United States. Merger evaluations typically are conducted more quickly in Europe, and it is easier in practice to prove that a European firm is dominant than it is to show that a U.S. firm has monopoly power. Both the European Union and the U.S. have been actively enforcing laws against price fixing, but Europe imposes only civil penalties, whereas the U.S. can impose prison sentences as well as fines.

In 1981 and early 1982, American Airlines and Braniff Airways were competing fiercely with each other for passengers. A fare war broke out as the firms undercut each other's prices to capture market share. On February 21, 1982, Robert Crandall, president and CEO of American, made a phone call to Howard Putnam, president and chief executive of Braniff. To Crandall's later surprise, the call had been taped. It went like this:[19]

Crandall: I think it's dumb as hell for Christ's sake, all right, to sit here and pound the @!#$%&! out of each other and neither one of us making a @!#$%&! dime.

Putnam: Well . . .

Crandall: I mean, you know, @!#$%&!, what the hell is the point of it?

Putnam: But if you're going to overlay every route of American's on top of every route that Braniff has—I just can't sit here and allow you to bury us without giving our best effort.

Crandall: Oh sure, but Eastern and Delta do the same thing in Atlanta and have for years.

Putnam: Do you have a suggestion for me?

[19]According to the *New York Times*, February 24, 1983.

Crandall: Yes, I have a suggestion for you. Raise your @!#$%&! fares 20 percent. I'll raise mine the next morning.

Putnam: Robert, we. . .

Crandall: You'll make more money and I will, too.

Putnam: We can't talk about pricing!

Crandall: Oh @!#$%&!, Howard. We can talk about any @!#$%&! thing we want to talk about.

Crandall was wrong. Corporate executives cannot talk about anything they want. Talking about prices and agreeing to fix them is a clear violation of Section 1 of the Sherman Act. Putnam must have known this because he promptly rejected Crandall's suggestion. After learning about the call, the Justice Department filed a suit accusing Crandall of violating the antitrust laws by proposing to fix prices.

However, *proposing* to fix prices is not enough to violate Section 1 of the Sherman Act: For the law to be violated, the two parties must *agree* to collude. Therefore, because Putnam had rejected Crandall's proposal, Section 1 was not violated. The court later ruled, however, that a proposal to fix prices could be an attempt to monopolize part of the airline industry and, if so, would violate Section 2 of the Sherman Act. American Airlines promised the Justice Department never again to engage in such activity.

Over the past decade, Microsoft Corporation has grown to become the largest computer software company in the world. Its Windows operating system has over 94 percent of the worldwide market for personal computer operating systems. Microsoft also dominates the office productivity market: Its Office Suite, which includes Word (word processing), Excel (spreadsheets), and Powerpoint (presentations), held over a 95-percent worldwide market share in 2006.

Microsoft's incredible success has been due in good part to the creative technological and marketing decisions of the company and its CEO, Bill Gates. Is there anything wrong as a matter of either economics or law with being so successful and dominant? It all depends. Under the antitrust laws, efforts by firms to restrain trade or to engage in activities that inappropriately maintain monopolies are illegal. Did Microsoft engage in anticompetitive, illegal practices?

The U.S. Government said yes; Microsoft disagreed. In October 1998, the Antitrust Division of the U.S. Department of Justice (DOJ) put Microsoft's behavior to the test: It filed suit, raising a broad set of issues that created the most significant antitrust law suit of the past two decades. The ensuing trial ended in June 1999, but it wasn't until early in 2003 that a settlement between the government and Microsoft was finalized. Here is a brief road map of some of the DOJ's major claims and Microsoft's responses.

- **DOJ claim:** Microsoft has a great deal of market power in the market for PC operating systems—enough to meet the legal definition of monopoly power.
MS response: Microsoft does not meet the legal test for monopoly power because it faces significant threats from potential competitors that offer or will offer platforms to compete with Windows.
- **DOJ claim:** Microsoft viewed Netscape's Internet browser (Netscape Navigator) as a threat to its monopoly over the PC operating system market. The threat existed because Netscape's browser includes Sun's Java software, which can run programs that have been written for *any* operating system, including those that compete with Windows, such as Apple, Unix, and Linux. In violation of Section 1 of the Sherman Act, Microsoft entered into exclusionary agreements with computer manufacturers and Internet service providers with the objective of raising the cost to Netscape of making its browser available to consumers. This action impaired Netscape's ability to compete fairly with Microsoft's Internet Explorer for the browser business.
MS response: The contracts were not unduly restrictive. In any case, Microsoft unilaterally agreed to stop most of them.
- **DOJ claim:** In violation of Section 2 of the Sherman Act, Microsoft engaged in practices designed to maintain its monopoly in the market for desktop PC operating systems. Most importantly, it tied its browser to the Windows 98 operating system, even though doing so was technically unnecessary and provides little or no benefit to consumers. This action was predatory because it made it difficult or impossible for Netscape and other firms to successfully offer competing products.
MS response: There are benefits to incorporating the browser functionality into the operating system. Not being allowed to integrate new functionality into an operating system will discourage innovation. Offering consumers a choice between separate or integrated browsers would cause confusion in the marketplace.
- **DOJ claim:** In violation of Section 2 of the Sherman Act, Microsoft attempted to divide the browser business with Netscape and engaged in similar conduct with both Apple Computer and Intel.
MS response: Microsoft's meetings with Netscape, Apple, and Intel were for valid business reasons. Indeed, it is useful for consumers and firms to agree on common standards and protocols in developing computer software.

These are some of the highlights of an eight-month trial that was hard-fought on a range of economic issues. The District Court reached its findings regarding the facts of the case in November 1999 and the legal conclusions in April 2000. It found that Microsoft did have monopoly power in the market for PC operating systems. The Court concluded further that Microsoft had viewed Netscape as a threat and that in responding to that threat, it had engaged in a series of anticompetitive acts to protect and extend its operating system monopoly. The court deemed these actions to violate Section 2 of the Sherman Act. However, the Court also found that the exclusionary agreements with computer manufacturers and Internet service providers had not foreclosed competition sufficiently to violate Section 1 of the Sherman Act. Microsoft's appeal to the Circuit Court of Appeals for the District of Columbia was decided in June 2001. The Appellate Court supported the District Court's conclusions that Microsoft was a monopoly and had engaged in anticompetitive practices to protect that monopoly. However, the Court left undecided whether including Internet Explorer in the operating system was itself illegal.

Since this decision, the DOJ and Microsoft agreed to settle the case. Among other things, the agreement required Microsoft (1) to give computer manufacturers the ability to offer its operating system without Internet Explorer and (2) to include competing browser programs when loading the Windows operating system onto the machines they sell. Microsoft also agreed to a program that would monitor its compliance with the terms of the settlement. Despite opposition from critics who believed the remedy insufficient, the settlement was approved by the Appellate Court in 2004, putting an end to this landmark antitrust case in the United States.

Microsoft's problems did not end with the U.S. settlement, however. In 2004, the European Commission ordered Microsoft to pay $610 million in fines for its anticompetitive practices and to produce a version of Windows without the Windows Media Player to be sold alongside its standard editions. In addition, numerous private lawsuits were brought in the United States, with most settling for substantial sums of money.

SUMMARY

1. Market power is the ability of sellers or buyers to affect the price of a good.
2. Market power comes in two forms. When sellers charge a price that is above marginal cost, we say that they have monopoly power, which we measure by the extent to which price exceeds marginal cost. When buyers can obtain a price below their marginal value of the good, we say they have monopsony power, which we measure by the extent to which marginal value exceeds price.
3. Monopoly power is determined in part by the number of firms competing in a market. If there is only one firm—a pure monopoly—monopoly power depends entirely on the elasticity of market demand. The less elastic the demand, the more monopoly power the firm will have. When there are several firms, monopoly power also depends on how the firms interact. The more aggressively they compete, the less monopoly power each firm will have.
4. Monopsony power is determined in part by the number of buyers in a market. If there is only one buyer—a pure monopsony—monopsony power depends on the elasticity of market supply. The less elastic the supply, the more monopsony power the buyer will have. When there are several buyers, monopsony power also depends on how aggressively they compete for supplies.
5. Market power can impose costs on society. Because monopoly and monopsony power both cause production to fall below the competitive level, there is a deadweight loss of consumer and producer surplus. There can be additional social costs from rent seeking.
6. Sometimes, scale economies make pure monopoly desirable. But the government will still want to regulate price to maximize social welfare.
7. More generally, we rely on the antitrust laws to prevent firms from obtaining excessive market power.

QUESTIONS FOR REVIEW

1. A monopolist is producing at a point at which marginal cost exceeds marginal revenue. How should it adjust its output to increase profit?
2. We write the percentage markup of price over marginal cost as $(P - MC)/P$. For a profit-maximizing monopolist, how does this markup depend on the elasticity of demand? Why can this markup be viewed as a measure of monopoly power?
3. Why is there no market supply curve under conditions of monopoly?
4. Why might a firm have monopoly power even if it is not the only producer in the market?

5. What are some of the different types of barriers to entry that give rise to monopoly power? Give an example of each.
6. What factors determine the amount of monopoly power an individual firm is likely to have? Explain each one briefly.
7. Why is there a social cost to monopoly power? If the gains to producers from monopoly power could be redistributed to consumers, would the social cost of monopoly power be eliminated? Explain briefly.
8. Why will a monopolist's output increase if the government forces it to lower its price? If the government wants to set a price ceiling that maximizes the monopolist's output, what price should it set?
9. How should a monopsonist decide how much of a product to buy? Will it buy more or less than a competitive buyer? Explain briefly.
10. What is meant by the term "monopsony power"? Why might a firm have monopsony power even if it is not the only buyer in the market?
11. What are some sources of monopsony power? What determines the amount of monopsony power an individual firm is likely to have?
12. Why is there a social cost to monopsony power? If the gains to buyers from monopsony power could be redistributed to sellers, would the social cost of monopsony power be eliminated? Explain briefly.
13. How do the antitrust laws limit market power in the United States? Give examples of major provisions of these laws.
14. Explain briefly how the U.S. antitrust laws are actually enforced.

EXERCISES

1. Will an increase in the demand for a monopolist's product always result in a higher price? Explain. Will an increase in the supply facing a monopsonist buyer always result in a lower price? Explain.
2. Caterpillar Tractor, one of the largest producers of farm machinery in the world, has hired you to advise it on pricing policy. One of the things the company would like to know is how much a 5-percent increase in price is likely to reduce sales. What would you need to know to help the company with this problem? Explain why these facts are important.
3. A monopolist firm faces a demand with constant elasticity of –2.0. It has a constant marginal cost of $20 per unit and sets a price to maximize profit. If marginal cost should increase by 25 percent, would the price charged also rise by 25 percent?
4. A firm faces the following average revenue (demand) curve:

$$P = 120 - 0.02Q$$

where Q is weekly production and P is price, measured in cents per unit. The firm's cost function is given by $C = 60Q + 25{,}000$. Assume that the firm maximizes profits.
   a. What is the level of production, price, and total profit per week?
   b. If the government decides to levy a tax of 14 cents per unit on this product, what will be the new level of production, price, and profit?
5. The following table shows the demand curve facing a monopolist who produces at a constant marginal cost of $10:

Price	Quantity
18	0
16	4
14	8
12	12
10	16
8	20
6	24
4	28
2	32
0	36

   a. Calculate the firm's marginal revenue curve.
   b. What are the firm's profit-maximizing output and price? What is its profit?
   c. What would the equilibrium price and quantity be in a competitive industry?
   d. What would the social gain be if this monopolist were forced to produce and price at the competitive equilibrium? Who would gain and lose as a result?
6. Suppose that an industry is characterized as follows:

$C = 100 + 2q^2$	each firm's total cost function
$MC = 4q$	firm's marginal cost function
$P = 90 - 2Q$	industry demand curve
$MR = 90 - 4Q$	industry marginal revenue curve

   a. If there is *only one firm* in the industry, find the monopoly price, quantity, and level of profit.
   b. Find the price, quantity, and level of profit if the industry is competitive.
   c. Graphically illustrate the demand curve, marginal revenue curve, marginal cost curve, and average cost curve. Identify the difference between the profit level of the monopoly and the profit level of the competitive industry in two different ways. Verify that the two are numerically equivalent.
7. Suppose a profit-maximizing monopolist is producing 800 units of output and is charging a price of $40 per unit.
   a. If the elasticity of demand for the product is –2, find the marginal cost of the last unit produced.
   b. What is the firm's percentage markup of price over marginal cost?
   c. Suppose that the average cost of the last unit produced is $15 and the firm's fixed cost is $2000. Find the firm's profit.
8. A firm has two factories, for which costs are given by:

$$\text{Factory \#1: } C_1(Q_1) = 10Q_1^2$$

$$\text{Factory \#2: } C_2(Q_2) = 20Q_2^2$$

The firm faces the following demand curve:

$$P = 700 - 5Q$$

where Q is total output—i.e., $Q = Q_1 + Q_2$.
   a. On a diagram, draw the marginal cost curves for the two factories, the average and marginal revenue curves, and the total marginal cost curve (i.e., the marginal cost of producing $Q = Q_1 + Q_2$). Indicate the profit-maximizing output for each factory, total output, and price.
   b. Calculate the values of Q_1, Q_2, Q, and P that maximize profit.
   c. Suppose that labor costs increase in Factory 1 but not in Factory 2. How should the firm adjust (i.e., raise, lower, or leave unchanged) the following: Output in Factory 1? Output in Factory 2? Total output? Price?
9. A drug company has a monopoly on a new patented medicine. The product can be made in either of two plants. The costs of production for the two plants are $MC_1 = 20 + 2Q_1$ and $MC_2 = 10 + 5Q_2$. The firm's estimate of demand for the product is $P = 20 - 3(Q_1 + Q_2)$. How much should the firm plan to produce in each plant? At what price should it plan to sell the product?
10. One of the more important antitrust cases of the 20th century involved the Aluminum Company of America (Alcoa) in 1945. At that time, Alcoa controlled about 90 percent of primary aluminum production in the United States, and the company had been accused of monopolizing the aluminum market. In its defense, Alcoa argued that although it indeed controlled a large fraction of the primary market, secondary aluminum (i.e., aluminum produced from the recycling of scrap) accounted for roughly 30 percent of the total supply of aluminum and that many competitive firms were engaged in recycling. Therefore, Alcoa argued, it did not have much monopoly power.
   a. Provide a clear argument *in favor* of Alcoa's position.
   b. Provide a clear argument *against* Alcoa's position.
   c. The 1945 decision by Judge Learned Hand has been called "one of the most celebrated judicial opinions of our time." Do you know what Judge Hand's ruling was?
11. A monopolist faces the demand curve $P = 11 - Q$, where P is measured in dollars per unit and Q in thousands of units. The monopolist has a constant average cost of $6 per unit.
   a. Draw the average and marginal revenue curves and the average and marginal cost curves. What are the monopolist's profit-maximizing price and quantity? What is the resulting profit? Calculate the firm's degree of monopoly power using the Lerner index.
   b. A government regulatory agency sets a price ceiling of $7 per unit. What quantity will be produced, and what will the firm's profit be? What happens to the degree of monopoly power?
   c. What price ceiling yields the largest level of output? What is that level of output? What is the firm's degree of monopoly power at this price?
12. Michelle's Monopoly Mutant Turtles (MMMT) has the exclusive right to sell Mutant Turtle t-shirts in the United States. The demand for these t-shirts is $Q = 10{,}000/P^2$. The firm's short-run cost is $SRTC = 2000 + 5Q$, and its long-run cost is $LRTC = 6Q$.
   a. What price should MMMT charge to maximize profit in the short run? What quantity does it sell, and how much profit does it make? Would it be better off shutting down in the short run?
   b. What price should MMMT charge in the long run? What quantity does it sell and how much profit does it make? Would it be better off shutting down in the long run?
   c. Can we expect MMMT to have lower marginal cost in the short run than in the long run? Explain why.
13. You produce widgets for sale in a perfectly competitive market at a market price of $10 per widget. Your widgets are manufactured in two plants, one in Massachusetts and the other in Connecticut. Because of labor problems in Connecticut, you are forced to raise wages there, so that marginal costs in that plant increase. In response to this, should you shift production and produce more in your Massachusetts plant?
14. The employment of teaching assistants (TAs) by major universities can be characterized as a monopsony. Suppose the demand for TAs is $W = 30{,}000 - 125n$,

where W is the wage (as an annual salary) and n is the number of TAs hired. The supply of TAs is given by $W = 1000 + 75n$.

a. If the university takes advantage of its monopsonist position, how many TAs will it hire? What wage will it pay?

b. If, instead, the university faced an infinite supply of TAs at the annual wage level of \$10,000, how many TAs would it hire?

*15. Dayna's Doorstops, Inc. (DD) is a monopolist in the doorstop industry. Its cost is $C = 100 - 5Q + Q^2$, and demand is $P = 55 - 2Q$.

a. What price should DD set to maximize profit? What output does the firm produce? How much profit and consumer surplus does DD generate?

b. What would output be if DD acted like a perfect competitor and set $MC = P$? What profit and consumer surplus would then be generated?

c. What is the deadweight loss from monopoly power in part (a)?

d. Suppose the government, concerned about the high price of doorstops, sets a maximum price at \$27. How does this affect price, quantity, consumer surplus, and DD's profit? What is the resulting deadweight loss?

e. Now suppose the government sets the maximum price at \$23. How does this decision affect price, quantity, consumer surplus, DD's profit, and deadweight loss?

f. Finally, consider a maximum price of \$12. What will this do to quantity, consumer surplus, profit, and deadweight loss?

*16. There are 10 households in Lake Wobegon, Minnesota, each with a demand for electricity of $Q = 50 - P$. Lake Wobegon Electric's (LWE) cost of producing electricity is $TC = 500 + Q$.

a. If the regulators of LWE want to make sure that there is no deadweight loss in this market, what price will they force LWE to charge? What will output be in that case? Calculate consumer surplus and LWE's profit with that price.

b. If regulators want to ensure that LWE doesn't lose money, what is the lowest price they can impose? Calculate output, consumer surplus, and profit. Is there any deadweight loss?

c. Kristina knows that deadweight loss is something that this small town can do without. She suggests that each household be required to pay a fixed amount just to receive any electricity at all, and then a per-unit charge for electricity. Then LWE can break even while charging the price calculated in part (a). What fixed amount would each household have to pay for Kristina's plan to work? Why can you be sure that no household will choose instead to refuse the payment and go without electricity?

17. A certain town in the Midwest obtains all of its electricity from one company, Northstar Electric. Although the company is a monopoly, it is owned by the citizens of the town, all of whom split the profits equally at the end of each year. The CEO of the company claims that because all of the profits will be given back to the citizens, it makes economic sense to charge a monopoly price for electricity. True or false? Explain.

18. A monopolist faces the following demand curve:

$$Q = 144/P^2$$

where Q is the quantity demanded and P is price. Its *average variable* cost is

$$AVC = Q^{1/2}$$

and its *fixed cost* is 5.

a. What are its profit-maximizing price and quantity? What is the resulting profit?

b. Suppose the government regulates the price to be no greater than \$4 per unit. How much will the monopolist produce? What will its profit be?

c. Suppose the government wants to set a ceiling price that induces the monopolist to produce the largest possible output. What price will accomplish this goal?

Pricing with Market Power

11

As we explained in Chapter 10, market power is quite common. Many industries have only a few producers, so that each producer has some monopoly power. And many firms, as buyers of raw materials, labor, or specialized capital goods, have some monopsony power in the markets for these factor inputs. The problem faced by the managers of these firms is *how to use their market power most effectively*. They must decide how to set prices, choose quantities of factor inputs, and determine output in both the short and long run to maximize profit.

