CHAPTER FOURTEEN

# Monopoly

A *monopoly* is a single firm that serves an entire market. This single firm faces the market demand curve for its output. Using its knowledge of this demand curve, the monopoly makes a decision on how much to produce. Unlike the perfectly competitive firm's output decision (which has no effect on market price), the monopoly's output decision will, in fact, determine the good's price. In this sense, monopoly markets and markets characterized by perfect competition are polar-opposite cases.

DEFINITION

**Monopoly.** A *monopoly* is a single supplier to a market. This firm may choose to produce at any point on the market demand curve.

At times it is more convenient to treat monopolies as having the power to set prices. Technically, a monopoly can choose that point on the market demand curve at which it prefers to operate. It may choose either market price or quantity, but not both. In this chapter we will usually assume that monopolies choose the quantity of output that maximizes profits and then settle for the market price that the chosen output level yields. It would be a simple matter to rephrase the discussion in terms of price setting, and in some places we shall do so.

## BARRIERS TO ENTRY

The reason a monopoly exists is that other firms find it unprofitable or impossible to enter the market. Therefore, *barriers to entry* are the source of all monopoly power. If other firms could enter a market, then the firm would, by definition, no longer be a monopoly. There are two general types of barriers to entry: technical barriers and legal barriers.

### Technical barriers to entry

A primary technical barrier is that the production of the good in question may exhibit decreasing marginal (and average) costs over a wide range of output levels. The technology of production is such that relatively large-scale firms are low-cost producers. In this situation (which is sometimes referred to as *natural monopoly*), one firm may find it profitable to drive others out of the industry by cutting prices. Similarly, once a monopoly has been established, entry will be difficult because any new firm must produce at relatively low levels of output and therefore at relatively high average costs. It is important to stress that the range of declining costs need only be "large" relative to the market in question. Declining costs on some absolute scale are not necessary. For example, the production and delivery of concrete does not exhibit declining marginal costs over a broad range of output when compared with the total U.S. market. However, in any particular small town, declining marginal costs may permit a monopoly to be established. The high costs of transportation in this industry tend to isolate one market from another.

Another technical basis of monopoly is special knowledge of a low-cost productive technique. The monopoly has an incentive to keep its technology secret; but unless this technology is protected by a patent (see next paragraph), this may be extremely difficult. Ownership of unique resources—such as mineral deposits or land locations, or the possession of unique managerial talents—may also be a lasting basis for maintaining a monopoly.

## Legal barriers to entry

Many pure monopolies are created as a matter of law rather than as a matter of economic conditions. One important example of a government-granted monopoly position is the legal protection of a product by a patent or copyright. Prescription drugs, computer chips, and Disney animated movies are examples of profitable products that are shielded (for a time) from direct competition by potential imitators. Because the basic technology for these products is uniquely assigned to one firm, a monopoly position is established. The defense made of such a governmentally granted monopoly is that the patent and copyright system makes innovation more profitable and therefore acts as an incentive. Whether the benefits of such innovative behavior exceed the costs of having monopolies is an open question that has been much debated by economists.

A second example of a legally created monopoly is the awarding of an exclusive franchise to serve a market. These franchises are awarded in cases of public utility (gas and electric) service, communications services, the post office, some television and radio station markets, and a variety of other situations. Usually the restriction of entry is combined with a regulatory cap on the price the franchised monopolist is allowed to charge. The argument usually put forward in favor of creating these franchised monopolies is that the industry in question is a natural monopoly: average cost is diminishing over a broad range of output levels, and minimum average cost can be achieved only by organizing the industry as a monopoly. The public utility and communications industries are often considered good examples. Certainly, that does appear to be the case for local electricity and telephone service where a given network probably exhibits declining average cost up to the point of universal coverage. But recent deregulation in telephone services and electricity generation show that, even for these industries, the natural monopoly rationale may not be all-inclusive. In other cases, franchises may be based largely on political rationales. This seems to be true for the postal service in the United States and for a number of nationalized industries (airlines, radio and television, banking) in other countries.