Managers of firms with market power have a harder job than those who manage perfectly competitive firms. A firm that is perfectly competitive in output markets has no influence over market price. As a result, its managers need worry only about the cost side of the firm's operations, choosing output so that price is equal to marginal cost. But the managers of a firm with monopoly power must also worry about the characteristics of demand. Even if they set a single price for the firm's output, they must obtain at least a rough estimate of the elasticity of demand to determine what that price (and corresponding output level) should be. Furthermore, firms can often do much better by using a more complicated pricing strategy—for example, charging different prices to different customers. To design such pricing strategies, managers need ingenuity and even more information about demand.

This chapter explains how firms with market power set prices. We begin with the basic objective of every pricing strategy: capturing consumer surplus and converting it into additional profit for the firm. Then we discuss how this goal can be achieved using *price discrimination*—charging different prices to different customers, sometimes for the same product and sometimes for small variations in the product. Because price discrimination is widely practiced in one form or another, it is important to understand how it works.

Next, we discuss the *two-part tariff*—requiring customers to pay in advance for the right to purchase units of a good at a later time (and at additional cost). The classic example of this is an amusement park, where customers pay a fee to enter and then additional fees for each ride. Although amusement parks may seem like a rather specialized market, there are many other examples of two-part tariffs: the price of a Gillette razor, which gives the owner the opportunity to purchase Gillette razor blades; a tennis club, where members pay an annual fee and then an hourly rate for court time; or the monthly subscription cost of long-distance telephone service, which gives users the opportunity to make long-distance calls, paying by the minute as they do so.

We will also discuss *bundling*, a pricing strategy that involves tying products together and selling them as a package. For example: a personal computer that comes bundled with several software packages; a

CHAPTER OUTLINE

LIST OF EXAMPLES

one-week vacation in which the airfare, rental car, and hotel are bundled and sold at a single package price; or a luxury car, in which the sun roof, power windows, and leather seats are "standard" features.

Finally, we will examine the use of *advertising* by firms with market power. As we will see, deciding how much money to spend on advertising requires information about demand and is closely related to the firm's pricing decision. We will derive a simple rule of thumb for determining the profit-maximizing advertising-to-sales ratio.

11.1 CAPTURING CONSUMER SURPLUS

All the pricing strategies that we will examine have one thing in common: *They are means of capturing consumer surplus and transferring it to the producer*. You can see this more clearly in Figure 11.1. Suppose the firm sold all its output at a single price. To maximize profit, it would pick a price P^* and corresponding output Q^* at the intersection of its marginal cost and marginal revenue curves. Although the firm would then be profitable, its managers might still wonder if they could make it even more profitable.

Consumer surplus is explained in §4.4 and reviewed in §9.1.

They know that some customers (in region A of the demand curve) would pay more than P^*. But raising the price would mean losing some customers, selling less, and earning smaller profits. Similarly, other potential customers are not buying the firm's product because they will not pay a price as high as P^*. Many

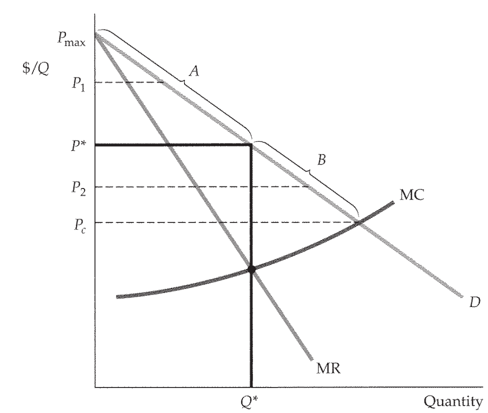

FIGURE 11.1 Capturing Consumer Surplus

If a firm can charge only one price for all its customers, that price will be P^* and the quantity produced will be Q^*. Ideally, the firm would like to charge a higher price to consumers willing to pay more than P^*, thereby capturing some of the consumer surplus under region A of the demand curve. The firm would also like to sell to consumers willing to pay prices lower than P^*, but only if doing so does not entail lowering the price to other consumers. In that way, the firm could also capture some of the surplus under region B of the demand curve.

of them, however, would pay prices higher than the firm's marginal cost. (These customers are in region B of the demand curve.) By lowering its price, the firm could sell to some of these customers. Unfortunately, it would then earn less revenue from its existing customers, and again profits would shrink.

How can the firm capture the consumer surplus (or at least part of it) from its customers in region A, and perhaps also sell profitably to some of its potential customers in region B? Charging a single price clearly will not do the trick. However, the firm might charge different prices to different customers, according to where the customers are along the demand curve. For example, some customers in the upper end of region A would be charged the higher price P_1, some in region B would be charged the lower price P_2, and some in between would be charged P^*. This is the basis of **price discrimination**: charging different prices to different customers. The problem, of course, is to identify the different customers, and to get them to pay different prices. We will see how this can be done in the next section.

• **price discrimination** Practice of charging different prices to different consumers for similar goods.

The other pricing techniques that we will discuss in this chapter—two-part tariffs and bundling—also expand the range of a firm's market to include more customers and to capture more consumer surplus. In each case, we will examine both the amount by which the firm's profit can be increased and the effect on consumer welfare. (As we will see, when there is a high degree of monopoly power, these pricing techniques can sometimes make both consumers and the producer better off.) We turn first to price discrimination.

11.2 PRICE DISCRIMINATION

Price discrimination can take three broad forms, which we call first-, second-, and third-degree price discrimination. We will examine them in turn.

First-Degree Price Discrimination

Ideally, a firm would like to charge a different price to each of its customers. If it could, it would charge each customer the maximum price that the customer is willing to pay for each unit bought. We call this maximum price the customer's **reservation price**. The practice of charging each customer his or her reservation price is called perfect **first-degree price discrimination**.[1] Let's see how it affects the firm's profit.

• **reservation price** Maximum price that a customer is willing to pay for a good.

• **first-degree price discrimination** Practice of charging each customer her reservation price.

First, we need to know the profit that the firm earns when it charges only the single price P^* in Figure 11.2. To find out, we can add the profit on each incremental unit produced and sold, up to the total quantity Q^*. This incremental profit is the marginal revenue less the marginal cost for each unit. In Figure 11.2, this marginal revenue is highest and marginal cost lowest for the first unit. For each additional unit, marginal revenue falls and marginal cost rises. Thus the firm produces the total output Q^*, at which point marginal revenue and marginal cost are equal.

In §8.3, we explain that a firm's profit-maximizing output is the output at which marginal revenue is equal to marginal cost.

If we add up the profits on each incremental unit produced, we obtain the firm's **variable profit**; the firm's profit, ignoring its fixed costs. In Figure 11.2, variable profit is given by the *yellow-shaded* area between the marginal revenue and

• **variable profit** Sum of profits on each incremental unit produced by a firm; i.e., profit ignoring fixed costs.

[1] We are assuming that each customer buys one unit of the good. If a customer buys more than one unit, the firm will have to charge different prices for each of the units.

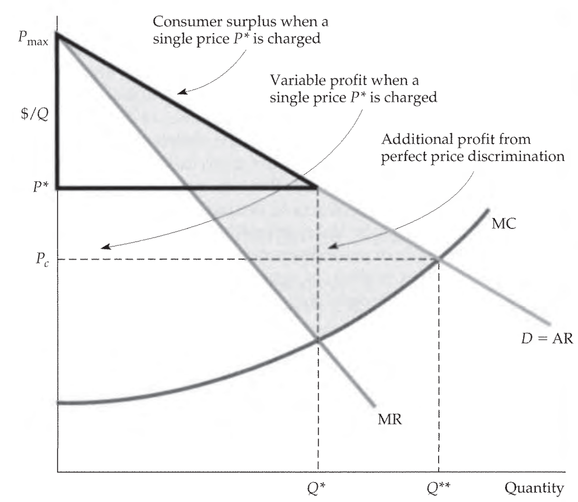

FIGURE 11.2 Additional Profit from Perfect First-Degree Price Discrimination

Because the firm charges each consumer her reservation price, it is profitable to expand output to Q^{**}. When only a single price, P^*, is charged, the firm's variable profit is the area between the marginal revenue and marginal cost curves. With perfect price discrimination, this profit expands to the area between the demand curve and the marginal cost curve.

marginal cost curves.[2] Consumer surplus, which is the area between the average revenue curve and the price P^* that customers pay, is outlined as a black triangle.

Perfect Price Discrimination What happens if the firm can perfectly price discriminate? Because each consumer is charged exactly what he or she is willing to pay, the marginal revenue curve is no longer relevant to the firm's output decision. Instead, the incremental revenue earned from each additional unit sold is simply the price paid for that unit; it is therefore given by the demand curve.

Since price discrimination does not affect the firm's cost structure, the cost of each additional unit is again given by the firm's marginal cost curve. Therefore, *the additional profit from producing and selling an incremental unit is now the difference between demand and marginal cost*. As long as demand exceeds marginal cost, the firm can increase its profit by expanding production. It will do so until it produces a total output Q^{**}. At Q^{**}, demand is equal to marginal cost, and producing any more reduces profit.

[2]Recall from Chapter 10 that because total profit π is the difference between total revenue R and total cost C, incremental profit is just $\Delta\pi = \Delta R - \Delta C = MR - MC$. Variable profit is found by summing all the $\Delta\pi$s, and thus it is the area between the MR and MC curves. This ignores fixed costs, which are independent of the firm's output and pricing decisions. Thus, total profit equals variable profit minus fixed cost.

Variable profit is now given by the area between the demand and marginal cost curves.[3] Observe from Figure 11.2 how the firm's profit has increased. (The additional profit resulting from price discrimination is shown by the purple-shaded area.) Note also that because every customer is being charged the maximum amount that he or she is willing to pay, all consumer surplus has been captured by the firm.

Imperfect Price Discrimination In practice, perfect first-degree price discrimination is almost never possible. First, it is usually impractical to charge each and every customer a different price (unless there are only a few customers). Second, a firm usually does not know the reservation price of each customer. Even if it could ask how much each customer would be willing to pay, it probably would not receive honest answers. After all, it is in the customers' interest to claim that they would pay very little.

Sometimes, however, firms can discriminate imperfectly by charging a few different prices based on estimates of customers' reservation prices. This practice is often used by professionals, such as doctors, lawyers, accountants, or architects, who know their clients reasonably well. In such cases, the client's willingness to pay can be assessed and fees set accordingly. For example, a doctor may offer a reduced fee to a low-income patient whose willingness to pay or insurance coverage is low but charge higher fees to upper-income or better-insured patients. And an accountant, having just completed a client's tax returns, is in an excellent position to estimate how much the client is willing to pay for the service.

Another example is a car salesperson, who typically works with a 15-percent profit margin. The salesperson can give part of this margin away to the customer by making a "deal," or can insist that the customer pay the full sticker price. A good salesperson knows how to size up customers: A customer who is likely to look elsewhere for a car is given a large discount (from the salesperson's point of view, a small profit is better than no sale and no profit), but the customer in a hurry is offered little or no discount. In other words, *a successful car salesperson knows how to price discriminate!*

Still another example is college and university tuition. Colleges don't charge different tuition rates to different students in the same degree programs. Instead, they offer financial aid, in the form of scholarships or subsidized loans, which reduces the *net* tuition that the student must pay. By requiring those who seek aid to disclose information about family income and wealth, colleges can link the amount of aid to ability (and hence willingness) to pay. Thus students who are financially well off pay more for their education, while students who are less well off pay less.

Figure 11.3 illustrates imperfect first-degree price discrimination. If only a single price were charged, it would be P_4^*. Instead, six different prices are charged, the lowest of which, P_6, is set at about the point where marginal cost intersects the demand curve. Note that those customers who would not have been willing to pay a price of P_4^* or greater are actually better off in this situation—they are now in the market and may be enjoying at least some consumer surplus. In fact, if price discrimination brings enough new customers into the market, consumer welfare can increase to the point that both the producer and consumers are better off.

[3]Incremental profit is again $\Delta\pi = \Delta R - \Delta C$, but ΔR is given by the price to each customer (i.e., the average revenue curve), so $\Delta\pi = AR - MC$. Variable profit is the sum of these $\Delta\pi$s and is given by the area between the AR and MC curves.

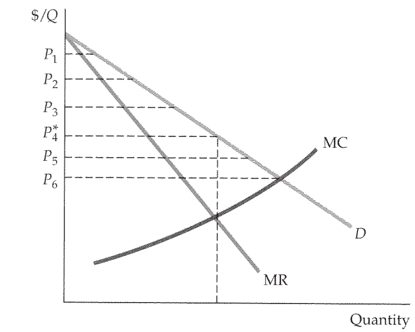

FIGURE 11.3 First-Degree Price Discrimination in Practice

Firms usually don't know the reservation price of every consumer, but sometimes reservation prices can be roughly identified. Here, six different prices are charged. The firm earns higher profits, but some consumers may also benefit. With a single price P_4^*, there are fewer consumers. The consumers who now pay P_5 or P_6 enjoy a surplus.

Second-Degree Price Discrimination

In some markets, as each consumer purchases many units of a good over any given period, his reservation price declines with the number of units purchased. Examples include water, heating fuel, and electricity. Consumers may each purchase a few hundred kilowatt-hours of electricity a month, but their willingness to pay declines with increasing consumption. The first 100 kilowatt-hours may be worth a lot to the consumer—operating a refrigerator and providing for minimal lighting. Conservation becomes easier with the additional units and may be worthwhile if the price is high. In this situation, a firm can discriminate according to the quantity consumed. This is called **second-degree price discrimination**, and it works by charging different prices for different quantities of the same good or service.

• **second-degree price discrimination** Practice of charging different prices per unit for different quantities of the same good or service.

Quantity discounts are an example of second-degree price discrimination. A single roll of Kodak film might be priced at $5, while a box containing four rolls of the same film might be priced at $14, making the average price per roll $3.50. Similarly, the price per ounce for breakfast cereal is likely to be smaller for the 24-ounce box than for the 16-ounce box.

Another example of second-degree price discrimination is *block pricing* by electric power companies, natural gas utilities, and municipal water companies. With **block pricing**, the consumer is charged different prices for different quantities or "blocks" of a good. If scale economies cause average and marginal costs to decline, the government agency that controls rates may encourage block pricing. Because it leads to expanded output and greater scale economies, this policy can increase consumer welfare while allowing for greater profit to the company: While prices are reduced overall, the savings from the lower unit cost still permits the company to increase its profit.

• **block pricing** Practice of charging different prices for different quantities or "blocks" of a good.

Figure 11.4 illustrates second-degree price discrimination for a firm with declining average and marginal costs. If a single price were charged, it would be P_0, and the quantity produced would be Q_0. Instead, three different prices are

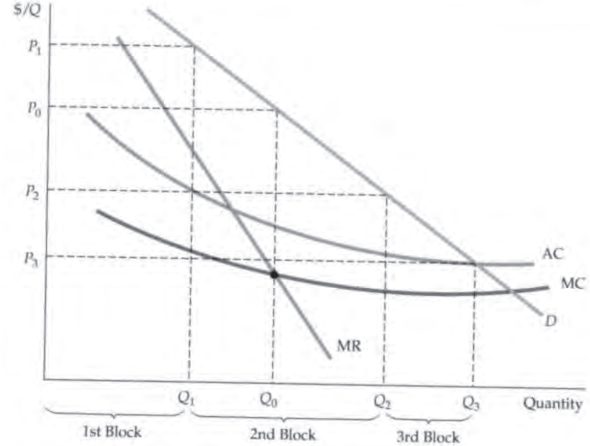

FIGURE 11.4 Second-Degree Price Discrimination

Different prices are charged for different quantities, or "blocks," of the same good. Here, there are three blocks, with corresponding prices P_1, P_2, and P_3. There are also economies of scale, and average and marginal costs are declining. Second-degree price discrimination can then make consumers better off by expanding output and lowering cost.

charged, based on the quantities purchased. The first block of sales is priced at P_1, the second at P_2, and the third at P_3.

Third-Degree Price Discrimination

A well-known liquor company has what seems to be a strange pricing practice. The company produces a vodka that it advertises as one of the smoothest and best-tasting available. This vodka is called "Three Star Golden Crown" and sells for about $16 a bottle.[4] However, the company also takes some of this same vodka and bottles it under the name "Old Sloshbucket," which is sold for about $8 a bottle. Why does it do this? Has the president of the company been spending too much time near the vats?

Perhaps, but this company is also practicing **third-degree price discrimination**, and it does so because the practice is profitable. This form of price discrimination divides consumers into two or more groups with separate demand curves for each group. It is the most prevalent form of price discrimination, and examples abound: regular versus "special" airline fares; premium versus nonpremium brands of liquor, canned food or frozen vegetables; discounts to students and senior citizens; and so on.

• **third-degree price discrimination** Practice of dividing consumers into two or more groups with separate demand curves and charging different prices to each group.

Creating Consumer Groups In each case, some characteristic is used to divide consumers into distinct groups. For many goods, for example, students and senior citizens are usually willing to pay less on average than the rest of the population

[4] We have changed the names to protect the innocent.

(because their incomes are lower), and identity can be readily established (via a college ID or driver's license). Likewise, to separate vacationers from business travelers (whose companies are usually willing to pay higher fares), airlines can put restrictions on special low-fare tickets, such as requiring advance purchase or a Saturday night stay. With the liquor company, or the premium versus nonpremium (e.g., supermarket label) brand of food, the label itself divides consumers; many consumers are willing to pay more for a name brand even though the nonpremium brand is identical or nearly identical (and might be manufactured by the same company that produced the premium brand).

If third-degree price discrimination is feasible, how should the firm decide what price to charge each group of consumers? Let's think about this in two steps.

1. We know that however much is produced, total output should be divided between the groups of customers so that marginal revenues for each group are equal. Otherwise, the firm would not be maximizing profit. For example, if there are two groups of customers and the marginal revenue for the first group, MR_1, exceeds the marginal revenue for the second group, MR_2, the firm could clearly do better by shifting output from the second group to the first. It would do this by lowering the price to the first group and raising the price to the second group. Thus, whatever the two prices, they must be such that the marginal revenues for the different groups are equal.

2. We know that *total* output must be such that the marginal revenue for each group of consumers is equal to the marginal cost of production. Again, if this were not the case, the firm could increase its profit by raising or lowering total output (and lowering or raising its prices to both groups). For example, suppose that marginal revenues were the same for each group of consumers but that marginal revenue exceeded marginal cost. The firm could then make a greater profit by increasing its total output. It would lower its prices to both groups of consumers, so that marginal revenues for each group would fall (but would still be equal to each other) and would approach marginal cost.

Let's look at this problem algebraically. Let P_1 be the price charged to the first group of consumers, P_2 the price charged to the second group, and $C(Q_T)$ the total cost of producing output $Q_T = Q_1 + Q_2$. Total profit is then

$$\pi = P_1Q_1 + P_2Q_2 - C(Q_T)$$

The firm should increase its sales to each group of consumers, Q_1 and Q_2, until the incremental profit from the last unit sold is zero. First, we set incremental profit for sales to the first group of consumers equal to zero:

$$\frac{\Delta\pi}{\Delta Q_1} = \frac{\Delta(P_1Q_1)}{\Delta Q_1} - \frac{\Delta C}{\Delta Q_1} = 0$$

Here, $\Delta(P_1Q_1)/\Delta Q_1$ is the incremental revenue from an extra unit of sales to the first group of consumers (i.e., MR_1). The next term, $\Delta C/\Delta Q_1$, is the incremental cost of producing this extra unit—i.e., marginal cost, MC. We thus have

$$MR_1 = MC$$

Similarly, for the second group of consumers, we must have

$$MR_2 = MC$$

Putting these relations together, we see that prices and output must be set so that

$$MR_1 = MR_2 = MC \qquad \textbf{(11.1)}$$

Again, marginal revenue must be equal across groups of consumers and must equal marginal cost.