## Creation of barriers to entry

Although some barriers to entry may be independent of the monopolist's own activities, other barriers may result directly from those activities. For example, firms may develop unique products or technologies and take extraordinary steps to keep these from being copied by competitors. Or firms may buy up unique resources to prevent potential entry. The De Beers cartel, for example, controls a large fraction of the world's diamond mines. Finally, a would-be monopolist may enlist government aid in devising barriers to entry. It may lobby for legislation that restricts new entrants to "maintain an orderly market" or for health and safety regulations that raise potential entrants' costs. Because the monopolist has both special knowledge of its business and significant incentives to pursue these goals, it may have considerable success in creating such barriers to entry.

The attempt by a monopolist to erect barriers to entry may involve real resource costs. Maintaining secrecy, buying unique resources, and engaging in political lobbying are all costly activities. A full analysis of monopoly should involve not only questions of cost minimization and output choice (as under perfect competition) but also an analysis of profit-maximizing creation of entry barriers. However, we will not provide a detailed investigation of such questions here.[1] Instead, we will take the presence of a single supplier on the market, and this single firm's cost function, as given.

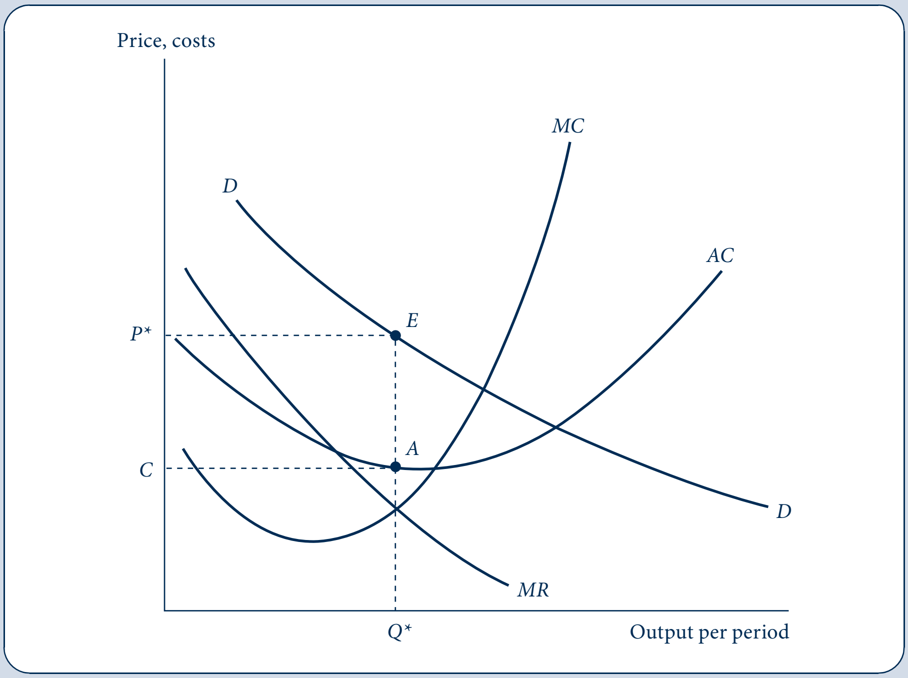
# PROFIT MAXIMIZATION AND OUTPUT CHOICE

To maximize profits, a monopoly will choose to produce that output level for which marginal revenue is equal to marginal cost. Because the monopoly, in contrast to a perfectly competitive firm, faces a negatively sloped market demand curve, marginal revenue will be less than the market price. To sell an additional unit, the monopoly must lower its price on all units to be sold if it is to generate the extra demand necessary to absorb this marginal unit. The profit-maximizing output level for a firm is then the level $Q^*$ in Figure 14.1. At that level, marginal revenue is equal to marginal costs, and profits are maximized.

**FIGURE 14.1**

**Profit Maximization and Price Determination for a Monopoly**

A profit-maximizing monopolist produces that quantity for which marginal revenue is equal to marginal cost. In the diagram this quantity is given by $Q^*$, which will yield a price of $P^*$ in the market. Monopoly profits can be read as the rectangle of $P^*EAC$.




[1] For a simple treatment, see R. A. Posner, "The Social Costs of Monopoly and Regulation," *Journal of Political Economy* 83 (August 1975): 807–27.

Given the monopoly's decision to produce $Q^*$, the demand curve $D$ indicates that a market price of $P^*$ will prevail. This is the price that demanders as a group are willing to pay for the output of the monopoly. In the market, an equilibrium price–quantity combination of $P^*$, $Q^*$ will be observed. Assuming $P^* > AC$, this output level will be profitable, and the monopolist will have no incentive to alter output levels unless demand or cost conditions change. Hence we have the following principle.

**OPTIMIZATION PRINCIPLE**

**Monopolist's output.** A monopolist will choose to produce that output for which marginal revenue equals marginal cost. Because the monopolist faces a downward-sloping demand curve, market price will exceed marginal revenue and the firm's marginal cost at this output level.

## The inverse elasticity rule, again

In Chapter 11 we showed that the assumption of profit maximization implies that the gap between a price of a firm's output and its marginal cost is inversely related to the price elasticity of the demand curve faced by the firm. Applying Equation 11.14 to the case of monopoly yields

$$\frac{P - MC}{P} = -\frac{1}{e_{Q,P}}, \qquad \textbf{(14.1)}$$

where now we use the elasticity of demand for the entire market ($e_{Q,P}$) because the monopoly is the sole supplier of the good in question. This observation leads to two general conclusions about monopoly pricing. First, a monopoly will choose to operate only in regions in which the *market* demand curve is elastic ($e_{Q,P} < -1$). If demand were inelastic, then marginal revenue would be negative and thus could not be equated to marginal cost (which presumably is positive). Equation 14.1 also shows that $e_{Q,P} > -1$ implies an (implausible) negative marginal cost.

A second implication of Equation 14.1 is that the firm's "markup" over marginal cost (measured as a fraction of price) depends inversely on the elasticity of market demand. For example, if $e_{Q,P} = -2$, then Equation 14.1 shows that $P = 2MC$, whereas if $e_{Q,P} = -10$, then $P = 1.11MC$. Notice also that if the elasticity of demand were constant along the entire demand curve, the proportional markup over marginal cost would remain unchanged in response to changes in input costs. Therefore, market price moves proportionally to marginal cost: Increases in marginal cost will prompt the monopoly to increase its price proportionally, and decreases in marginal cost will cause the monopoly to reduce its price proportionally. Even if elasticity is not constant along the demand curve, it seems clear from Figure 14.1 that increases in marginal cost will increase price (although not necessarily in the same proportion). As long as the demand curve facing the monopoly is downward sloping, upward shifts in $MC$ will prompt the monopoly to reduce output and thereby obtain a higher price.[2] We will examine all these relationships mathematically in Examples 14.1 and 14.2.

## Monopoly profits

Total profits earned by the monopolist can be read directly from Figure 14.1. These are shown by the rectangle $P^*EAC$ and again represent the profit per unit (price minus average cost) times the number of units sold. These profits will be positive if market price exceeds average total cost. If $P^* < AC$, however, then the monopolist can operate only at a long-term loss and will decline to serve the market.

[2]The comparative statics of a shift in the demand curve facing the monopolist are not so clear, however, and no unequivocal prediction about price can be made. For an analysis of this issue, see the discussion that follows and Problem 14.4.

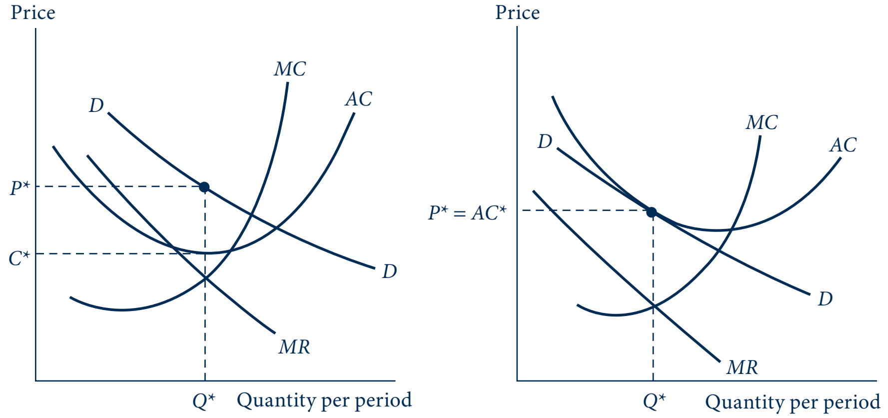
Because (by assumption) no entry is possible into a monopoly market, the monopolist's positive profits can exist even in the long run. For this reason, some authors refer to the profits that a monopoly earns in the long run as *monopoly rents.* These profits can be regarded as a return to that factor that forms the basis of the monopoly (e.g., a patent, a favorable location, or a dynamic entrepreneur); hence another possible owner might be willing to pay that amount in rent for the right to the monopoly. The potential for profits is the reason why some firms pay other firms for the right to use a patent and why concessioners at sporting events (and on some highways) are willing to pay for the right to the concession. To the extent that monopoly rights are given away at less than their true market value (as in radio and television licensing), the wealth of the recipients of those rights is increased.