Determining Relative Prices Managers may find it easier to think in terms of the relative prices that should be charged to each group of consumers and to relate these prices to the elasticities of demand. Recall from Section 10.1 that we can write marginal revenue in terms of the elasticity of demand:

In our discussion of a rule of thumb for pricing in §10.1, we explained that a profit-maximizing firm chooses an output at which its marginal revenue is equal to the price of the product plus the ratio of the price to the price elasticity of demand.

$$MR = P(1 + 1/E_d)$$

Thus $MR_1 = P_1(1 + 1/E_1)$ and $MR_2 = P_2(1 + 1/E_2)$, where E_1 and E_2 are the elasticities of demand for the firm's sales in the first and second markets, respectively. Now equating MR_1 and MR_2 as in equation (11.1) gives the following relationship that must hold for the prices:

$$\frac{P_1}{P_2} = \frac{(1+1/E_2)}{(1+1/E_1)} \qquad \textbf{(11.2)}$$

As you would expect, the higher price will be charged to consumers with the lower demand elasticity. For example, if the elasticity of demand for consumers in group 1 is −2 and the elasticity for consumers in group 2 is −4, we will have $P_1/P_2 = (1 - 1/4)/(1 - 1/2) = (3/4)/(1/2) = 1.5$. In other words, the price charged to the first group of consumers should be 1.5 times as high as the price charged to the second group.

Figure 11.5 illustrates third-degree price discrimination. Note that the demand curve D_1 for the first group of consumers is less elastic than the curve for the second

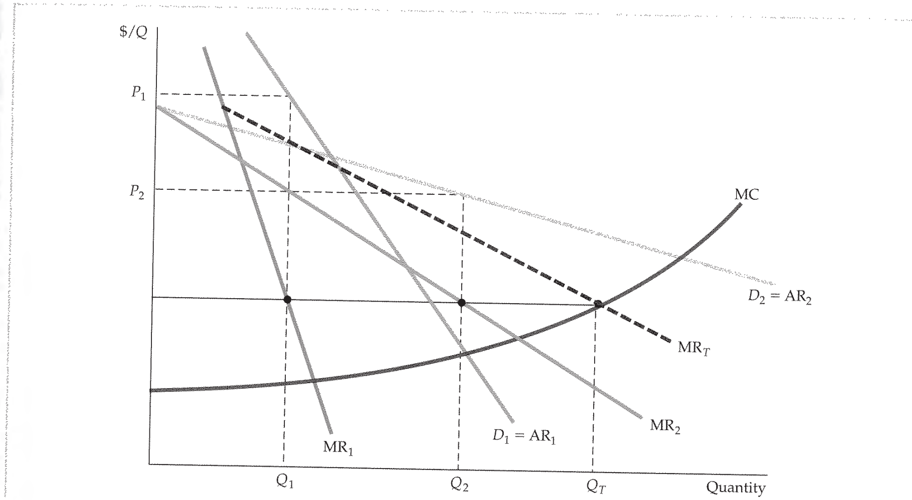

FIGURE 11.5 Third-Degree Price Discrimination

Consumers are divided into two groups, with separate demand curves for each group. The optimal prices and quantities are such that the marginal revenue from each group is the same and equal to marginal cost. Here group 1, with demand curve D_1, is charged P_1, and group 2, with the more elastic demand curve D_2, is charged the lower price P_2. Marginal cost depends on the total quantity produced Q_T. Note that Q_1 and Q_2 are chosen so that $MR_1 = MR_2 = MC$.

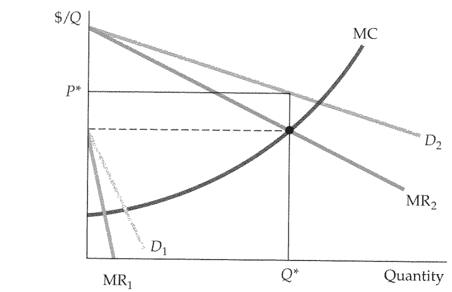

FIGURE 11.6 No Sales to Smaller Market

Even if third-degree price discrimination is feasible, it may not pay to sell to both groups of consumers if marginal cost is rising. Here the first group of consumers, with demand D_1, are not willing to pay much for the product. It is unprofitable to sell to them because the price would have to be too low to compensate for the resulting increase in marginal cost.

group; thus the price charged to the first group is higher. The total quantity produced, $Q_T = Q_1 + Q_2$, is found by summing the marginal revenue curves MR_1 and MR_2 horizontally, which yields the dashed curve MR_T, and finding its intersection with the marginal cost curve. Because MC must equal MR_1 and MR_2, we can draw a horizontal line leftward from this intersection to find the quantities Q_1 and Q_2.

It may not always be worthwhile for the firm to try to sell to more than one group of consumers. In particular, if demand is small for the second group and marginal cost is rising steeply, the increased cost of producing and selling to this group may outweigh the increase in revenue. In Figure 11.6, the firm is better off charging a single price P^* and selling only to the larger group of consumers: The additional cost of serving the smaller market would outweigh the additional revenue that might come from selling to it.

EXAMPLE 11.1 The Economics of Coupons and Rebates

Producers of processed foods and related consumer goods often issue coupons that let customers buy products at discounts. These coupons are usually distributed as part of an advertisement for the product. They may appear in newspapers or magazines or in promotional mailings. For example, a coupon for a particular breakfast cereal might be worth 50 cents toward the purchase of a box of the cereal. Why do firms issue these coupons? Why not just lower the price of the product and thereby save the costs of printing and collecting the coupons?

Coupons provide a means of price discrimination. Studies show that only about 20 to 30 percent of all consumers regularly bother to clip, save, and use coupons. These consumers tend to be more sensitive to price than those who ignore coupons. They generally have more price-elastic demands and lower reservation prices. By issuing coupons, therefore, a cereal company can separate its customers into two groups and, in effect, charge the more price-sensitive customers a lower price than the other customers.

Rebate programs work the same way. For example, Kodak ran a program in which a consumer could mail in a form together with the proof of purchase of three rolls of film and receive a rebate of $1.50. Why not just lower the price of film by 50 cents a roll? Because only those consumers with relatively price-sensitive demands bother to send in the materials and request rebates. Again, the program is a means of price discrimination.

Can consumers really be divided into distinct groups in this way? Table 11.1 shows the results of a statistical study in which, for a variety of products, price elasticities of demand were estimated for users and nonusers of coupons.[5] This study confirms that users of coupons tend to have more price-sensitive demands. It also shows the extent to which the elasticities differ for the two groups of consumers and how the difference varies from one product to another.

By themselves, these elasticity estimates do not tell a firm what price to set and how large a discount to offer because they pertain to *market demand*, not to the demand for the firm's particular brand. For example, Table 11.1 indicates that the elasticity of demand for cake mix is –0.21 for nonusers of coupons and –0.43

TABLE 11.1 Price Elasticities of Demand for Users versus Nonusers of Coupons

	PRICE ELASTICITY	
Product	**Nonusers**	**Users**
Toilet tissue	–0.60	–0.66
Stuffing/dressing	–0.71	–0.96
Shampoo	–0.84	–1.04
Cooking/salad oil	–1.22	–1.32
Dry mix dinners	–0.88	–1.09
Cake mix	–0.21	–0.43
Cat food	–0.49	–1.13
Frozen entrees	–0.60	–0.95
Gelatin	–0.97	–1.25
Spaghetti sauce	–1.65	–1.81
Creme rinse/conditioner	–0.82	–1.12
Soups	–1.05	–1.22
Hot dogs	–0.59	–0.77

[5]The study is by Chakravarthi Narasimhan, "A Price Discrimination Theory of Coupons," *Marketing Science* (Spring 1984). A recent study of coupons for breakfast cereals finds that contrary to the predictions of the price-discrimination model, shelf prices for cereals tend to be lower during periods when coupons are more widely available. This might occur because couponing spurs more price competition among cereal manufacturers. See Aviv Nevo and Catherine Wolfram, "Prices and Coupons for Breakfast Cereals," *RAND Journal of Economics* 33 (2002): 319–39.

for users. But the elasticity of demand for any of the five or six major brands of cake mix on the market will be much larger than either of these numbers—about five or six times as large, as a rule of thumb.[6] So for any one brand of cake mix—say, Pillsbury—the elasticity of demand for users of coupons might be about −2.4, versus about −1.2 for nonusers. From equation (11.2), therefore, we can determine that the price to nonusers of coupons should be about 1.5 times the price to users. In other words, if a box of cake mix sells for $3.00, the company should offer coupons that give a $1.00 discount.

EXAMPLE 11.2 Airline Fares

Travelers are often amazed at the variety of fares available for round-trip flights from New York to Los Angeles. Recently, for example, the first-class fare was above $2000; the regular (unrestricted) economy fare was about $1700, and special discount fares (often requiring the purchase of a ticket two weeks in advance and/or a Saturday night stayover) could be bought for as little as $400. Although first-class service is not the same as economy service with a minimum stay requirement, the difference would not seem to warrant a price that is seven times as high. Why do airlines set such fares?

These fares provide a profitable form of price discrimination. The gains from discriminating are large because different types of customers, with very different elasticities of demand, purchase these different types of tickets. Table 11.2 shows price (and income) elasticities of demand for three categories of service within the United States: first class, unrestricted coach, and discounted tickets (which often have restrictions and may be partly nonrefundable).

Note that the demand for discounted fares is about two or three times as price elastic as first-class or unrestricted coach service. Why the difference? While discounted tickets are usually used by families and other leisure travelers, first-class and unrestricted coach tickets are more often bought by business travelers, who have little choice about when they travel and whose companies pick up the tab. Of course, these elasticities pertain to market demand, and with several airlines competing for customers, the elasticities of demand for each airline will be larger. But the *relative* sizes of elasticities across the three categories of service should be about the same. When elasticities of demand differ so widely, it should not be surprising that airlines set such different fares for different categories of service.

TABLE 11.2 Elasticities of Demand for Air Travel

	FARE CATEGORY		
Elasticity	**First Class**	**Unrestricted Coach**	**Discounted**
Price	−0.3	−0.4	−0.9
Income	1.2	1.2	1.8

[6]This rule of thumb applies if interfirm competition can be described by the Cournot model, which we will discuss in Chapter 12.

Airline price discrimination has become increasingly sophisticated. A wide variety of fares is available, depending on how far in advance the ticket is bought, the percentage of the fare that is refundable if the trip is changed or cancelled, and whether the trip includes a weekend stay.[7] The objective of the airlines has been to discriminate more finely among travelers with different reservation prices. As one industry executive puts it, "You don't want to sell a seat to a guy for $69 when he is willing to pay $400."[8] At the same time, an airline would rather sell a seat for $69 than leave it empty.

11.3 INTERTEMPORAL PRICE DISCRIMINATION AND PEAK-LOAD PRICING

Two other closely related forms of price discrimination are important and widely practiced. The first of these is **intertemporal price discrimination**: separating consumers with different demand functions into different groups by charging different prices at different points in time. The second is **peak-load pricing**: charging higher prices during peak periods when capacity constraints cause marginal costs to be high. Both of these strategies involve charging different prices at different times, but the reasons for doing so are somewhat different in each case. We will take each in turn.

• **intertemporal price discrimination** Practice of separating consumers with different demand functions into different groups by charging different prices at different points in time.

• **peak-load pricing** Practice of charging higher prices during peak periods when capacity constraints cause marginal costs to be high.

Intertemporal Price Discrimination

The objective of intertemporal price discrimination is to divide consumers into high-demand and low-demand groups by charging a price that is high at first but falls later. To see how this strategy works, think about how an electronics company might price new, technologically advanced equipment, such as high-performance digital cameras or LCD television monitors. In Figure 11.7, D_1 is the (inelastic) demand curve for a small group of consumers who value the product highly and do not want to wait to buy it (e.g., photography buffs who want the latest camera). D_2 is the demand curve for the broader group of consumers who are more willing to forgo the product if the price is too high. The strategy, then, is to offer the product initially at the high price P_1, selling mostly to consumers on demand curve D_1. Later, after this first group of consumers has bought the product, the price is lowered to P_2, and sales are made to the larger group of consumers on demand curve D_2.[9]

There are other examples of intertemporal price discrimination. One involves charging a high price for a first-run movie and then lowering the price after the movie has been out a year. Another, practiced almost universally by publishers, is to charge a high price for the hardcover edition of a book and then to release the paperback version at a much lower price about a year later. Many people

[7]Airlines also allocate the number of seats on each flight that will be available for each fare category. The allocation is based on the total demand and mix of passengers expected for each flight, and can change as the departure of the flight nears and estimates of demand and passenger mix change.

[8]"The Art of Devising Air Fares," *New York Times*, March 4, 1987.

[9]The prices of new electronic products also come down over time because costs fall as producers start to achieve greater scale economies and move down the learning curve. But even if costs did not fall, producers can make more money by first setting high prices and then reducing them over time, thereby discriminating and capturing consumer surplus.

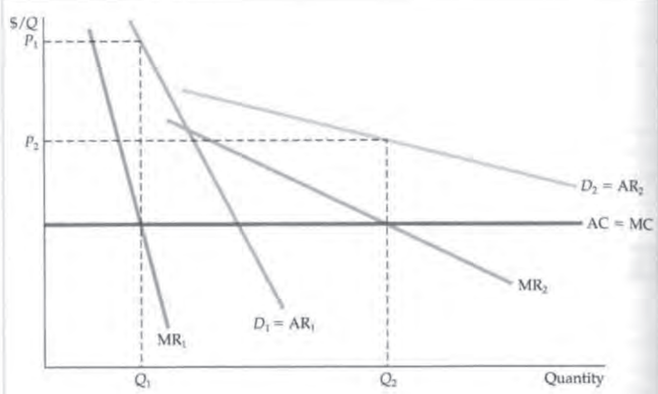

FIGURE 11.7 Intertemporal Price Discrimination

Consumers are divided into groups by changing the price over time. Initially, the price is high. The firm captures surplus from consumers who have a high demand for the good and who are unwilling to wait to buy it. Later the price is reduced to appeal to the mass market.

think that the lower price of the paperback is due to a much lower cost of production, but this is not true. Once a book has been edited and typeset, the marginal cost of printing an additional copy, whether hardcover or paperback, is quite low, perhaps a dollar or so. The paperback version is sold for much less not because it is much cheaper to print but because high-demand consumers have already purchased the hardbound edition. The remaining consumers—paperback buyers—generally have more elastic demands.

Peak-Load Pricing

Peak-load pricing also involves charging different prices at different points in time. Rather than capturing consumer surplus, however, the objective is to increase economic efficiency by charging consumers prices that are close to marginal cost.

For some goods and services, demand peaks at particular times—for roads and tunnels during commuter rush hours, for electricity during late summer afternoons, and for ski resorts and amusement parks on weekends. Marginal cost is also high during these peak periods because of capacity constraints. Prices should thus be higher during peak periods.

This is illustrated in Figure 11.8, where D_1 is the demand curve for the peak period and D_2 the demand curve for the nonpeak period. The firm sets marginal revenue equal to marginal cost for each period, obtaining the high price P_1 for the peak period and the lower price P_2 for the nonpeak period, selling corresponding quantities Q_1 and Q_2. This strategy increases the firm's profit above what it would be if it charged one price for all periods. It is also more efficient: The sum of producer and consumer surplus is greater because prices are closer to marginal cost.

In §9.2, we explain that economic efficiency means that aggregate consumer and producer surplus is maximized.

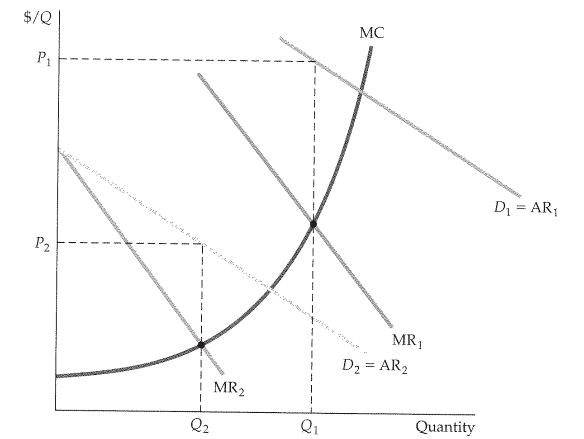

FIGURE 11.8 Peak-Load Pricing

Demands for some goods and services increase sharply during particular times of the day or year. Charging a higher price P_1 during the peak periods is more profitable for the firm than charging a single price at all times. It is also more efficient because marginal cost is higher during peak periods.

The efficiency gain from peak-load pricing is important. If the firm were a regulated monopolist (e.g., an electric utility), the regulatory agency should set the prices P_1 and P_2 at the points where the *demand* curves, D_1 and D_2, intersect the marginal cost curve, rather than where the marginal revenue curves intersect marginal cost. In that case, consumers realize the entire efficiency gain.

Note that peak-load pricing is different from third-degree price discrimination. With third-degree price discrimination, marginal revenue must be equal for each group of consumers and equal to marginal cost. Why? Because the costs of serving the different groups are not independent. For example, with unrestricted versus discounted air fares, increasing the number of seats sold at discounted fares affects the cost of selling unrestricted tickets—marginal cost rises rapidly as the airplane fills up. But this is not so with peak-load pricing (or for that matter, with most instances of intertemporal price discrimination). Selling more tickets for ski lifts or amusement parks on a weekday does not significantly raise the cost of selling tickets on the weekend. Similarly, selling more electricity during off-peak periods will not significantly increase the cost of selling electricity during peak periods. As a result, price and sales in each period can be determined independently by setting marginal cost equal to marginal revenue for each period.

Movie theaters, which charge more for evening shows than for matinees, are another example. For most movie theaters, the marginal cost of serving customers during the matinee is independent of marginal cost during the evening. The owner of a movie theater can determine the optimal prices for the evening and matinee shows independently, using estimates of demand and marginal cost in each period.

EXAMPLE 11.3 How to Price a Best-Selling Novel

Publishing both hardbound and paperback editions of a book allows publishers to price discriminate. As they do with most goods, consumers differ considerably in their willingness to pay for books. For example, some consumers want to buy a new bestseller as soon as it is released, even if the price is $25. Other consumers, however, will wait a year until the book is available in paperback for $10. But how does a publisher decide that $25 is the right price for the new hardbound edition and $10 is the right price for the paperback edition? And how long should it wait before bringing out the paperback edition?

The key is to divide consumers into two groups, so that those who are willing to pay a high price do so and *only* those unwilling to pay a high price wait and buy the paperback. This means that significant time must be allowed to pass before the paperback is released. If consumers know that the paperback will be available within a few months, they will have little incentive to buy the hardbound edition.[10] On the other hand, if the publisher waits too long to bring out the paperback edition, interest will wane and the market will dry up. As a result, publishers typically wait 12 to 18 months before releasing paperback editions.

What about price? Setting the price of the hardbound edition is difficult: Except for a few authors whose books always seem to sell, publishers have little data with which to estimate demand for a book that is about to be published. Often, they can judge only from the past sales of similar books. But usually only aggregate data are available for each category of book. Most new novels, therefore, are released at similar prices. It is clear, however, that those consumers willing to wait for the paperback edition have demands that are far more elastic than those of bibliophiles. It is not surprising, then, that paperback editions sell for so much less than hardbacks.[11]

11.4 THE TWO-PART TARIFF

• **two-part tariff** Form of pricing in which consumers are charged both an entry and a usage fee.

The **two-part tariff** is related to price discrimination and provides another means of extracting consumer surplus. It requires consumers to pay a fee up front for the right to buy a product. Consumers then pay an additional fee for each unit of the product they wish to consume. The classic example of this strategy is an amusement park.[12] You pay an admission fee to enter, and you also pay a certain amount for each ride. The owner of the park must decide whether to charge a high entrance fee and a low price for the rides or, alternatively, to admit people for free but charge high prices for the rides.

The two-part tariff has been applied in many settings: tennis and golf clubs (you pay an annual membership fee plus a fee for each use of a court or round of golf); the rental of large mainframe computers (a flat monthly fee plus a fee for each unit of processing time consumed); telephone service (a monthly hook-up fee plus a fee for minutes of usage). The strategy also applies to the sale of products like safety razors (you pay for the razor, which lets you consume the blades that fit that brand of razor).

The problem for the firm is how to set the *entry fee* (which we denote by T) versus the *usage fee* (which we denote by P). Assuming that the firm has some market power, should it set a high entry fee and low usage fee, or vice versa? To solve this problem, we need to understand the basic principles involved.

Single Consumer Let's begin with the artificial but simple case illustrated in Figure 11.9. Suppose there is only one consumer in the market (or many consumers with identical demand curves). Suppose also that the firm knows this consumer's demand curve. Now, remember that the firm wants to capture as much consumer surplus as possible. In this case, the solution is straightforward: Set the usage fee P equal to marginal cost and the entry fee T equal to the total consumer surplus for each consumer. Thus, the consumer pays T^* (or a bit less) to use the product, and $P^* = \text{MC}$ per unit consumed. With the fees set in this way, the firm captures *all* the consumer surplus as its profit.

Two Consumers Now suppose that there are two different consumers (or two groups of identical consumers). The firm, however, can set only *one* entry fee and one usage fee. It would thus no longer want to set the usage fee equal to marginal cost. If it did, it could make the entry fee no larger than the consumer surplus of the consumer with the smaller demand (or else it would lose that consumer), and this would not yield a maximum profit. Instead, the firm should set the usage fee *above* marginal cost and then set the entry fee equal to the remaining consumer surplus of the consumer with the smaller demand.

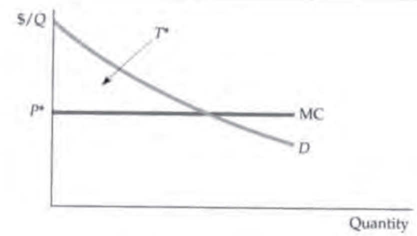

FIGURE 11.9 Two-Part Tariff with a Single Consumer

The consumer has demand curve D. The firm maximizes profit by setting usage fee P equal to marginal cost and entry fee T^* equal to the entire surplus of the consumer.

[10]Some consumers will buy the hardbound edition even if the paperback is already available because it is more durable and more attractive on a bookshelf. This must be taken into account when setting prices, but it is of secondary importance compared with intertemporal price discrimination.

[11]Hardbound and paperback editions are often published by different companies. The author's agent auctions the rights to the two editions, but the contract for the paperback specifies a delay to protect the sales of the hardbound edition. The principle still applies, however. The length of the delay and the prices of the two editions are chosen to price discriminate intertemporally.

[12]This pricing strategy was first analyzed by Walter Oi, "A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly," *Quarterly Journal of Economics* (February 1971): 77–96.

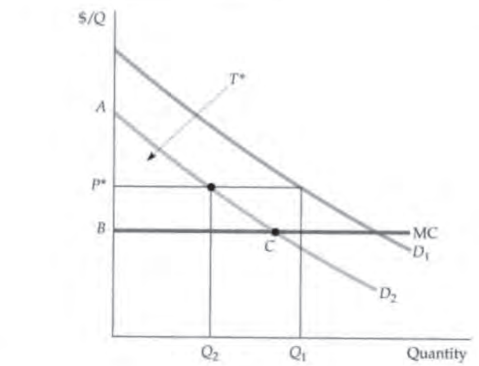

FIGURE 11.10 Two-Part Tariff with Two Consumers

The profit-maximizing usage fee P^* will exceed marginal cost. The entry fee T^* is equal to the surplus of the consumer with the smaller demand. The resulting profit is $2T^* + (P^* - \text{MC})(Q_1 + Q_2)$. Note that this profit is larger than twice the area of triangle *ABC*.

Figure 11.10 illustrates this. With the optimal usage fee at P^* greater than MC, the firm's profit is $2T^* + (P^* - \text{MC})(Q_1 + Q_2)$. (There are two consumers, and each pays T^*.) You can verify that this profit is more than twice the area of triangle *ABC*, the consumer surplus of the consumer with the smaller demand when P = MC. To determine the exact values of P^* and T^*, the firm would need to know (in addition to its marginal cost) the demand curves D_1 and D_2. It would then write down its profit as a function of P and T and choose the two prices that maximize this function. (See Exercise 10 for an example of how to do this.)

Many Consumers Most firms, however, face a variety of consumers with different demands. Unfortunately, there is no simple formula to calculate the optimal two-part tariff in this case, and some trial-and-error experiments might be required. But there is always a trade-off: A lower entry fee means more entrants and thus more profit from sales of the item. On the other hand, as the entry fee becomes smaller and the number of entrants larger, the profit derived from the entry fee will fall. The problem, then, is to pick an entry fee that results in the optimum number of entrants—that is, the fee that allows for maximum profit. In principle, we can do this by starting with a price for sales of the item P, finding the optimum entry fee T, and then estimating the resulting profit. The price P is then changed, and the corresponding entry fee calculated, along with the new profit level. By iterating this way, we can approach the optimal two-part tariff.

Figure 11.11 illustrates this principle. The firm's profit π is divided into two components, each of which is plotted as a function of the entry fee T, assuming a fixed sales price P. The first component, π_a, is the profit from the entry fee and is equal to the revenue $n(T)T$, where $n(T)$ is the number of entrants. (Note that a high T implies a small n.) Initially, as T is increased from zero, revenue $n(T)T$ rises. Eventually, however, further increases in T will make n so small that $n(T)T$ falls. The second component, π_s, is the profit from sales of the item itself at price P and is equal to $(P - \text{MC})Q$, where Q is the rate at which entrants purchase the item. The larger the number of entrants n, the larger Q will be. Thus π_s falls when T is increased because a higher T reduces n.

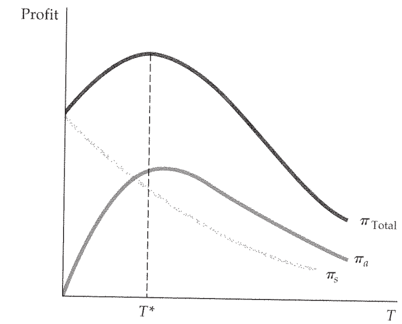

FIGURE 11.11 Two-Part Tariff with Many Different Consumers

Total profit π is the sum of the profit from the entry fee π_a and the profit from sales π_s. Both π_a and π_s depend on T, the entry fee. Therefore

$$\pi = \pi_a + \pi_s = n(T)T + (P - \text{MC})Q(n)$$

where n is the number of entrants, which depends on the entry fee T, and Q is the rate of sales, which is greater the larger is n. Here T^* is the profit-maximizing entry fee, given P. To calculate optimum values for P and T, we can start with a number for P, find the optimum T, and then estimate the resulting profit. P is then changed and the corresponding T recalculated, along with the new profit level.

Starting with a number for P, we determine the optimal (profit-maximizing) T^*. We then change P, find a new T^*, and determine whether profit is now higher or lower. This procedure is repeated until profit has been maximized.

Obviously, more data are needed to design an optimal two-part tariff than to choose a single price. Knowing marginal cost and the aggregate demand curve is not enough. It is impossible (in most cases) to determine the demand curve of every consumer, but one would at least like to know by how much individual demands differ from one another. If consumers' demands for your product are fairly similar, you would want to charge a price P that is close to marginal cost and make the entry fee T large. This is the ideal situation from the firm's point of view because most of the consumer surplus could then be captured. On the other hand, if consumers have different demands for your product, you would probably want to set P well above marginal cost and charge a lower entry fee T. In that case, however, the two-part tariff is a less effective means of capturing consumer surplus; setting a single price may do almost as well.

At Disneyland in California and Walt Disney World in Florida, the strategy is to charge a high entry fee and charge nothing for the rides. This policy makes sense because consumers have reasonably similar demands for Disney vacations. Most people visiting the parks plan daily budgets (including expenditures for food and beverages) that, for most consumers, do not differ very much.

Firms are perpetually searching for innovative pricing strategies, and a few have devised and introduced a two-part tariff with a "twist"—the entry fee T entitles the customer to a certain number of free units. For example, if you buy a Gillette razor, several blades are usually included in the package. The monthly lease fee for a mainframe computer usually includes some free usage before usage is charged. This twist lets the firm set a higher entry fee T without losing as many small customers. Because these small customers might pay little or nothing for usage under this scheme, the higher entry fee will capture their surplus without driving them out of the market, while also capturing more of the surplus of the large customers.

EXAMPLE 11.4 Polaroid Cameras

In 1971, Polaroid introduced its SX-70 camera. This camera was sold, not leased, to consumers. Nevertheless, because film was sold separately, Polaroid could apply a two-part tariff to the pricing of the SX-70. Let's see how this pricing strategy gave Polaroid greater profits than would have been possible if its camera had used ordinary roll film, and how Polaroid might have determined the optimal prices for each part of its two-part tariff.

Why did the pricing of Polaroid's cameras and film involve a two-part tariff? Because Polaroid had a monopoly on both its camera and the film, only Polaroid film could be used in the camera. Consumers bought the camera and film to take instant pictures: The camera was the "entry fee" that provided access to the consumption of instant pictures, which was what consumers ultimately demanded.[13] In this sense, the price of the camera was like the entry fee at an amusement park. However, while the marginal cost of allowing someone entry into the park is close to zero, the marginal cost of producing a camera is significantly above zero, and thus had to be taken into account when designing the two-part tariff.

It was important that Polaroid have a monopoly on the film as well as the camera. If the camera had used ordinary roll film, competitive forces would have pushed the price of film close to its marginal cost. If all consumers had identical demands, Polaroid could still have captured all the consumer surplus by setting a high price for the camera (equal to the surplus of each consumer). But in practice, consumers were heterogeneous, and the optimal two-part tariff required a price for the film well above marginal cost.

How should Polaroid have selected its prices for the camera and film? It could have begun with some analytical spadework. Its profit is given by

$$\pi = PQ + nT - C_1(Q) - C_2(n)$$

where P is the price of the film, T the price of the camera, Q the quantity of film sold, n the number of cameras sold, and $C_1(Q)$ and $C_2(n)$ the costs of producing film and cameras, respectively.

[13]We are simplifying here. In fact, some consumers obtain utility just from owning the camera, even if they take few or no pictures. Adults, like children, enjoy new toys and can obtain pleasure from the mere possession of a technologically innovative good.

Polaroid wanted to maximize its profit π, taking into account that Q and n depend on P and T. Given a heterogeneous base of potential consumers, managers might initially have guessed at this dependence on P and T, drawing on knowledge of related products. Later, they may have gotten a better understanding of demand and of how Q and n depend on P and T as they accumulated data from the firm's sales experience. They may have found knowledge of C_1 and C_2 easier to come by, perhaps from engineering and statistical studies (as discussed in Chapter 7).

Given some initial guesses or estimates for $Q(P)$, $n(T)$, $C_1(Q)$, and $C_2(n)$, Polaroid could have calculated the profit-maximizing prices P and T. It could also have determined how sensitive these prices were to uncertainty over demand and cost. This knowledge could have provided a guideline for trial-and-error pricing experiments. Over time these experiments would also have told Polaroid more about demand and cost, so that it could refine its two-part tariff accordingly.[14]

In 1999, Polaroid introduced its I-Zone camera and film, which takes matchbook-size pictures. The camera was priced at $25 and the film at $7 per pack. In 2003, Polaroid's One Step cameras sold for $30 to $50 and used Polaroid 600 film, which was priced at about $14 per pack of 10 pictures. Polaroid's higher-end Spectra cameras sold for $60 to over $100 and used Spectra film, priced at about $13 per pack. These film prices were well above marginal cost, reflecting the considerable heterogeneity of consumer demands.

EXAMPLE 11.5 Pricing Cellular Phone Service

Most telephone service is priced using a two-part tariff: a monthly access fee, which may include some free minutes, plus a per-minute charge for additional minutes. This is also true for cellular phone service, which has grown explosively, both in the United States and around the world. In the case of cellular service, providers have taken the two-part tariff and turned it into an art form.

In most parts of the United States, consumers can choose among four national network providers—Verizon, T-Mobile, AT&T, and Sprint. These providers compete among themselves for customers, but each has some market power. This market power arises in part from oligopolistic pricing and output decisions, as we will explain in Chapters 12 and 13. Market power also arises because consumers face *switching costs*: When they sign up for a cellular plan, they must typically make a commitment to stay for at least one year, and breaking the contract is quite expensive. Most service providers impose a penalty upwards of $200 for early termination.

Because providers have market power, they must think carefully about profit-maximizing pricing strategies. The two-part tariff provides an ideal means by which cellular providers can capture consumer surplus and turn it into profit.

Table 11.3 shows cellular rate plans (for 2007) offered by Verizon Wireless, T-Mobile, and AT&T. The plans are structured in similar ways, so let's focus on

[14]Setting prices for a product such as a Polaroid camera is clearly not a simple matter. We have ignored the *dynamic* behavior of cost and demand: namely, how production costs fall as the firm moves down its learning curve and how demand changes over time as the market becomes saturated.

TABLE 11.3 Cellular Rate Plans (2007)

Anytime Minutes	Monthly Access Fee	Unlimited Nights/ Weekends	Per-Minute Rate After Allowance
A. Verizon: America's Choice Basic			
450	$39.99	Included	$0.45
900	$59.99	Included	$0.40
1350	$79.99	Included	$0.35
2000	$99.99	Included	$0.25
4000	$149.99	Included	$0.25
6000	$199.99	Included	$0.20
B. T-Mobile Individual Plans			
300	$29.99	Unlimited weekends, not weeknights	$0.40
1000	$39.99	Included	$0.40
1500	$59.99	Included	$0.40
2500	$99.99	Included	$0.30
5000	$129.99	Included	$0.30
C. AT&T Individual Plans			
450	$39.99	Includes 5000 minutes	$0.45
900	$59.99	Included	$0.40
1350	$79.99	Included	$0.35
2000	$99.99	Included	$0.25
4000	$149.99	Included	$0.25
6000	$199.99	Included	$0.20

Note: T-Mobile plans do not include any mobile-to-mobile minutes; for T-Mobile these calls are charged from the Anytime Minutes. All other plans include unlimited mobile to mobile minutes.

the Verizon plan. The least expensive Verizon plan has a monthly access charge of $39.99 and includes 450 "anytime" minutes (i.e., 450 minutes of talk time per month that can be used at any hour of the day). The plan also includes an unlimited amount of talk time during nights and weekends (periods when demand is generally much lower). A subscriber who uses more than the 450 "anytime" minutes is charged $0.45 for each additional minute. A customer who uses her cell phone more frequently could sign up for a more expensive plan, e.g., one that costs $59.99 per month but includes 900 "anytime" minutes and a charge of $0.40 for additional minutes. And if you, the reader, use your cell phone constantly (and thus have time for little else), you could sign up for a plan that includes 6000 "anytime" minutes, at a monthly cost of $199.99.

Why do cellular phone providers offer several different types of plans and options within each? Why don't they simply offer a single two-part tariff with a monthly access charge and a per-minute usage charge? Offering several different plans and options allows companies to combine third-degree price discrimination

with the two-part tariff. The plans are structured so that consumers sort themselves into groups based on their plan choices. A different two-part tariff is then applied to each group.

To see how this sorting works, consider the plan choices of different types of consumers. People who use a cell phone only occasionally will want to spend as little as possible on the service and will choose the least expensive plan (with the fewest "anytime" minutes). The most expensive plans are best suited to very heavy users (perhaps a salesperson who travels extensively and makes call throughout the day), who will want to minimize their per-minute cost. Other plans are better suited to consumers with moderate calling needs.

Consumers will choose a plan that best matches their needs. Thus they will sort themselves into groups, and the consumers in each group will be relatively homogeneous in terms of demands for cellular service. Remember that the two-part tariff works best when consumers have identical or very similar demands. (Recall from Figure 11.9 that with identical consumers, the two-part tariff can be used to capture *all* consumer surplus.) Creating a situation in which consumers sort themselves into groups in this way makes best use of the two-part tariff.

*11.5 BUNDLING

You have probably seen the 1939 film *Gone with the Wind*. It is a classic that is nearly as popular now as it was then.[15] Yet we would guess that you have not seen *Getting Gertie's Garter*, a flop that the same company (MGM, a division of Loews) also distributed. And we would also guess that you did not know that these two films were priced in what was then an unusual and innovative way.[16]

Movie theaters that leased *Gone with the Wind* also had to lease *Getting Gertie's Garter*. (Movie theaters pay the film companies or their distributors a daily or weekly fee for the films they lease.) In other words, these two films were **bundled**—i.e., sold as a package. Why would the film company do this?

• **bundling** Practice of selling two or more products as a package.

You might think that the answer is obvious: *Gone with the Wind* was a great film and *Gertie* was a lousy film, so bundling the two forced movie theaters to lease *Gertie*. But this answer doesn't make economic sense. Suppose a theater's reservation price (the maximum price it will pay) for *Gone with the Wind* is $12,000 per week, and its reservation price for *Gertie* is $3000 per week. Then the most it would pay for *both* films is $15,000, whether it takes the films individually or as a package.

Bundling makes sense when *customers have heterogeneous demands* and when the firm cannot price discriminate. With films, different movie theaters serve different groups of patrons and therefore different theaters may face different demands for films. For example, different theaters might appeal to different age groups, who in turn have different relative film preferences.

[15] Adjusted for inflation, *Gone with the Wind* was also the largest grossing film of all time. *Titanic*, released in 1997, made $601 million. *Gone with the Wind* grossed $81.5 million in 1939 dollars, which is equivalent to $941 million in 1997 dollars.

[16] For those readers who claim to know all this, our final trivia question is: Who played the role of Gertie in *Getting Gertie's Garter*?

To see how a film company can use customer heterogeneity to its advantage, suppose that there are *two* movie theaters and that their reservation prices for our two films are as follows:

	Gone with the Wind	Getting Gertie's Garter
Theater *A*	\$12,000	\$3000
Theater *B*	\$10,000	\$4000

If the films are rented separately, the maximum price that could be charged for *Wind* is \$10,000 because charging more would exclude Theater *B*. Similarly, the maximum price that could be charged for *Gertie* is \$3000. Charging these two prices would yield \$13,000 from each theater, for a total of \$26,000 in revenue. But suppose the films are *bundled*. Theater *A* values the *pair* of films at \$15,000 (\$12,000 + \$3000), and Theater *B* values the pair at \$14,000 (\$10,000 + \$4000). Therefore, we can charge each theater \$14,000 for the pair of films and earn a total revenue of \$28,000. Clearly, we can earn more revenue (\$2000 more) by bundling the films.