Although a monopoly may earn positive profits in the long run,[3] the size of such profits will depend on the relationship between the monopolist's average costs and the demand for its product. Figure 14.2 illustrates two situations in which the demand, marginal revenue, and marginal cost curves are rather similar. As Equation 14.1 suggests, the price-marginal cost markup is about the same in these two cases. But average costs in Figure 14.2a are considerably lower than in Figure 14.2b. Although the profit-maximizing decisions are similar in the two cases, the level of profits ends up being different. In Figure 14.2a the monopolist's price ($P^*$) exceeds the average cost of producing $Q^*$ (labeled $AC^*$) by a large extent, and significant profits are obtained. In Figure 14.2b, however, $P^* = AC^*$ and the monopoly earns zero economic profits, the largest amount possible in this case. Hence large profits from a monopoly are not inevitable, and the actual extent of economic profits may not always be a good guide to the significance of monopolistic influences in a market.

**FIGURE 14.2**

**Monopoly Profits Depend on the Relationship between the Demand and Average Cost Curves**

Both monopolies in this figure are equally "strong" if by this we mean they produce similar divergences between market price and marginal cost. However, because of the location of the demand and average cost curves, it turns out that the monopoly in (a) earns high profits, whereas that in (b) earns no profits. Consequently, the size of profits is not a measure of the strength of a monopoly.




(a) Monopoly with large profits

(b) Zero-profit monopoly

[3]As in the competitive case, the profit-maximizing monopolist would be willing to produce at a loss in the short run as long as market price exceeds average variable cost.

## There is no monopoly supply curve

In the theory of perfectly competitive markets presented in Part 4, it was possible to speak of an industry supply curve. We constructed the long-run supply curve by allowing the market demand curve to shift and observing the supply curve that was traced out by the series of equilibrium price–quantity combinations. This type of construction is not possible for monopolistic markets. With a fixed market demand curve, the supply "curve" for a monopoly will be only one point—namely, that price–quantity combination for which $MR = MC$. If the demand curve should shift, then the marginal revenue curve would also shift, and a new profit-maximizing output would be chosen. However, connecting the resulting series of equilibrium points on the market demand curves would have little meaning. This locus might have a strange shape, depending on how the market demand curve's elasticity (and its associated $MR$ curve) changes as the curve is shifted. In this sense the monopoly firm has no well-defined "supply curve." Each demand curve is a unique profit-maximizing opportunity for a monopolist.

### EXAMPLE 14.1 Calculating Monopoly Output

Suppose the market for Olympic-quality Frisbees ($Q$, measured in Frisbees bought per year) has a linear demand curve of the form

$$Q = 2{,}000 - 20P \tag{14.2}$$

or

$$P = 100 - \frac{Q}{20}, \tag{14.3}$$

and let the costs of a monopoly Frisbee producer be given by

$$C(Q) = 0.05Q^2 + 10{,}000. \tag{14.4}$$

To maximize profits, this producer chooses that output level for which $MR = MC$. To solve this problem we must phrase both $MR$ and $MC$ as functions of $Q$ alone. Toward this end, write total revenue as

$$P \cdot Q = 100Q - \frac{Q^2}{20}. \tag{14.5}$$

Consequently,

$$MR = 100 - \frac{Q}{10} = MC = 0.1Q \tag{14.6}$$

and

$$Q^* = 500, \quad P^* = 75. \tag{14.7}$$

At the monopoly's preferred output level,

$$\begin{aligned} C(Q) &= 0.05(500)^2 + 10{,}000 = 22{,}500, \\ AC &= \frac{22{,}500}{500} = 45. \end{aligned} \tag{14.8}$$

Using this information, we can calculate profits as

$$\pi = (P^* - AC) \cdot Q^* = (75 - 45) \cdot 500 = 15{,}000. \tag{14.9}$$

Observe that at this equilibrium there is a large markup between price (75) and marginal cost ($MC = 0.1Q = 50$). Yet as long as entry barriers prevent a new firm from producing Olympic-quality Frisbees, this gap and positive economic profits can persist indefinitely.

**QUERY:** How would an increase in fixed costs from 10,000 to 12,500 affect the monopoly's output plans? How would profits be affected? Suppose total costs shifted to $C(Q) = 0.075Q^2 + 10{,}000$. How would the equilibrium change?

## EXAMPLE 14.2 Monopoly with Simple Demand Curves

We can derive a few simple facts about monopoly pricing in cases where the demand curve facing the monopoly takes a simple algebraic form and the firm has constant marginal costs (i.e., $C(Q) = cQ$ and $MC = c$).

**Linear demand.** Suppose that the inverse demand function facing the monopoly is of the linear form $P = a - bQ$. In this case, $PQ = aQ - bQ^2$ and $MR = dPQ/dQ = a - 2bQ$. Hence profit maximization requires that

$$MR = a - 2bQ = MC = c \quad \text{or} \quad Q = \frac{a - c}{2b}. \tag{14.10}$$

Inserting this solution for the profit-maximizing output level back into the inverse demand function yields a direct relationship between price and marginal cost:

$$P = a - bQ = a - \frac{a - c}{2} = \frac{a + c}{2}. \tag{14.11}$$

An interesting implication is that, in this linear case, $dP/dc = 1/2$. That is, only half of the amount of any increase in marginal cost will show up in the market price of the monopoly product.[4]

**Constant elasticity demand.** If the demand curve facing the monopoly takes the constant elasticity form $Q = aP^e$ (where $e$ is the price elasticity of demand), then we know $MR = P(1 + 1/e)$, and thus profit maximization requires

$$P\left(1 + \frac{1}{e}\right) = c \quad \text{or} \quad P = c\left(\frac{e}{1 + e}\right). \tag{14.12}$$

Because it must be the case that $e < -1$ for profit maximization, price will clearly exceed marginal cost, and this gap will be larger the closer $e$ is to $-1$. Notice also that $dP/dc = e/(1 + e)$ and so any given increase in marginal cost will increase price by more than this amount. Of course, as we have already pointed out, the proportional increase in marginal cost and price will be the same. That is, $e_{P,c} = dP/dc \cdot c/P = 1$.

**QUERY:** The demand function in both cases is shifted by the parameter $a$. Discuss the effects of such a shift for both linear and constant elasticity demand. Explain your results intuitively.

# MONOPOLY AND RESOURCE ALLOCATION

In Chapter 13 we briefly mentioned why the presence of monopoly distorts the allocation of resources. Because the monopoly produces a level of output for which $MC = MR < P$, the market price of its good no longer conveys accurate information about production costs. Hence consumers' decisions will no longer reflect true opportunity costs of production, and resources will be misallocated. In this section we explore this misallocation in some detail in a partial equilibrium context.

## Basis of comparison

To evaluate the allocational effect of a monopoly, we need a precisely defined basis of comparison. A particularly useful comparison is provided by a perfectly competitive industry. It is convenient to think of a monopoly as arising from the "capture" of such a competitive industry and to treat the individual firms that constituted the competitive industry as now

[4]Notice that when $c = 0$, we have $P = a/2$. That is, price should be halfway between zero and the price intercept of the demand curve.

being single plants in the monopolist's empire. A prototype case would be John D. Rockefeller's purchase of most of the U.S. petroleum refineries in the late nineteenth century and his decision to operate them as part of the Standard Oil trust. We can then compare the performance of this monopoly with the performance of the previously competitive industry to arrive at a statement about the welfare consequences of monopoly.