Relative Valuations

Why is bundling more profitable than selling the films separately? Because (in this example) the *relative valuations* of the two films are reversed. In other words, although both theaters would pay much more for *Wind* than for *Gertie*, Theater *A* would pay more than Theater *B* for *Wind* (\$12,000 vs. \$10,000), while Theater *B* would pay more than Theater *A* for *Gertie* (\$4000 vs. \$3000). In technical terms, we say that the demands are *negatively correlated*—the customer willing to pay the most for *Wind* is willing to pay the least for *Gertie*. To see why this is critical, suppose demands were *positively correlated*—that is, Theater *A* would pay more for *both* films:

	Gone with the Wind	Getting Gertie's Garter
Theater *A*	\$12,000	\$4000
Theater *B*	\$10,000	\$3000

The most that Theater *A* would pay for the pair of films is now \$16,000, but the most that Theater *B* would pay is only \$13,000. Thus if we bundled the films, the maximum price that could be charged for the package is \$13,000, yielding a total revenue of \$26,000, the same as by renting the films separately.

Now, suppose a firm is selling two different goods to many consumers. To analyze the possible advantages of bundling, we will use a simple diagram to describe the preferences of the consumers in terms of their reservation prices and their consumption decisions given the prices charged. In Figure 11.12 the horizontal axis is r_1, which is the reservation price of a consumer for good 1, and the vertical axis is r_2, which is the reservation price for good 2. The figure shows the reservation prices for three consumers. Consumer *A* is willing to pay up to \$3.25 for good 1 and up to \$6 for good 2; consumer *B* is willing to pay up to \$8.25 for good 1 and up to \$3.25 for good 2; and consumer *C* is willing to pay up to \$10 for each of the goods. In general, the reservation prices for any number of consumers can be plotted this way.

Suppose that there are many consumers and that the products are sold separately, at prices P_1 and P_2, respectively. Figure 11.13 shows how consumers can

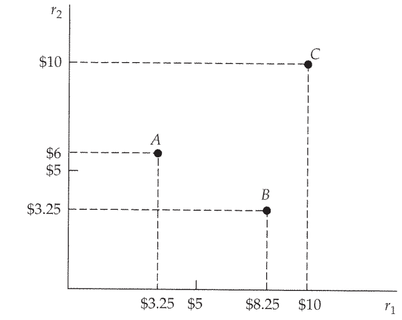

FIGURE 11.12 Reservation Prices

Reservation prices r_1 and r_2 for two goods are shown for three consumers, labeled *A*, *B*, and *C*. Consumer *A* is willing to pay up to \$3.25 for good 1 and up to \$6 for good 2.

be divided into groups. Consumers in region I of the graph have reservation prices that are above the prices being charged for each of the goods, so they will buy both goods. Consumers in region II have a reservation price for good 2 that is above P_2, but a reservation price for good 1 that is below P_1; they will buy

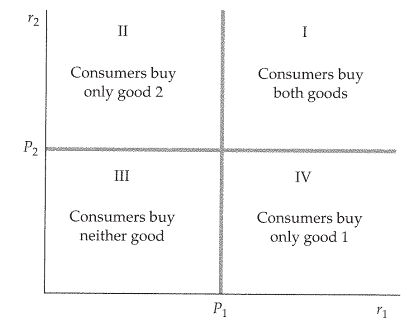

FIGURE 11.13 Consumption Decisions When Products Are Sold Separately

The reservation prices of consumers in region I exceed the prices P_1 and P_2 for the two goods, so these consumers buy both goods. Consumers in regions II and IV buy only one of the goods, and consumers in region III buy neither good.

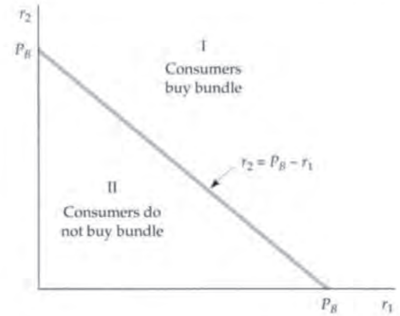

FIGURE 11.14 Consumption Decisions When Products Are Bundled

Consumers compare the *sum* of their reservation prices $r_1 + r_2$, with the price of the bundle P_B. They buy the bundle only if $r_1 + r_2$ is at least as large as P_B.

only good 2. Similarly, consumers in region IV will buy only good 1. Finally, consumers in region III have reservation prices below the prices charged for each of the goods, and so will buy neither.

Now suppose the goods are sold only as a bundle, for a total price of P_B. We can then divide the graph into two regions, as in Figure 11.14. Any given consumer will buy the bundle only if its price is less than or equal to the sum of that consumer's reservation prices for the two goods. The dividing line is therefore the equation $P_B = r_1 + r_2$ or, equivalently, $r_2 = P_B - r_1$. Consumers in region I have reservation prices that add up to more than P_B, so they will buy the bundle. Consumers in region II, who have reservation prices that add up to less than P_B, will not buy the bundle.

Depending on the prices, some of the consumers in region II of Figure 11.14 might have bought one of the goods if they had been sold separately. These consumers are lost to the firm, however, when it sells the goods only as a bundle. The firm, then, must determine whether it can do better by bundling.

In general, the effectiveness of bundling depends on the extent to which demands are negatively correlated. In other words, it works best when consumers who have a high reservation price for good 1 have a low reservation price for good 2, and vice versa. Figure 11.15 shows two extremes. In part (a), each point represents the two reservation prices of a consumer. Note that the demands for the two goods are perfectly positively correlated—consumers with a high reservation price for good 1 also have a high reservation price for good 2. If the firm bundles and charges a price $P_B = P_1 + P_2$, it will make the same profit that it would make by selling the goods separately at prices P_1 and P_2. In part (b), on the other hand, demands are perfectly negatively correlated—a higher reservation price for good 2 implies a proportionately lower one for good 1. In this case, bundling is the ideal strategy. By charging the price P_B the firm can capture *all* the consumer surplus.

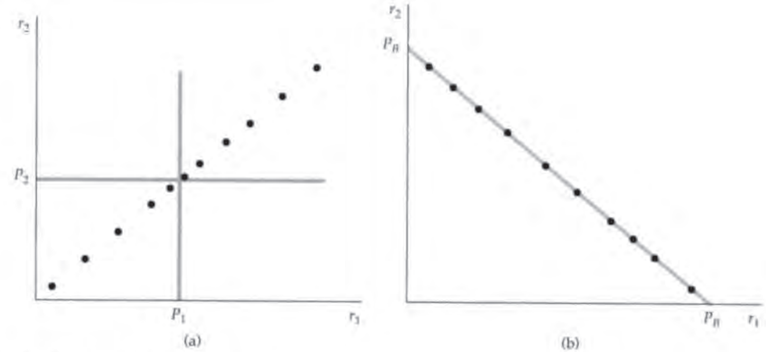

FIGURE 11.15 Reservation Prices

In (a), because demands are perfectly positively correlated, the firm does not gain by bundling: It would earn the same profit by selling the goods separately. In (b), demands are perfectly negatively correlated. Bundling is the ideal strategy—all the consumer surplus can be extracted.

Figure 11.16, which shows the movie example that we introduced at the beginning of this section, illustrates how the demands of the two movie theaters are negatively correlated. (Theater *A* will pay relatively more for *Gone with the Wind*, but Theater *B* will pay relatively more for *Getting Gertie's Garter*.) This makes it more profitable to rent the films as a bundle priced at $14,000.

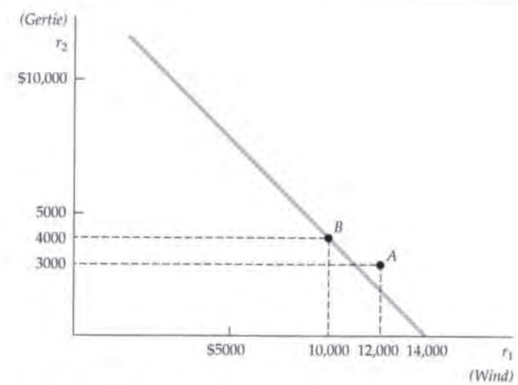

FIGURE 11.16 Movie Example

Consumers *A* and *B* are two movie theaters. The diagram shows their reservation prices for the films *Gone with the Wind* and *Getting Gertie's Garter*. Because the demands are negatively correlated, bundling pays.

Mixed Bundling

• **mixed bundling** Selling two or more goods both as a package and individually.

• **pure bundling** Selling products only as a package.

So far, we have assumed that the firm has two options: to sell the goods either separately or as a bundle. But there is a third option, called **mixed bundling**. As the name suggests, the firm offers its products *both* separately and as a bundle, with a package price below the sum of the individual prices. (We use the term **pure bundling** to refer to the strategy of selling the products *only* as a bundle.) Mixed bundling is often the ideal strategy when demands are only somewhat negatively correlated and/or when marginal production costs are significant. (Thus far, we have assumed that marginal production costs are zero.)

In Figure 11.17, mixed bundling is the most profitable strategy. Although demands are perfectly negatively correlated, there are significant marginal costs. (The marginal cost of producing good 1 is $20, and the marginal cost of producing good 2 is $30.) We have four consumers, labeled *A* through *D*. Now, let's compare three strategies:

1. Selling the goods separately at prices $P_1 = \$50$ and $P_2 = \$90$
2. Selling the goods only as a bundle at a price of $100
3. Mixed bundling, whereby the goods are offered separately at prices $P_1 = P_2 =$ $89.95, or as a bundle at a price of $100.

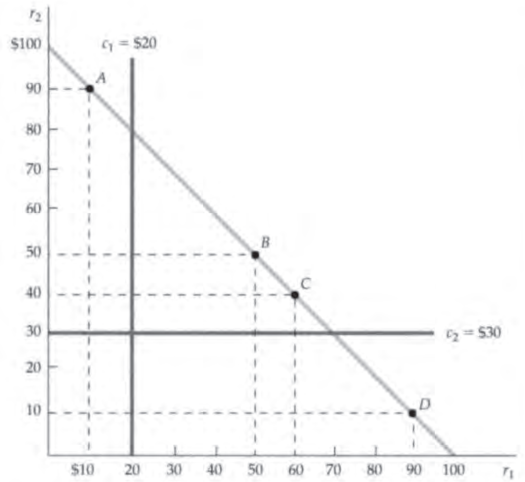

FIGURE 11.17 Mixed versus Pure Bundling

With positive marginal costs, mixed bundling may be more profitable than pure bundling. Consumer *A* has a reservation price for good 1 that is below marginal cost c_1, and consumer *D* has a reservation price for good 2 that is below marginal cost c_2. With mixed bundling, consumer *A* is induced to buy only good 2, and consumer *D* is induced to buy only good 1, thus reducing the firm's cost.

TABLE 11.4 Bundling Example

	P_1	P_2	P_B	Profit
Sold separately	$50	$90	—	$150
Pure bundling	—	—	$100	$200
Mixed bundling	$89.95	$89.95	$100	$229.90

Table 11.4 shows these three strategies and the resulting profits. (You can try other prices for P_1, P_2, and P_B to verify that those given in the table maximize profit for each strategy.) When the goods are sold separately, only consumers *B*, *C*, and *D* buy good 1, and only consumer *A* buys good 2; total profit is 3($50 – $20) + 1($90 – $30) = $150. With pure bundling, all four consumers buy the bundle for $100, so that total profit is 4($100 – $20 – $30) = $200. As we should expect, pure bundling is better than selling the goods separately because consumers' demands are negatively correlated. But what about mixed bundling? Consumer *D* buys only good 1 for $89.95, consumer *A* buys only good 2 for $89.95, and consumers *B* and *C* buy the bundle for $100. Total profit is now ($89.95 – $20) + ($89.95 – $30) + 2($100 – $20 – $30) = $229.90.[17]

In this case, mixed bundling is the most profitable strategy, even though demands are perfectly negatively correlated (i.e., all four consumers have reservation prices on the line $r_2 = 100 - r_1$). Why? For each good, marginal production cost exceeds the reservation price of one consumer. For example, consumer *A* has a reservation price of $90 for good 2 but a reservation price of only $10 for good 1. Because the cost of producing a unit of good 1 is $20, the firm would prefer that consumer *A* buy only good 2, not the bundle. It can achieve this goal by offering good 2 separately for a price just below consumer *A*'s reservation price, while also offering the bundle at a price acceptable to consumers *B* and *C*.

Mixed bundling would *not* be the preferred strategy in this example if marginal costs were zero: In that case, there would be no benefit in excluding consumer *A* from buying good 1 and consumer *D* from buying good 2. We leave it to you to demonstrate this (see Exercise 12).[18]

If marginal costs are zero, mixed bundling can still be more profitable than pure bundling if consumers' demands are not perfectly negatively correlated. (Recall that in Figure 11.17, the reservation prices of the four consumers are perfectly negatively correlated.) This is illustrated by Figure 11.18, in which we have modified the example of Figure 11.17. In Figure 11.18, marginal costs are zero, but the reservation prices for consumers *B* and *C* are now higher. Once again, let's compare three strategies: selling the two goods separately, pure bundling, and mixed bundling.

Table 11.5 shows the optimal prices and the resulting profits for each strategy. (Once again, you should try other prices for P_1, P_2, and P_B to verify that those given in the table maximize profit for each strategy.) When the goods are sold

[17] Note that in the mixed bundling strategy, goods 1 and 2 are priced at $89.95 rather than at $90. If they were priced at $90, consumers *A* and *D* would be indifferent between buying a single good and buying the bundle, and if they buy the bundle, total profit will be lower.

[18] Sometimes a firm with monopoly power will find it profitable to bundle its product with the product of another firm; see Richard L. Schmalensee, "Commodity Bundling by Single-Product Monopolies," *Journal of Law and Economics* 25 (April 1982): 67–71. Bundling can also be profitable when the products are substitutes or complements. See Arthur Lewbel, "Bundling of Substitutes or Complements," *International Journal of Industrial Organization* 3 (1985): 101–7.

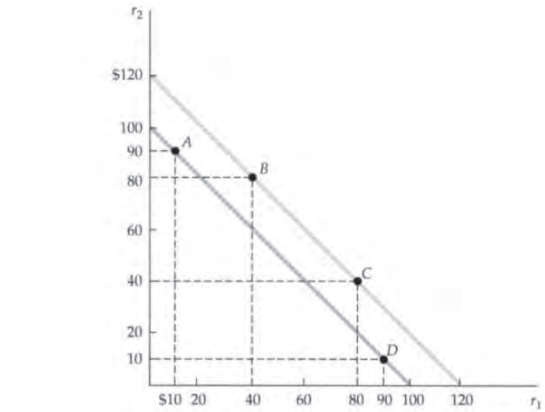

FIGURE 11.18 **Mixed Bundling with Zero Marginal Costs**

If marginal costs are zero, and if consumers' demands are not perfectly negatively correlated, mixed bundling is still more profitable than pure bundling. In this example, consumers *B* and *C* are willing to pay $20 more for the bundle than are consumers *A* and *D*. With pure bundling, the price of the bundle is $100. With mixed bundling, the price of the bundle can be increased to $120 and consumers *A* and *D* can still be charged $90 for a single good.

separately, only consumers *C* and *D* buy good 1, and only consumers *A* and *B* buy good 2; total profit is thus $320. With pure bundling, all four consumers buy the bundle for $100, so that total profit is $400. As expected, pure bundling is better than selling the goods separately because consumers' demands are negatively correlated. But mixed bundling is better still. With mixed bundling, consumer *A* buys only good 2, consumer *D* buys only good 1, and consumers *B* and *C* buy the bundle at a price of $120. Total profit is now $420.

Why does mixed bundling give higher profits than pure bundling even though marginal costs are zero? The reason is that demands are not perfectly negatively correlated: The two consumers who have high demands for both goods (*B* and *C*) are willing to pay more for the bundle than are consumers *A* and *D*. With mixed bundling, therefore, we can increase the price of the bundle (from $100 to $120), sell this bundle to two consumers, and charge the remaining consumers $90 for a single good.

TABLE 11.5 Mixed Bundling with Zero Marginal Costs

	P_1	P_2	P_B	Profit
Sell separately	$80	$80	—	$320
Pure bundling	—	—	$100	$400
Mixed bundling	$90	$90	$120	$420

Bundling in Practice

Bundling is a widely used pricing strategy. When you buy a new car, for example, you can purchase such options as power windows, power seats, or a sunroof separately, or you can purchase a "luxury package" in which these options are bundled. Manufacturers of luxury cars (such as Lexus, BMW, or Infiniti) tend to include such "options" as standard equipment; this practice is pure bundling. For more moderately priced cars, however, these items are optional, but are usually offered as part of a bundle. Automobile companies must decide which items to include in such bundles and how to price them.

Another example is vacation travel. If you plan a vacation to Europe, you might make your own hotel reservations, buy an airplane ticket, and order a rental car. Alternatively, you might buy a vacation package in which airfare, land arrangements, hotels, and even meals are all bundled together.

Still another example is cable television. Cable operators typically offer a basic service for a low monthly fee, plus individual "premium" channels, such as Cinemax, Home Box Office, and the Disney Channel, on an individual basis for additional monthly fees. However, they also offer packages in which two or more premium channels are sold as a bundle. Bundling cable channels is profitable because demands are negatively correlated. How do we know that? Given that there are only 24 hours in a day, the time that a consumer spends watching HBO is time that cannot be spent watching the Disney Channel. Thus consumers with high reservation prices for some channels will have relatively low reservation prices for others.

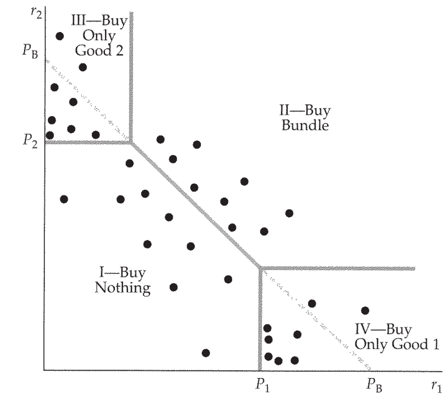

FIGURE 11.19 **Mixed Bundling in Practice**

The dots in this figure are estimates of reservation prices for a representative sample of consumers. A company could first choose a price for the bundle, P_B, such that a diagonal line connecting these prices passes roughly midway through the dots. The company could then try individual prices P_1 and P_2. Given P_1, P_2, and P_B, profits can be calculated for this sample of consumers. Managers can then raise or lower P_1, P_2, and P_B and see whether the new pricing leads to higher profits. This procedure is repeated until total profit is roughly maximized.

How can a company decide whether to bundle its products, and determine the profit-maximizing prices? Most companies do not know their customers' reservation prices. However, by conducting market surveys, they may be able to estimate the distribution of reservation prices, and then use this information to design a pricing strategy.

This is illustrated in Figure 11.19. The dots are estimates of reservation prices or a representative sample of consumers (obtained, say, from a market survey). The company might first choose a price for the bundle, P_B, such that a diagonal line connecting these prices passes roughly midway through the dots in the figure. It could then try individual prices P_1 and P_2. Given P_1, P_2, and P_B, we can separate consumers into four regions, as shown in the figure. Consumers in Region I buy nothing (because $r_1 < P_1$, $r_2 < P_2$, and $r_1 + r_2 < P_B$). Consumers in Region II buy the bundle (because $r_1 + r_2 > P_B$). Consumers in Region III buy only good 2 (because $r_2 > P_2$ but $r_1 < P_B - P_2$). Likewise, consumers in Region IV buy only good 1. Given this distribution, we can calculate the resulting profits. We can then raise or lower P_1, P_2, and P_B and see whether doing so leads to higher profits. This can be done repeatedly (on a computer) until prices are found that roughly maximize total profit.