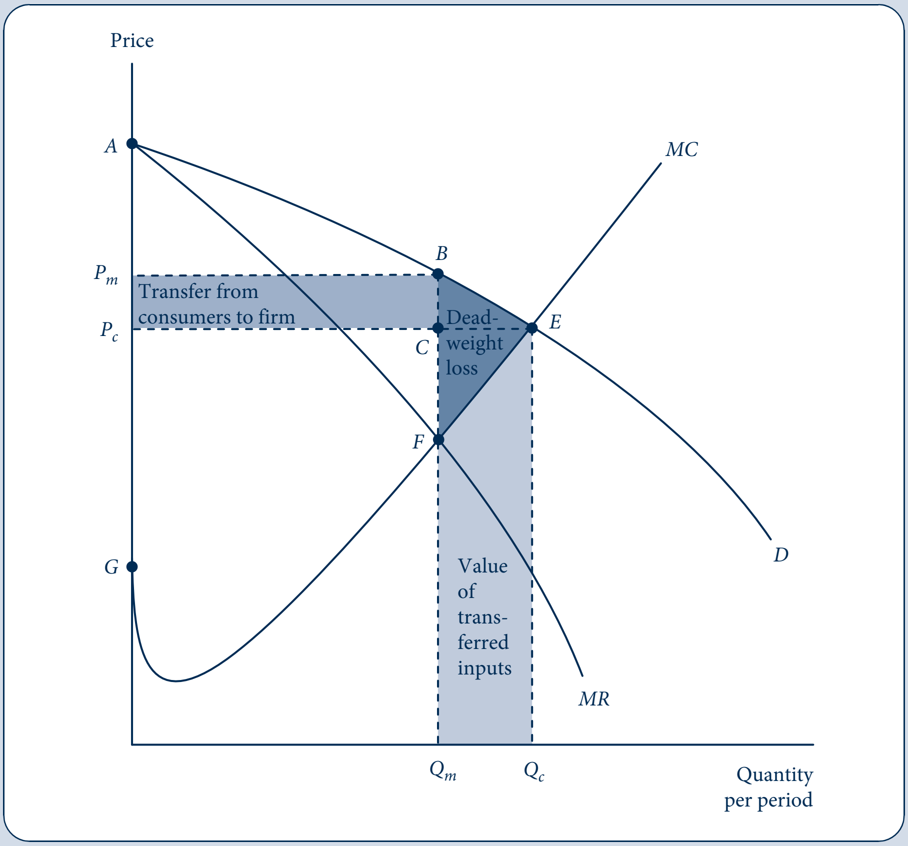
## A graphical analysis

Figure 14.3 provides a graphical analysis of the welfare effects of monopoly. If this market were competitive, output would be $Q_c$—that is, production would occur where price is equal to long-run average and marginal cost. Under a simple single-price monopoly, output would be $Q_m$ because this is the level of production for which marginal revenue is equal to marginal cost. The restriction in output from $Q_c$ to $Q_m$ represents the misallocation brought about through monopolization. The total value of resources released by this output restriction is shown in Figure 14.3 as area $FEQ_cQ_m$. Essentially, the monopoly closes down some of the plants that were operating in the competitive case. These transferred inputs can be productively used elsewhere; thus, area $FEQ_cQ_m$ is not a social loss.

**FIGURE 14.3**

**Allocational and Distributional Effects of Monopoly**

Monopolization of this previously competitive market would cause output to be reduced from $Q_c$ to $Q_m$. Productive inputs worth $FEQ_cQ_m$ are reallocated to the production of other goods. Consumer surplus equal to $P_mBCP_c$ is transferred into monopoly profits. Deadweight loss is given by $BEF$.



The restriction in output from $Q_c$ to $Q_m$ involves a total loss in consumer surplus of $P_mBEP_c$. Part of this loss, $P_mBCP_c$, is transferred to the monopoly as increased profit. Another part of the consumers' loss, *BEC*, is not transferred to anyone but is a pure deadweight loss in the market. A second source of deadweight loss is given by area *CEF*. This is an area of lost producer surplus that does not get transferred to another source.[5] The total deadweight loss from both sources is area *BEF*, sometimes called the *deadweight loss triangle* because of its roughly triangular shape. The gain $P_mBCP_c$ in monopoly profit from an increased price more than compensates for its loss of producer surplus *CEF* from the output reduction so that, overall, the monopolist finds reducing output from $Q_c$ to $Q_m$ to be profitable.

To illustrate the nature of this deadweight loss, consider Example 14.1, in which we calculated an equilibrium price of \$75 and a marginal cost of \$50. This gap between price and marginal cost is an indication of the efficiency-improving trades that are forgone through monopolization. Undoubtedly, there is a would-be buyer who is willing to pay, say, \$60 for an Olympic Frisbee but not \$75. A price of \$60 would more than cover all the resource costs involved in Frisbee production, but the presence of the monopoly prevents such a mutually beneficial transaction between Frisbee users and the providers of Frisbee-making resources. For this reason, the monopoly equilibrium is not Pareto optimal—an alternative allocation of resources would make all parties better off. Economists have made many attempts to estimate the overall cost of these deadweight losses in actual monopoly situations. Most of these estimates are rather small when viewed in the context of the whole economy.[6] Allocational losses are larger, however, for some narrowly defined industries.

## EXAMPLE 14.3 Welfare Losses and Elasticity

The allocational effects of monopoly can be characterized fairly completely in the case of constant marginal costs and a constant price elasticity demand curve. To do so, assume again that constant marginal (and average) costs for a monopolist are given by $c$ and that the demand curve has a constant elasticity form of

$$Q = P^e, \tag{14.13}$$

where $e$ is the price elasticity of demand ($e < -1$). We know the competitive price in this market will be

$$P_c = c \tag{14.14}$$

and the monopoly price is given by

$$P_m = \frac{c}{1 + 1/e}. \tag{14.15}$$

The consumer surplus associated with any price ($P_0$) can be computed as

$$\begin{aligned} CS &= \int_{P_0}^{\infty} Q(P)dP \\ &= \int_{P_0}^{\infty} P^e dP \\ &= \frac{P^{e+1}}{e+1}\bigg|_{P_0}^{\infty} \\ &= -\frac{P_0^{e+1}}{e+1}. \end{aligned} \tag{14.16}$$

[5]More precisely, region *CEF* represents lost producer surplus (equivalently, lost profit) if output were reduced holding prices constant at $P_c$. To understand how to measure producer surplus on a graph, review the section on producer surplus in Chapter 11, especially Figure 11.4.

[6]The classic study is A. Harberger, "Monopoly and Resource Allocation," *American Economic Review* (May 1954): 77–87. Harberger estimates that such losses constitute about 0.1 percent of gross national product.

Hence under perfect competition we have

$$CS_c = -\frac{c^{e+1}}{e+1} \tag{14.17}$$

and, under monopoly,

$$CS_m = -\frac{\left(\dfrac{c}{1+1/e}\right)^{e+1}}{e+1}. \tag{14.18}$$

Taking the ratio of these two surplus measures yields

$$\frac{CS_m}{CS_c} = \left(\frac{1}{1+1/e}\right)^{e+1}. \tag{14.19}$$

If $e = -2$, for example, then this ratio is 1/2: consumer surplus under monopoly is half what it is under perfect competition. For more elastic cases this figure decreases a bit (because output restrictions under monopoly are more significant). For elasticities closer to $-1$, the ratio increases.

**Profits.** The transfer from consumer surplus into monopoly profits can also be computed fairly easily in this case. Monopoly profits are given by

$$\begin{aligned} \pi_m &= P_m Q_m - cQ_m = \left(\frac{c}{1+1/e} - c\right) Q_m \\ &= \left(\frac{-c/e}{1+1/e}\right) \cdot \left(\frac{c}{1+1/e}\right)^{e} = -\left(\frac{c}{1+1/e}\right)^{e+1} \cdot \frac{1}{e}. \end{aligned} \tag{14.20}$$

Dividing this expression by Equation 14.17 yields

$$\frac{\pi_m}{CS_c} = \left(\frac{e+1}{e}\right)\left(\frac{1}{1+1/e}\right)^{e+1} = \left(\frac{e}{1+e}\right)^{e}. \tag{14.21}$$

For $e = -2$ this ratio is 1/4. Hence one fourth of the consumer surplus enjoyed under perfect competition is transferred into monopoly profits. Therefore, the deadweight loss from monopoly in this case is also a fourth of the level of consumer surplus under perfect competition.