EXAMPLE 11.6 The Complete Dinner versus à la Carte: A Restaurant's Pricing Problem

Many restaurants offer both complete dinners and à la carte menus. Why? Most customers go out to eat knowing roughly how much they are willing to spend for dinner (and choose the restaurant accordingly). Diners, however, have different preferences. For example, some value appetizers highly but could happily skip dessert. Others attach little value to the appetizer but regard dessert as essential. And some customers attach moderate values to both appetizers and desserts. What pricing strategy lets the restaurant capture as much consumer surplus as possible from these heterogeneous customers? The answer, of course, is mixed bundling.

For a restaurant, mixed bundling means offering both complete dinners (the appetizer, main course, and dessert come as a package) and an à la carte menu (the customer buys the appetizer, main course, and dessert separately). This strategy allows the à la carte menu to be priced to capture consumer surplus from customers who value some dishes much more highly than others. (Such customers would correspond to consumers *A* and *D* in Figure 11.17 (page 418).) At the same time, the complete dinner retains those customers who have lower variations in their reservation prices for different dishes (e.g., customers who attach moderate values to both appetizers and desserts).

For example, if the restaurant expects to attract customers willing to spend about \$20 for dinner, it might charge about \$5 for appetizers, \$14 for a typical main dish, and \$4 for dessert. It could also offer a complete dinner, which includes an appetizer, main course, and dessert, for \$20. Then, the customer who loves dessert but couldn't care less about an appetizer will order only the main dish and dessert, and spend \$18 (saving the restaurant the cost of preparing an appetizer). At the same time, another customer who attaches a moderate value (say, \$3 or \$3.50) to both the appetizer and dessert will buy the complete dinner.

TABLE 11.6 Mixed Bundling at McDonald's (2007)

Individual Item	Price	Meal (Includes Soda and Fries)	Unbundled Price	Price of Bundle	Savings
Chicken Sandwich	\$3.49	Chicken Sandwich	\$7.77	\$5.89	\$1.88
Filet-O-Fish	\$2.59	Filet-O-Fish	\$6.87	\$4.89	\$1.98
Big Mac	\$2.99	Big Mac	\$7.27	\$5.29	\$1.98
Quarter Pounder	\$3.09	Quarter Pounder	\$7.37	\$5.39	\$1.98
Double Quarter Pounder	\$3.69	Double Quarter Pounder	\$7.97	\$5.99	\$1.98
10-piece Chicken McNuggets	\$3.89	10-piece Chicken McNuggets	\$8.17	\$6.19	\$1.98
Large French Fries	\$2.29				
Large Soda	\$1.99				

You don't have to go to an expensive French restaurant to experience mixed bundling. Table 11.6 shows the prices of some individual items at a Boston-area McDonald's, as well as the prices of "super meals" that include meat or fish items along with a large order of French fries and a large soda. Note that you can buy a Big Mac, a large fries, and a large soda separately for a total of \$7.27, or you can buy them as a bundle for \$5.29. You say you don't care for fries? Then just buy the Big Mac and large soda separately, for a total of \$4.98, which is \$0.31 less than the price of the bundle.

Unfortunately for consumers, perhaps, creative pricing is sometimes more important than creative cooking for the financial success of a restaurant. Successful restaurateurs know their customers' demand characteristics and use that knowledge to design a pricing strategy that extracts as much consumer surplus as possible.

Tying

Tying is a general term that refers to any requirement that products be bought or sold in some combination. Pure bundling is a common form of tying, but tying can also take other forms. For example, suppose a firm sells a product (such as a copying machine) that requires the consumption of a secondary product (such as paper). The consumer who buys the first product is also required to buy the secondary product from the same company. This requirement is usually imposed through a contract. Note that this is different from the examples of bundling discussed earlier. In those examples, the consumer might have been happy to buy just one of the products. In this case, however, the first product is useless without access to the secondary product.

• **tying** Practice of requiring a customer to purchase one good in order to purchase another.

Why might firms use this kind of pricing practice? One of the main benefits of tying is that it often allows a firm to *meter demand* and thereby practice price discrimination more effectively. During the 1950s, for example, when Xerox had a monopoly on copying machines but not on paper, customers who leased Xerox copiers also had to buy Xerox paper. This allowed Xerox to meter consumption (customers who used a machine intensively bought more paper), and thereby apply a two-part tariff to the pricing of its machines. Also during the 1950s, IBM required customers who leased its mainframe computers to use paper computer cards made only by IBM. By pricing cards well above marginal cost, IBM was effectively charging higher prices for computer usage to customers with larger demands.[19]

Tying can also be used to extend a firm's market power. As we discussed in Example 10.6 (page 385), in 1998 the Department of Justice brought suit against Microsoft, claiming that the company had tied its Internet Explorer Web browser to its Windows 98 operating system in order to maintain its monopoly power in the market for PC operating systems.

Tying can have other uses. An important one is to protect customer goodwill connected with a brand name. This is why franchises are often required to purchase inputs from the franchiser. For example, Mobil Oil requires its service stations to sell only Mobil motor oil, Mobil batteries, and so on. Similarly, until recently, a McDonald's franchisee had to purchase all materials and supplies—from the hamburgers to the paper cups—from McDonald's, thus ensuring product uniformity and protecting the brand name.[20]

*11.6 ADVERTISING

We have seen how firms can utilize their market power when making pricing decisions. Pricing is important for a firm, but most firms with market power have another important decision to make: how much to advertise. In this section, we will see how firms with market power can make profit-maximizing advertising decisions, and how those decisions depend on the characteristics of demand for the firm's product.[21]

For simplicity, we will assume that the firm sets only one price for its product. We will also assume that having done sufficient market research, it knows how its quantity demanded depends on both its price P *and* its advertising expenditures in dollars A; that is, it knows $Q(P,A)$. Figure 11.20 shows the firm's demand and cost curves with and without advertising. AR and MR are the firm's average and marginal revenue curves when it does not advertise, and AC and MC are its average and marginal cost curves. It produces a quantity Q_0, where MR = MC, and receives a price P_0. Its profit per unit is the difference between P_0 and average cost, so its total profit π_0 is given by the gray-shaded rectangle.

In §7.1, marginal cost—the increase in cost that results from producing one extra unit of output—is distinguished from average cost—the cost per unit of output.

Now suppose the firm advertises. This causes its demand curve to shift out and to the right; the new average and marginal revenue curves are given by AR' and MR'. Advertising is a fixed cost, so the firm's average cost curve rises (to AC'). Marginal cost, however, remains the same. With advertising, the firm produces Q_1 (where MR' = MC) and receives a price P_1. Its total profit π_1, given by the purple-shaded rectangle, is now much larger.

Although the firm in Figure 11.20 is clearly better off when it advertises, the figure does not help us determine *how much* advertising it should do. It must choose its price P and advertising expenditure A to maximize profit, which is now given by:

$$\pi = PQ(P, A) - C(Q) - A$$

[19]Antitrust actions ultimately forced IBM to discontinue this pricing practice.

[20]In some cases, the courts have ruled that tying is not necessary to protect customer goodwill and is anticompetitive. Today, a McDonald's franchisee can buy supplies from any McDonald's-approved source. For a discussion of some of the antitrust issues involved in franchise tying, see Benjamin Klein and Lester F. Saft, "The Law and Economics of Franchise Tying Contracts," *Journal of Law and Economics* 28 (May 1985): 345–61.

[21]A perfectly competitive firm has little reason to advertise: By definition it can sell as much as it produces at a market price that it takes as given. That is why it would be unusual to see a producer of corn or soybeans advertise.

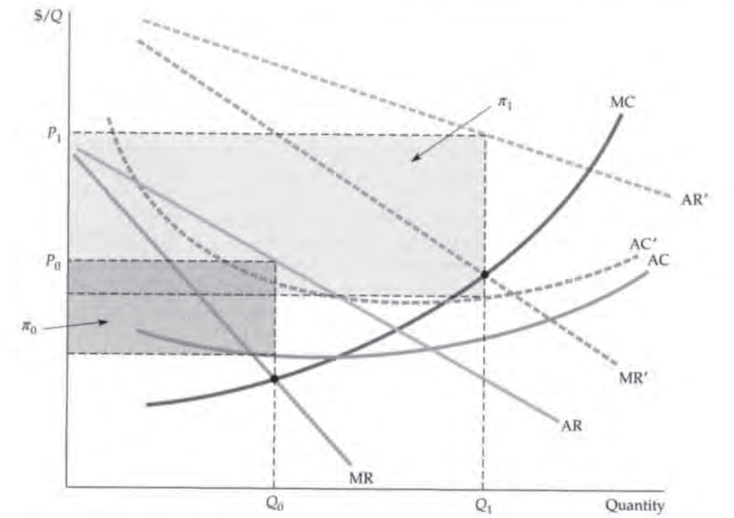

FIGURE 11.20 Effects of Advertising

AR and MR are average and marginal revenue when the firm doesn't advertise, and AC and MC are average and marginal cost. The firm produces Q_0 and receives a price P_0. Its total profit π_0 is given by the gray-shaded rectangle. If the firm advertises, its average and marginal revenue curves shift to the right. Average cost rises (to AC') but marginal cost remains the same. The firm now produces Q_1 (where MR' = MC), and receives a price P_1. Its total profit, π_1, is now larger.

Given a price, more advertising will result in more sales and thus more revenue. But what is the firm's profit-maximizing advertising expenditure? You might be tempted to say that the firm should increase its advertising expenditures until the last dollar of advertising just brings forth an additional dollar of revenue—that is, until the marginal revenue from advertising, $\Delta(P, Q)/\Delta A$, is just equal to 1. But as Figure 11.20 shows, this reasoning omits an important element. Remember that *advertising leads to increased output* (in the figure, output increased from Q_0 to Q_1). But increased output in turn means increased production costs, and this must be taken into account when comparing the costs and benefits of an extra dollar of advertising.

The correct decision is to increase advertising until the marginal revenue from an additional dollar of advertising, MR_{Ads}, just equals the *full* marginal cost of that advertising. That full marginal cost is the sum of the dollar spent directly on the advertising and the marginal production cost resulting from the increased sales that advertising brings about. Thus the firm should advertise up to the point that

$$MR_{Ads} = P\frac{\Delta Q}{\Delta A} = 1 + MC\frac{\Delta Q}{\Delta A} = \textit{full} \text{ marginal cost of advertising} \qquad (11.3)$$

This rule is often ignored by managers, who justify advertising budgets by comparing the expected benefits (i.e., added sales) only with the cost of the advertising. But additional sales mean increased production costs that must also be taken into account.[22]

A Rule of Thumb for Advertising

Like the rule MR = MC, equation (11.3) is sometimes difficult to apply in practice. In Chapter 10, we saw that MR = MC implies the following rule of thumb for pricing: $(P - MC)/P = -1/E_D$, where E_D is the firm's price elasticity of demand. We can combine this rule of thumb for pricing with equation (11.3) to obtain a rule of thumb for advertising.

In equation (10.1), we offer a rule of thumb for pricing for a profit-maximizing firm—the markup over marginal cost as a percentage of price should equal minus the inverse of the price elasticity of demand.

First, rewrite equation (11.3) as follows:

$$(P - \text{MC})\frac{\Delta Q}{\Delta A} = 1$$

Now multiply both sides of this equation by A/PQ, the **advertising-to-sales ratio**:

advertising-to-sales ratio Ratio of a firm's advertising expenditures to its sales.

$$\frac{P - \text{MC}}{P}\left[\frac{A}{Q}\frac{\Delta Q}{\Delta A}\right] = \frac{A}{PQ}$$

The term in brackets, $(A/Q)(\Delta Q/\Delta A)$, is the **advertising elasticity of demand**, the percentage change in the quantity demanded that results from a 1-percent increase in advertising expenditures. We will denote this elasticity by E_A. Because $(P - \text{MC})/P$ must equal $-1/E_P$, we can rewrite this equation as follows:

advertising elasticity of demand Percentage change in quantity demanded resulting from a 1-percent increase in advertising expenditures.

$$A/PQ = -(E_A/E_P) \qquad \textbf{(11.4)}$$

Equation (11.4) is a rule of thumb for advertising. It says that to maximize profit, the firm's advertising-to-sales ratio should be equal to minus the ratio of the advertising and price elasticities of demand. Given information (from, say, market research studies) on these two elasticities, the firm can use this rule to check that its advertising budget is not too small or too large.

To put this rule into perspective, assume that a firm is generating sales revenue of $1 million per year while allocating only $10,000 (1 percent of its revenues) to advertising. The firm knows that its advertising elasticity of demand is .2, so that a doubling of its advertising budget from $10,000 to $20,000 should increase sales by 20 percent. The firm also knows that the price elasticity of demand for its product is –4. Should it increase its advertising budget, knowing that with a price elasticity of demand of –4, its markup of price over marginal cost is substantial? The answer is yes; equation (11.4) tells us that the firm's advertising-to-sales ratio should be $-(.2/-4) = 5$ percent, so the firm should increase its advertising budget from $10,000 to $50,000.

This rule makes intuitive sense. It says firms should advertise a lot if (i) demand is very sensitive to advertising (E_A is large), or if (ii) demand is not very price elastic (E_P is small). Although (i) is obvious, why should firms advertise more when the price elasticity of demand is small? A small elasticity of demand implies a large markup of price over marginal cost. Therefore, the marginal profit from each extra unit sold is high. In this case, if advertising can help sell a few more units, it will be worth its cost.[23]

[22]To derive this result using calculus, differentiate $\pi(Q,A)$ with respect to A, and set the derivative equal to zero:

$$\partial\pi/\partial A = P(\partial Q/\partial A) - \text{MC}(\partial Q/\partial A) - 1 = 0$$

Rearranging gives equation (11.3).

EXAMPLE 11.7 Advertising in Practice

In Example 10.2 (page 364), we looked at the use of markup pricing by supermarkets, convenience stores, and makers of designer jeans. We saw in each case how the markup of price over marginal cost depended on the firm's price elasticity of demand. Now let's see why these firms, as well as producers of other goods, advertise as much (or as little) as they do.

First, supermarkets. We said that the price elasticity of demand for a typical supermarket is around –10. To determine the advertising-to-sales ratio, we also need to know the advertising elasticity of demand. This number can vary considerably depending on what part of the country the supermarket is located in and whether it is in a city, suburb, or rural area. A reasonable range, however, would be 0.1 to 0.3. Substituting these numbers into equation (11.4), we find that the manager of a typical supermarket should have an advertising budget of around 1 to 3 percent of sales—which is indeed what many supermarkets spend on advertising.

Convenience stores have lower price elasticities of demand (around –5), but their advertising-to-sales ratios are usually less than those for supermarkets (and are often zero). Why? Because convenience stores mostly serve customers who live nearby; they may need a few items late at night or may simply not want to drive to the supermarket. These customers already know about the convenience store and are unlikely to change their buying habits if the store advertises. Thus E_A is very small, and advertising is not worthwhile.

Advertising is quite important for makers of designer jeans, who will have advertising-to-sales ratios as high as 10 or 20 percent. Advertising helps to make consumers aware of the label and gives it an aura and image. We said that price elasticities of demand in the range of –3 to –4 are typical for the major labels, and advertising elasticities of demand can range from .3 to as high as 1. So, these levels of advertising would seem to make sense.

Laundry detergents have among the highest advertising-to-sales ratios of all products, sometimes exceeding 30 percent, even though demand for any one brand is at least as price elastic as it is for designer jeans. What justifies all the advertising? A very large advertising elasticity. The demand for any one brand of laundry detergent depends crucially on advertising; without it, consumers would have little basis for selecting that particular brand.[24]

[23]Advertising often affects the price elasticity of demand, and this fact must be taken into account. For some products, advertising broadens the market by attracting a large range of customers, or by creating a bandwagon effect. This is likely to make demand more price elastic than it would have been otherwise. (But E_A is likely to be large, so that advertising will still be worthwhile.) Sometimes advertising is used to differentiate a product from others (by creating an image, allure, or brand identification), making the product's demand less price elastic than it would otherwise be.

[24]For an overview of statistical approaches to estimating the advertising elasticity of demand, see Ernst R. Berndt, *The Practice of Econometrics* (Reading, MA: Addison-Wesley, 1991), ch. 8.

TABLE 11.7 Sales and Advertising Expenditures for Leading Brands of Over-the-Counter Drugs (in millions of dollars)

	Sales	Advertising	Ratio (%)
Pain Medications			
Tylenol	855	143.8	17
Advil	360	91.7	26
Bayer	170	43.8	26
Excedrin	130	26.7	21
Antacids			
Alka-Seltzer	160	52.2	33
Mylanta	135	32.8	24
Tums	135	27.6	20
Cold Remedies (decongestants)			
Benadryl	130	30.9	24
Sudafed	115	28.6	25
Cough Medicine			
Vicks	350	26.6	8
Robitussin	205	37.7	19
Halls	130	17.4	13

Source: New York Times, September 27, 1994.

Finally, Table 11.7 shows sales, advertising expenditures, and the ratio of the two for leading brands of over-the-counter drugs. Observe that overall, the ratios are quite high. As with laundry detergents, the advertising elasticity for name-brand drugs is very high. Alka-Seltzer, Mylanta, and Tums, for instance, are all antacids that do much the same thing. Sales depend on consumer identification with a particular brand, which requires advertising.

SUMMARY

1. Firms with market power are in an enviable position because they have the potential to earn large profits. Realizing that potential, however, may depend critically on pricing strategy. Even if the firm sets a single price, it needs an estimate of the elasticity of demand for its output. More complicated strategies, which can involve setting several different prices, require even more information about demand.
2. A pricing strategy aims to enlarge the customer base that the firm can sell to and capture as much consumer surplus as possible. There are a number of ways to do this, and they usually involve setting more than a single price.
3. Ideally, the firm would like to price discriminate perfectly—i.e., to charge each customer his or her reservation price. In practice, this is almost always impossible. On the other hand, various forms of imperfect price discrimination are often used to increase profits.
4. The two-part tariff is another means of capturing consumer surplus. Customers must pay an "entry" fee that allows them to buy the good at a per-unit price. The two-part tariff is most effective when customer demands are relatively homogeneous.
5. When demands are heterogeneous and negatively correlated, bundling can increase profits. With pure bundling, two or more different goods are sold only as a package. With mixed bundling, the customer can buy the goods individually or as a package. Mixed bundling can be more profitable than pure bundling if marginal costs are significant or if demands are not perfectly negatively correlated.
6. Bundling is a special case of tying, a requirement that products be bought or sold in some combination. Tying can be used to meter demand or to protect customer goodwill associated with a brand name.
7. Advertising can further increase profits. The profit-maximizing advertising-to-sales ratio is equal in magnitude to the ratio of the advertising and price elasticities of demand.

QUESTIONS FOR REVIEW

1. Suppose a firm can practice perfect, first-degree price discrimination. What is the lowest price it will charge, and what will its total output be?
2. How does a car salesperson practice price discrimination? How does the ability to discriminate correctly affect his or her earnings?
3. Electric utilities often practice second-degree price discrimination. Why might this improve consumer welfare?
4. Give some examples of third-degree price discrimination. Can third-degree price discrimination be effective if the different groups of consumers have different levels of demand but the same price elasticities?
5. Show why optimal, third-degree price discrimination requires that marginal revenue for each group of consumers equals marginal cost. Use this condition to explain how a firm should change its prices and total output if the demand curve for one group of consumers shifts outward, causing marginal revenue for that group to increase.
6. When pricing automobiles, American car companies typically charge a much higher percentage markup over cost for "luxury option" items (such as leather trim, etc.) than for the car itself or for more "basic" options such as power steering and automatic transmission. Explain why.
7. How is peak-load pricing a form of price discrimination? Can it make consumers better off? Give an example.
8. How can a firm determine an optimal two-part tariff if it has two customers with different demand curves? (Assume that it knows the demand curves.)
9. Why is the pricing of a Gillette safety razor a form of two-part tariff? Must Gillette be a monopoly producer of its blades as well as its razors? Suppose you were advising Gillette on how to determine the two parts of the tariff. What procedure would you suggest?
10. In the town of Woodland, California, there are many dentists but only one eye doctor. Are senior citizens more likely to be offered discount prices for dental exams or for eye exams? Why?
11. Why did MGM bundle *Gone with the Wind* and *Getting Gertie's Garter*? What characteristic of demands is needed for bundling to increase profits?
12. How does mixed bundling differ from pure bundling? Under what conditions is mixed bundling preferable to pure bundling? Why do many restaurants practice mixed bundling (by offering a complete dinner as well as an à la carte menu) instead of pure bundling?
13. How does tying differ from bundling? Why might a firm want to practice tying?
14. Why is it incorrect to advertise up to the point that the last dollar of advertising expenditures generates another dollar of sales? What is the correct rule for the marginal advertising dollar?
15. How can a firm check that its advertising-to-sales ratio is not too high or too low? What information does it need?

EXERCISES

1. Price discrimination requires the ability to sort customers and the ability to prevent arbitrage. Explain how the following can function as price discrimination schemes and discuss both sorting and arbitrage:
   a. Requiring airline travelers to spend at least one Saturday night away from home to qualify for a low fare.
   b. Insisting on delivering cement to buyers and basing prices on buyers' locations.
   c. Selling food processors along with coupons that can be sent to the manufacturer for a $10 rebate.
   d. Offering temporary price cuts on bathroom tissue.
   e. Charging high-income patients more than low-income patients for plastic surgery.
2. If the demand for drive-in movies is more elastic for couples than for single individuals, it will be optimal for theaters to charge one admission fee for the driver of the car and an extra fee for passengers. True or false? Explain.
3. In Example 11.1 (page 400), we saw how producers of processed foods and related consumer goods use coupons as a means of price discrimination. Although coupons are widely used in the United States, that is not the case in other countries. In Germany, coupons are illegal.
   a. Does prohibiting the use of coupons in Germany make German *consumers* better off or worse off?

b. Does prohibiting the use of coupons make German *producers* better off or worse off?

4. Suppose that BMW can produce any quantity of cars at a constant marginal cost equal to \$20,000 and a fixed cost of \$10 billion. You are asked to advise the CEO as to what prices and quantities BMW should set for sales in Europe and in the United States. The demand for BMWs in each market is given by

$$Q_E = 4{,}000{,}000 - 100P_E$$

and

$$Q_U = 1{,}000{,}000 - 20P_U$$

where the subscript E denotes Europe, the subscript U denotes the United States. Assume that BMW can restrict U.S. sales to authorized BMW dealers only.
   a. What quantity of BMWs should the firm sell in each market, and what should the price be in each market? What should the total profit be?
   b. If BMW were forced to charge the same price in each market, what would be the quantity sold in each market, the equilibrium price, and the company's profit?

5. A monopolist is deciding how to allocate output between two geographically separated markets (East Coast and Midwest). Demand and marginal revenue for the two markets are

$$P_1 = 15 - Q_1 \qquad MR_1 = 15 - 2Q_1$$
$$P_2 = 25 - 2Q_2 \qquad MR_2 = 25 - 4Q_2$$

The monopolist's total cost is $C = 5 + 3(Q_1 + Q_2)$. What are price, output, profits, marginal revenues, and deadweight loss (i) if the monopolist can price discriminate? (ii) if the law prohibits charging different prices in the two regions?

*6. Elizabeth Airlines (EA) flies only one route: Chicago–Honolulu. The demand for each flight is $Q = 500 - P$. EA's cost of running each flight is \$30,000 plus \$100 per passenger.
   a. What is the profit-maximizing price that EA will charge? How many people will be on each flight? What is EA's profit for each flight?
   b. EA learns that the fixed costs per flight are in fact \$41,000 instead of \$30,000. Will the airline stay in business for long? Illustrate your answer using a graph of the demand curve that EA faces, EA's average cost curve when fixed costs are \$30,000, and EA's average cost curve when fixed costs are \$41,000.
   c. Wait! EA finds out that two different types of people fly to Honolulu. Type A consists of business people with a demand of $Q_A = 260 - 0.4P$. Type B consists of students whose total demand is $Q_B = 240 - 0.6P$. Because the students are easy to spot, EA decides to charge them different prices. Graph each of these demand curves and their horizontal sum. What price does EA charge the students? What price does it charge other customers? How many of each type are on each flight?
   d. What would EA's profit be for each flight? Would the airline stay in business? Calculate the consumer surplus of each consumer group. What is the total consumer surplus?
   e. Before EA started price discriminating, how much consumer surplus was the Type A demand getting from air travel to Honolulu? Type B? Why did total consumer surplus decline with price discrimination, even though total quantity sold remained unchanged?

7. Many retail video stores offer two alternative plans for renting films:
   - **A two-part tariff:** Pay an annual membership fee (e.g., \$40) and then pay a small fee for the daily rental of each film (e.g., \$2 per film per day).
   - **A straight rental fee:** Pay no membership fee, but pay a higher daily rental fee (e.g., \$4 per film per day).

   What is the logic behind the two-part tariff in this case? Why offer the customer a choice of two plans rather than simply a two-part tariff?

8. Sal's satellite company broadcasts TV to subscribers in Los Angeles and New York. The demand functions for each of these two groups are

$$Q_{NY} = 60 - 0.25P_{NY}$$
$$Q_{LA} = 100 - 0.50P_{LA}$$

where Q is in thousands of subscriptions per year and P is the subscription price per year. The cost of providing Q units of service is given by

$$C = 1000 + 40Q$$

where $Q = Q_{NY} + Q_{LA}$.
   a. What are the profit-maximizing prices and quantities for the New York and Los Angeles markets?
   b. As a consequence of a new satellite that the Pentagon recently deployed, people in Los Angeles receive Sal's New York broadcasts and people in New York receive Sal's Los Angeles broadcasts. As a result, anyone in New York or Los Angeles can receive Sal's broadcasts by subscribing in either city. Thus Sal can charge only a single price. What price should he charge, and what quantities will he sell in New York and Los Angeles?
   c. In which of the above situations, (a) or (b), is Sal better off? In terms of consumer surplus, which situation do people in New York prefer and which do people in Los Angeles prefer? Why?

*9. You are an executive for Super Computer, Inc. (SC), which rents out super computers. SC receives a fixed rental payment per time period in exchange for the right to unlimited computing at a rate of P cents per second. SC has two types of potential customers of equal number—10 businesses and 10 academic institutions. Each business customer has the demand function $Q = 10 - P$, where Q is in millions of seconds per month; each academic institution has the demand $Q = 8 - P$. The marginal cost to SC of additional computing is 2 cents per second, regardless of volume.
   a. Suppose that you could separate business and academic customers. What rental fee and usage fee would you charge each group? What would be your profits?
   b. Suppose you were unable to keep the two types of customers separate and charged a zero rental fee. What usage fee would maximize your profits? What would be your profits?
   c. Suppose you set up one two-part tariff—that is, you set one rental and one usage fee that both business and academic customers pay. What usage and rental fees would you set? What would be your profits? Explain why price would not be equal to marginal cost.

10. As the owner of the only tennis club in an isolated wealthy community, you must decide on membership dues and fees for court time. There are two types of tennis players. "Serious" players have demand

$$Q_1 = 10 - P$$

where Q_1 is court hours per week and P is the fee per hour for each individual player. There are also "occasional" players with demand

$$Q_2 = 4 - 0.25P$$

Assume that there are 1000 players of each type. Because you have plenty of courts, the marginal cost of court time is zero. You have fixed costs of \$10,000 per week. Serious and occasional players look alike, so you must charge them the same prices.
   a. Suppose that to maintain a "professional" atmosphere, you want to limit membership to serious players. How should you set the *annual* membership dues and court fees (assume 52 weeks per year) to maximize profits, keeping in mind the constraint that only serious players choose to join? What would profits be (per week)?
   b. A friend tells you that you could make greater profits by encouraging both types of players to join. Is your friend right? What annual dues and court fees would maximize weekly profits? What would these profits be?
   c. Suppose that over the years, young, upwardly mobile professionals move to your community, all of whom are serious players. You believe there are now 3000 serious players and 1000 occasional players. Would it still be profitable to cater to the occasional player? What would be the profit-maximizing annual dues and court fees? What would profits be per week?

11. Look again at Figure 11.12 (p. 415), which shows the reservation prices of three consumers for two goods. Assuming that marginal production cost is zero for both goods, can the producer make the most money by selling the goods separately, by using pure bundling, or by using mixed bundling? What prices should be charged?

12. Look again at Figure 11.17 (p. 418). Suppose that the marginal costs c_1 and c_2 were zero. Show that in this case, pure bundling, not mixed bundling, is the most profitable pricing strategy. What price should be charged for the bundle? What will the firm's profit be?

13. Some years ago, an article appeared in the *New York Times* about IBM's pricing policy. The previous day, IBM had announced major price cuts on most of its small and medium-sized computers. The article said:

   > IBM probably has no choice but to cut prices periodically to get its customers to purchase more and lease less. If they succeed, this could make life more difficult for IBM's major competitors. Outright purchases of computers are needed for ever larger IBM revenues and profits, says Morgan Stanley's Ulric Weil in his new book, *Information Systems in the '80's*. Mr. Weil declares that IBM cannot revert to an emphasis on leasing.

   a. Provide a brief but clear argument in *support* of the claim that IBM should try "to get its customers to purchase more and lease less."
   b. Provide a brief but clear argument *against* this claim.
   c. What factors determine whether leasing or selling is preferable for a company like IBM? Explain briefly.

14. You are selling two goods, 1 and 2, to a market consisting of three consumers with reservation prices as follows:

	Reservation Price (\$)	
Consumer	For 1	For 2
A	20	100
B	60	60
C	100	20

The unit cost of each product is \$30.
   a. Compute the optimal prices and profits for (i) selling the goods separately, (ii) pure bundling, and (iii) mixed bundling.
   b. Which strategy would be most profitable? Why?

15. Your firm produces two products, the demands for which are independent. Both products are produced at zero marginal cost. You face four consumers (or groups of consumers) with the following reservation prices:

Consumer	Good 1(\$)	Good 2(\$)
A	25	100
B	40	80
C	80	40
D	100	25

   a. Consider three alternative pricing strategies: (i) selling the goods separately; (ii) pure bundling;

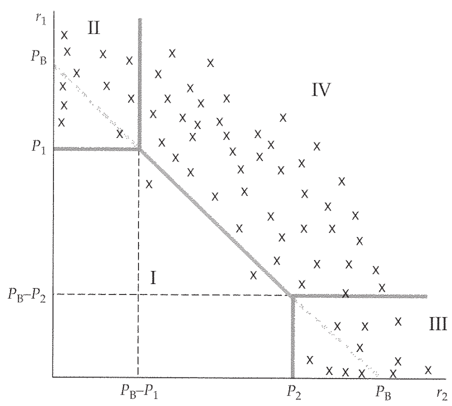

FIGURE 11.21 **Figure for Exercise 16**

(iii) mixed bundling. For *each strategy*, determine the optimal prices to be charged and the resulting profits. Which strategy would be best?

b. Now suppose that the production of each good entails a marginal cost of \$30. How does this information change your answers to (a)? Why is the optimal strategy now different?

16. A cable TV company offers, in addition to its basic service, two products: a Sports Channel (Product 1) and a Movie Channel (Product 2). Subscribers to the basic service can subscribe to these additional services individually at the monthly prices P_1 and P_2, respectively, or they can buy the two as a bundle for the price P_B, where $P_B < P_1 + P_2$. They can also forgo the additional services and simply buy the basic service. The company's marginal cost for these additional services is *zero*. Through market research, the cable company has estimated the reservation prices for these two services for a representative group of consumers in the company's service area. These reservation prices are plotted (as x's) in Figure 11.21, as are the prices P_1, P_2, and P_B that the cable company is currently charging. The graph is divided into regions I, II, III, and IV.

a. Which products, if any, will be purchased by the consumers in region I? In region II? In region III? In region IV? Explain briefly.

b. Note that as drawn in the figure, the reservation prices for the Sports Channel and the Movie Channel are negatively correlated. Why would you, or why would you not, expect consumers' reservation prices for cable TV channels to be negatively correlated?

c. The company's vice president has said: "Because the marginal cost of providing an additional channel is zero, mixed bundling offers no advantage over pure bundling. Our profits would be just as high if we offered the Sports Channel and the Movie Channel together as a bundle, and only as a bundle." Do you agree or disagree? Explain why.

d. Suppose the cable company continues to use mixed bundling to sell these two services. Based on the distribution of reservation prices shown in Figure 11.21, do you think the cable company should alter any of the prices that it is now charging? If so, how?

***17.** Consider a firm with monopoly power that faces the demand curve

$$P = 100 - 3Q + 4A^{1/2}$$

and has the total cost function

$$C = 4Q^2 + 10Q + A$$

where A is the level of advertising expenditures, and P and Q are price and output.

a. Find the values of A, Q, and P that maximize the firm's profit.

b. Calculate the Lerner index, $L = (P - MC)/P$, for this firm at its profit-maximizing levels of A, Q, and P.

Appendix to Chapter 11

TRANSFER PRICING IN THE INTEGRATED FIRM

So far, we have studied the firm's pricing decision assuming that it sells its output in an *outside market*, i.e., to consumers or to other firms. Many firms, however, are *vertically integrated*—they contain several divisions, with some divisions producing parts and components that other divisions use to produce the finished product.[1] For example, automobile companies have "upstream" divisions that produce engines, brakes, radiators, and other components that the "downstream" divisions use to produce the finished cars. *Transfer pricing* refers to the valuation of these parts and components within the firm. **Transfer prices** are internal prices at which the parts and components from upstream divisions are "sold" to downstream divisions. Transfer prices must be chosen correctly because they are the signals that divisional managers use to determine output levels.

• **transfer prices** Internal prices at which parts and components from upstream divisions are "sold" to downstream divisions within a firm.

This appendix shows how a profit-maximizing firm chooses its transfer prices and divisional output levels. We will also examine other issues raised by vertical integration. For example, suppose a computer firm's upstream division produces memory chips used by a downstream division to produce the final product. If other firms also produce these chips, should our firm obtain all its chips from the upstream division, or should it also buy some on the outside market? Should the upstream division produce more chips than the downstream division needs and sell the excess in the market? How should the firm coordinate its upstream and downstream divisions? In particular, can we design incentives for the divisions that help the firm to maximize its profits?

We begin with the simplest case: There is no outside market for the output of the upstream division—i.e., the upstream division produces a good that is neither produced nor used by any other firm. Next we consider what happens when there is an outside market for the upstream division's output.

Transfer Pricing When There Is No Outside Market

Consider a firm with three divisions: Two upstream divisions produce inputs to a downstream processing division. The two upstream divisions produce quantities Q_1 and Q_2 and have total costs $C_1(Q_1)$ and $C_2(Q_2)$. The downstream division produces a quantity Q using the production function

$$Q = f(K, L, Q_1, Q_2)$$

where K and L are capital and labor inputs, and Q_1 and Q_2 are the intermediate inputs from the upstream divisions. Excluding the costs of the inputs Q_1 and Q_2, the downstream division has a total production cost $C_d(Q)$. Total revenue from sales of the final product is $R(Q)$.

We assume there are *no outside markets* for the intermediate inputs Q_1 and Q_2; they can be used only by the downstream division. Then the firm has two problems:

1. What quantities Q_1, Q_2, and Q will maximize its profit?
2. Is there an incentive scheme that will decentralize the firm's management? In particular, is there a set of transfer prices P_1 and P_2, so that *if each division*

[1] A firm is *horizontally integrated* when it has several divisions that produce the same or closely related products. Many firms are both vertically and horizontally integrated.

maximizes its own divisional profit, the profit of the overall firm will also be maximized?

To solve these problems, we note that the firm's total profit is

$$\pi(Q) = R(Q) - C_d(Q) - C_1(Q_1) - C_2(Q_2) \quad \textbf{(A11.1)}$$

What is the level of Q_1 that maximizes this profit? It is the level at which *the cost of the last unit of Q_1 is just equal to the additional revenue it brings to the firm.* The cost of producing one extra unit of Q_1 is the marginal cost $\Delta C_1/\Delta Q_1 = MC_1$. How much extra revenue results from that one extra unit? An extra unit of Q_1 allows the firm to produce more final output Q of an amount $\Delta Q/\Delta Q_1 = MP_1$, the marginal product of Q_1. An extra unit of final output results in additional revenue $\Delta R/\Delta Q = MR$, but it also results in additional cost to the downstream division of an amount $\Delta C_d/\Delta Q = MC_d$. Thus the *net marginal revenue* NMR_1 that the firm earns from an extra unit of Q_1 is $(MR - MC_d)MP_1$. Setting this equal to the marginal cost of the unit, we obtain the following rule for profit maximization[2]:

In §10.1, we explain that a firm maximizes its profit at the output at which marginal revenue is equal to marginal cost.

$$NMR_1 = (MR - MC_d)MP_1 = MC_1 \quad \textbf{(A11.2)}$$

Going through the same steps for the second intermediate input gives

$$NMR_2 = (MR - MC_d)MP_2 = MC_2 \quad \textbf{(A11.3)}$$

Note from equations (A11.2) and (A11.3) that it is *incorrect* to determine the firm's final output level Q by setting marginal revenue equal to marginal cost for the downstream division—i.e., by setting $MR = MC_d$. Doing so ignores the cost of producing the intermediate input. (MR exceeds MC_d because this cost is positive.) Also, note that equations (A11.2) and (A11.3) are standard conditions of marginal analysis: The output of each upstream division should be such that its marginal cost is equal to its marginal contribution to the profit of the overall firm.

Now, what transfer prices P_1 and P_2 should be "charged" to the downstream division for its use of the intermediate inputs? Remember that if each of the three divisions uses these transfer prices to maximize its own divisional profit, the profit of the overall firm should be maximized. The two upstream divisions will maximize their divisional profits, π_1 and π_2, which are given by

$$\pi_1 = P_1Q_1 - C_1(Q_1)$$

and

$$\pi_2 = P_2Q_2 - C_2(Q_2)$$

Because the upstream divisions take P_1 and P_2 as given, they will choose Q_1 and Q_2 so that $P_1 = MC_1$ and $P_2 = MC_2$. Similarly, the downstream division will maximize

$$\pi(Q) = R(Q) - C_d(Q) - P_1Q_1 - P_2Q_2$$

Because the downstream division also takes P_1 and P_2 as given, it will choose Q_1 and Q_2 so that

$$(MR - MC_d)MP_1 = NMR_1 = P_1 \quad \textbf{(A11.4)}$$

[2]Using calculus, we can obtain this rule by differentiating equation (A11.1) with respect to Q_1:

$$\begin{aligned} d\pi/dQ_1 &= (dR/dQ)(\partial Q/\partial Q_1) - (dC_d/dQ)(\partial Q/\partial Q_1) - dC_1/dQ_1 \\ &= (MR - MC_d)MP_1 - MC_1 \end{aligned}$$

Setting $d\pi/dQ = 0$ to maximize profit gives equation (A11.2).

and

$$(MR - MC_d)MP_2 = NMR_2 = P_2 \quad \textbf{(A11.5)}$$

Note that by setting the transfer prices equal to the respective marginal costs ($P_1 = MC_1$ and $P_2 = MC_2$), the profit-maximizing conditions given by equations (A11.2) and (A11.3) will be satisfied. We therefore have a simple solution to the transfer pricing problem: *Set each transfer price equal to the marginal cost of the respective upstream division.* Then when each division is required to maximize its own profit, the quantities Q_1 and Q_2 that the upstream divisions will want to produce will be the same quantities that the downstream division will want to "buy," and they will maximize the firm's total profit.

To illustrate this graphically, suppose Race Car Motors, Inc., has two divisions. The upstream Engine Division produces engines, and the downstream Assembly Division puts together automobiles, using one engine (and a few other parts) in each car. In Figure A11.1, the average revenue curve AR is Race Car Motors' demand curve for cars. (Note that the firm has monopoly power in the automobile market.) MC_A is the marginal cost of assembling automobiles, *given the engines* (i.e., it does not include the cost of the engines). Because the car requires one engine, the marginal product of the engines is one. Thus the curve labeled $MR - MC_A$ is also the net marginal revenue curve for engines:

$$NMR_E = (MR - MC_A)MP_E = MR - MC_A$$

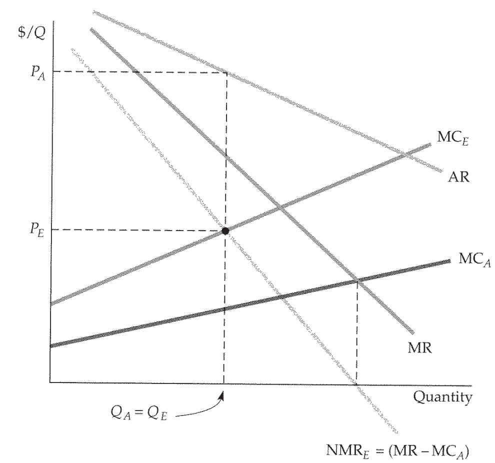

FIGURE A11.1 Race Car Motors, Inc.

The firm's upstream division should produce a quantity of engines Q_E that equates its marginal cost of engine production MC_E with the downstream division's net marginal revenue of engines NMR_E. Because the firm uses one engine in every car, NMR_E is the difference between the marginal revenue from selling cars and the marginal cost of assembling them, i.e., $MR - MC_A$. The optimal transfer price for engines P_E equals the marginal cost of producing them. Finished cars are sold at price P_A.

The profit-maximizing number of engines (and number of cars) is given by the intersection of the net marginal revenue curve NMR_E with the marginal cost curve for engines MC_E. Having determined the number of cars that it will produce, and knowing its divisional cost functions, the management of Race Car Motors can now set the transfer price P_E that correctly values the engines used to produce its cars. This is the transfer price that should be used to calculate divisional profit (and year-end bonuses for divisional managers).

Transfer Pricing with a Competitive Outside Market

Now suppose there is a *competitive* outside market for the intermediate good produced by an upstream division. Because the outside market is competitive, there is a single market price at which one can buy or sell the good. Therefore, *the marginal cost of the intermediate good is simply the market price.* Because the optimal transfer price must equal marginal cost, it must also equal the competitive market price.

To see this, suppose there is a competitive market for the engines that Race Car Motors produces. If the market price is low, Race Car Motors may want to buy some or all of its engines in the market; if it is high, it may want to sell engines in the market. Figure A11.2 illustrates the first case. For quantities below $Q_{E,1}$, the upstream division's marginal cost of producing engines MC_E is below the market price $P_{E,M}$; for quantities above $Q_{E,1}$, it is above the market price. The

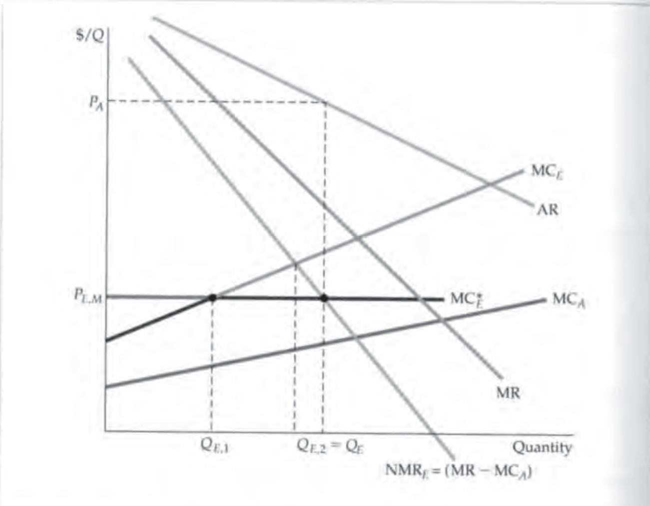

FIGURE A11.2 Buying Engines in a Competitive Outside Market

Race Car Motors' marginal cost of engines MC_E^* is the upstream division's marginal cost for quantities up to $Q_{E,1}$ and the market price $P_{E,M}$ for quantities above $Q_{E,1}$. The downstream division should use a total of $Q_{E,2}$ engines to produce an equal number of cars; in that case, the marginal cost of engines equals net marginal revenue. $Q_{E,2} - Q_{E,1}$ of these engines are bought in the outside market. The downstream division "pays" the upstream division the transfer price $P_{E,M}$ for the remaining $Q_{E,1}$ engines.

firm should obtain engines at the least cost, so the marginal cost of engines MC^*_E will be the upstream division's marginal cost for quantities up to $Q_{E,1}$ and the market price for quantities above $Q_{E,1}$. Note that Race Car Motors uses more engines and produces more cars than it would have had there been no outside engine market. The downstream division now buys $Q_{E,2}$ engines and produces an equal number of automobiles. However, it "buys" only $Q_{E,1}$ of these engines from the upstream division and the rest on the open market.

It might seem strange that Race Car Motors must go into the open market to buy engines that it can make itself. If it made all of its own engines, however, its marginal cost of producing them would exceed the competitive market price. Although the profit of the upstream division would be higher, *the total profit of the firm would be lower*.

Figure A11.3 shows the case where Race Car Motors *sells* engines in the outside market. Now the competitive market price $P_{E,M}$ is above the transfer price that the firm would have set had there been no outside market. In this case, although the upstream Engine Division produces $Q_{E,1}$ engines, only $Q_{E,2}$ engines are used by the downstream division to produce automobiles. The rest are sold in the outside market at the price $P_{E,M}$.

Note that compared with a situation in which there is no outside engine market, Race Car Motors is producing more engines but fewer cars. Why not produce this larger number of engines but use all of them to produce more cars?

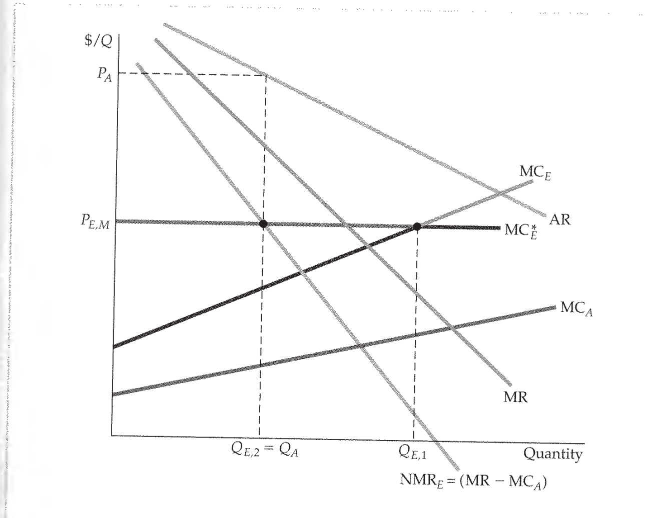

FIGURE A11.3 Selling Engines in a Competitive Outside Market

The optimal transfer price for Race Car Motors is again the market price $P_{E,M}$. This price is above the point at which MC_E intersects NMR_E, so the upstream division sells some of its engines in the outside market. The upstream division produces $Q_{E,1}$ engines, the quantity at which MC_E equals $P_{E,M}$. The downstream division uses only $Q_{E,2}$ of these engines, the quantity at which NMR_E equals $P_{E,M}$. Compared with Figure A11.1, in which there is no outside market, more engines but fewer cars are produced.

Because the engines are too valuable. On the margin, the net revenue that can be earned from selling them in the outside market is higher than the net revenue from using them to build additional cars.

Transfer Pricing with a Noncompetitive Outside Market

Now suppose there is an outside market for the output of the upstream division, but that market is not competitive—the firm has monopoly power. The same principles apply, but we must be careful when measuring net marginal revenue.

Suppose the engine produced by the upstream Engine Division is a special one that only Race Car Motors can make. There is, however, an outside market for this engine. Race Car Motors, therefore, can be a monopoly supplier to that market while also producing engines for its own use. What is the optimal transfer price for use of the engines by the downstream division, and at what price (if any) should engines be sold in the outside market?

We must find the firm's net marginal revenue from the sale of engines. In Figure A11.4, $D_{E,M}$ is the outside market demand curve for engines and $MR_{E,M}$

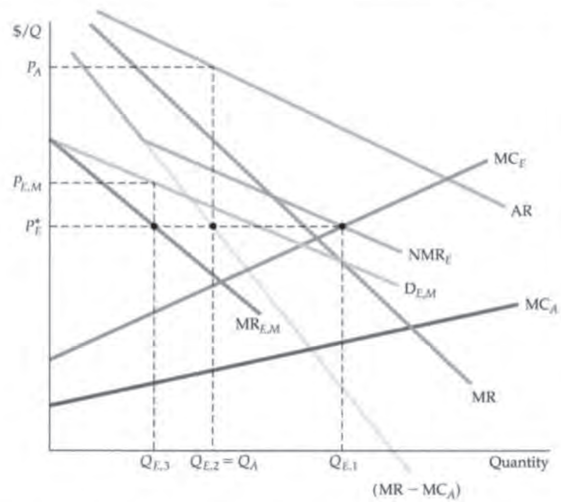

FIGURE A11.4 Race Car Motors—A Monopoly Supplier of Engines to an Outside Market

$D_{E,M}$ is the outside market demand curve for engines; $MR_{E,M}$ is the corresponding marginal revenue curve; $(MR - MC_A)$ is the net marginal revenue from the use of engines by the downstream division. The *total net marginal revenue curve for engines* NMR_E is the horizontal sum of these two marginal revenues. The optimal transfer price P^*_E and the quantity of engines that the upstream division produces, $Q_{E,1}$, are found where $MC_E = NMR_E$. $Q_{E,2}$ of these engines are used by the downstream division, the quantity at which the downstream division's net marginal revenue, $MR - MC_A$, is equal to the transfer price P^*_E. The remaining engines, $Q_{E,3}$, are sold in the outside market at the price $P_{E,M}$.

is the corresponding marginal revenue curve. Race Car Motors thus has two sources of marginal revenue from the production and sale of an additional engine: marginal revenue $MR_{E,M}$ from sales in the outside market and net marginal revenue $(MR - MC_A)$ from the use of the engines by the downstream division. By summing these two curves horizontally, we obtain the *total net marginal revenue curve for engines*; it is the green line labeled NMR_E.

The intersection of the marginal cost and total net marginal revenue curves gives the quantity of engines $Q_{E,1}$ that the upstream division should produce and the optimal transfer price P^*_E. Again, the optimal transfer price is equal to marginal cost. But note that only $Q_{E,2}$ of these engines are used by the downstream division to make cars. This is the quantity at which the downstream division's net marginal revenue, $MR - MC_A$, is equal to the transfer price P^*_E. The remaining engines $Q_{E,3}$ are sold in the outside market. However, they are not sold at the transfer price P^*_E. Instead the firm exercises its monopoly power and sells them at the higher price $P_{E,M}$.

Why pay the upstream division only P^*_E per engine when the firm is selling engines in the outside market at the higher price $P_{E,M}$? Because if the upstream division is paid more than P^*_E (and thereby encouraged to produce more engines), the marginal cost of engines will rise and exceed the net marginal revenue from their use by the downstream division. And if the price charged in the outside market were lowered, the marginal revenue from sales in that market would fall below marginal cost. At the prices P^*_E and $P_{E,M}$, marginal revenues and marginal cost are equal:

$$MR_{E,M} = (MR - MC_A) = MC_E$$

Sometimes a vertically integrated firm can buy components in an outside market in which it has *monopsony* power. Suppose, for example, that Race Car Motors can obtain engines from its upstream Engine Division, or can purchase them *as a monopsonist* in an outside market. Although we have not illustrated this case graphically, you should be able to see that in this case, the transfer price paid to the Engine Division will be *above* the price at which engines are bought in the outside market. Why "pay" the upstream division a price that is higher than that paid in the outside market? With monopsony power, purchasing one additional engine in the outside market incurs a *marginal expenditure* that is greater than the actual price per engine paid in that market. The marginal expenditure is higher because purchasing an additional unit raises the average expenditure paid *for all* units bought in the outside market.

In §10.5, we explain that when a buyer has monopsony power, its marginal expenditure curve lies above its average expenditure curve because the decision to buy an extra unit of the good raises the price that must be paid on all units.

A Numerical Example

Suppose Race Car Motors has the following demand for its automobiles:

$$P = 20{,}000 - Q$$

Its marginal revenue is thus

$$MR = 20{,}000 - 2Q$$

The downstream division's cost of assembling cars is

$$C_A(Q) = 8000Q$$

so that the division's marginal cost is $MC_A = 8000$. The upstream division's cost of producing engines is

$$C_E(Q_E) = 2Q_E^2$$

The division's marginal cost is thus $MC_E(Q_E) = 4Q_E$.

First, suppose there is *no outside market* for the engines. How many engines and cars should the firm produce? What should be the transfer price for engines? To solve this problem, we set the net marginal revenue for engines equal to the marginal cost of producing engines. Because each car has one engine, $Q_E = Q$. The net marginal revenue of engines is thus

$$NMR_E = MR - MC_A = 12{,}000 - 2Q_E$$

Now set NMR_E equal to MC_E:

$$12{,}000 - 2Q_E = 4Q_E$$

Thus $6Q_E = 12{,}000$ and $Q_E = 2000$. The firm should therefore produce 2000 engines and 2000 cars. The optimal transfer price is the marginal cost of these 2000 engines:

$$P_E = 4Q_E = \$8000$$

Second, suppose that engines can be bought or sold for \$6000 in an *outside competitive market*. This is below the \$8000 transfer price that is optimal when there is no outside market, so the firm should buy some engines outside. Its marginal cost of engines, and the optimal transfer price, is now \$6000. Set this \$6000 marginal cost equal to the net marginal revenue of engines:

$$6000 = NMR_E = 12{,}000 - 2Q_E$$

Thus the total quantity of engines and cars is now 3000. The company now produces more cars (and sells them at a lower price) because its cost of engines is lower. Also, since the transfer price for the engines is now \$6000, the upstream Engine Division supplies only 1500 engines (because $MC_E(1500) = \$6000$). The remaining 1500 engines are bought in the outside market.

Finally, suppose Race Car Motors is the only producer of these engines but can sell them in an outside market. Demand in the outside market is

$$P_{E,M} = 10{,}000 - Q_E$$

The marginal revenue from sales in the market is therefore

$$MR_{E,M} = 10{,}000 - 2Q_E$$

To determine the optimal transfer price, we find the *total* net marginal revenue by horizontally summing $MR_{E,M}$ with the net marginal revenue from "sales" to the downstream division, $12{,}000 - 2Q_E$, as in Figure A11.4. For outputs Q_E greater than 1000, this is

$$NMR_{E,Total} = 11{,}000 - Q_E$$

Now set this equal to the marginal cost of producing engines:

$$11{,}000 - Q_E = 4Q_E$$

The total quantity of engines produced should therefore be $Q_E = 2200$.

How many of these engines should go to the downstream division and how many to the outside market? Note that the marginal cost of producing these 2200 engines—and therefore the optimal transfer price—is $4Q_E = \$8800$. Set this price equal to the marginal revenue from sales in the outside market:

$$8800 = 10{,}000 - 2Q_E$$

or $Q_E = 600$. Therefore, 600 engines should be sold in the outside market. Finally, set this \$8800 transfer price equal to the net marginal revenue from "sales" to the downstream division:

$$8800 = 12{,}000 - 2Q_E$$

or $Q_E = 1600$. Thus 1600 engines should be supplied to the downstream division for use in the production of 1600 cars.

EXERCISES

1. Review the numerical example about Race Car Motors. Calculate the profit earned by the upstream division, the downstream division, and the firm as a whole in each of the three cases examined: (a) there is no outside market for engines; (b) there is a competitive market for engines in which the market price is \$6000; and (c) the firm is a monopoly supplier of engines to an outside market. In which case does Race Car Motors earn the most profit? In which case does the upstream division earn the most? The downstream division?
2. Ajax Computer makes a computer for climate control in office buildings. The company uses a microprocessor produced by its upstream division, along with other parts bought in outside competitive markets. The microprocessor is produced at a constant marginal cost of \$500, and the marginal cost of assembling the computer (including the cost of the other parts) by the downstream division is a constant \$700. The firm has been selling the computer for \$2000, and until now there has been no outside market for the microprocessor.
   a. Suppose an outside market for the microprocessor develops and that Ajax has monopoly power in that market, selling microprocessors for \$1000 each. Assuming that demand for the microprocessor is unrelated to the demand for the Ajax computer, what transfer price should Ajax apply to the microprocessor for its use by the downstream computer division? Should production of computers be increased, decreased, or left unchanged? Explain briefly.
   b. How would your answer to (a) change if the demands for the computer and the microprocessors were competitive; i.e., if some of the people who buy the microprocessors use them to make climate control systems of their own?
3. Reebok produces and sells running shoes. It faces a market demand schedule $P = 11 - 1.5Q_s$, where Q_s is the number of pairs of shoes sold and P is the price in dollars per pair of shoes. Production of each pair of shoes requires 1 square yard of leather. The leather is shaped and cut by the Form Division of Reebok. The cost function for leather is

$$TC_L = 1 + Q_L + 0.5Q_L^2$$

where Q_L is the quantity of leather (in square yards) produced. Excluding leather, the cost function for running shoes is

$$TC_s = 2Q_s$$

   a. What is the optimal transfer price?
   b. Leather can be bought and sold in a competitive market at the price of $P_F = 1.5$. In this case, how much leather should the Form Division supply internally? How much should it supply to the outside market? Will Reebok buy any leather in the outside market? Find the optimal transfer price.
   c. Now suppose the leather is unique and of extremely high quality. Therefore, the Form Division may act as a monopoly supplier to the outside market as well as a supplier to the downstream division. Suppose the outside demand for leather is given by $P = 32 - Q_L$. What is the optimal transfer price for the use of leather by the downstream division? At what price, if any, should leather be sold to the outside market? What quantity, if any, will be sold to the outside market?
4. The House Products Division of Acme Corporation manufactures and sells digital clock radios. A major component is supplied by the electronics division of Acme. The cost functions for the radio and the electronic component divisions are, respectively,

$$TC_r = 30 + 2Q_r$$
$$TC_c = 70 + 6Q_c + Q_c^2$$

Note that TC_r does not include the cost of the component. Manufacture of one radio set requires the use of one electronic component. Market studies show that the firm's demand curve for the digital clock radio is given by

$$P_r = 108 - Q_r$$

   a. If there is no outside market for the components, how many of them should be produced to maximize profits for Acme as a whole? What is the optimal transfer price?
   b. If other firms are willing to purchase in the outside market the component manufactured by the electronics division (which is the only supplier of this product), what is the optimal transfer price? Why? What price should be charged in the outside market? Why? How many units will the electronics division supply internally and to the outside market? Why? (*Note:* The demand for components in the outside market is $P_c = 72 - 1.5Q_c$.)