**QUERY:** Suppose $e = -1.5$. What fraction of consumer surplus is lost through monopolization? How much is transferred into monopoly profits? Why do these results differ from the case $e = -2$?

# MONOPOLY, PRODUCT QUALITY, AND DURABILITY

The market power enjoyed by a monopoly may be exercised along dimensions other than the market price of its product. If the monopoly has some leeway in the type, quality, or diversity of the goods it produces, then it would not be surprising for the firm's decisions to differ from those that might prevail under a competitive organization of the market. Whether a monopoly will produce higher-quality or lower-quality goods than would be produced under competition is unclear, however. It all depends on the firm's costs and the nature of consumer demand.

## A formal treatment of quality

Suppose consumers' willingness to pay for quality ($X$) is given by the inverse demand function $P(Q, X)$, where

$$\frac{\partial P}{\partial Q} < 0 \quad \text{and} \quad \frac{\partial P}{\partial X} > 0.$$

If the costs of producing $Q$ and $X$ are given by $C(Q, X)$, the monopoly will choose $Q$ and $X$ to maximize

$$\pi = P(Q, X)Q - C(Q, X). \tag{14.22}$$

The first-order conditions for a maximum are

$$\frac{\partial \pi}{\partial Q} = P(Q, X) + Q\frac{\partial P}{\partial Q} - C_Q = 0, \tag{14.23}$$

$$\frac{\partial \pi}{\partial X} = Q\frac{\partial P}{\partial X} - C_X = 0. \tag{14.24}$$

The first of these equations repeats the usual rule that marginal revenue equals marginal cost for output decisions. The second equation states that, when $Q$ is appropriately set, the monopoly should choose that level of quality for which the marginal revenue attainable from increasing the quality of its output by one unit is equal to the marginal cost of making such an increase. As might have been expected, the assumption of profit maximization requires the monopolist to proceed to the margin of profitability along all the dimensions it can. Notice, in particular, that the marginal demander's valuation of quality per unit is multiplied by the monopolist's output level when determining the profit-maximizing choice.

The level of product quality chosen under competitive conditions will also be the one that maximizes net social welfare:

$$SW = \int_0^{Q^*} P(Q, X)\, dQ - C(Q, X), \tag{14.25}$$

where $Q^*$ is the output level determined through the competitive process of marginal cost pricing, given $X$. Differentiation of Equation 14.25 with respect to $X$ yields the first-order condition for a maximum:

$$\frac{\partial SW}{\partial X} = \int_0^{Q^*} P_X(Q, X)\, dQ - C_X = 0. \tag{14.26}$$

The monopolist's choice of quality in Equation 14.24 targets the marginal consumer. The monopolist cares about the marginal consumer's valuation of quality because increasing the attractiveness of the product to the marginal consumer is how it increases sales. The perfectly competitive market ends up providing a quality level in Equation 14.26, maximizing total consumer surplus (the total after subtracting the cost of providing that quality level), which is the same as the quality level that maximizes consumer surplus for the *average* consumer.[7] Therefore, even if a monopoly and a perfectly competitive industry choose the same output level, they might opt for differing quality levels because each is

[7] The average marginal valuation ($AV$) of product quality is given by

$$AV = \int_0^{Q^*} P_X(Q, X)\, dQ/Q.$$

Hence $Q \cdot AV = C_x$ is the quality rule adopted to maximize net welfare under perfect competition. Compare this with Equation 14.24.

concerned with a different margin in its decision making. Only by knowing the specifics of the problem is it possible to predict the direction of these differences. For an example, see Problem 14.9; more detail on the theory of product quality and monopoly is provided in Problem 14.11.

## The durability of goods

Much of the research on the effect of monopolization on quality has focused on durable goods. These are goods such as automobiles, houses, or refrigerators that provide services to their owners over several periods rather than being completely consumed soon after they are bought. The element of time that enters into the theory of durable goods leads to many interesting problems and paradoxes. Initial interest in the topic started with the question of whether monopolies would produce goods that lasted as long as similar goods produced under perfect competition. The intuitive notion that monopolies would "underproduce" durability (just as they choose an output below the competitive level) was soon shown to be incorrect by the Australian economist Peter Swan in the early 1970s.[8]

Swan's insight was to view the demand for durable goods as the demand for a flow of services (i.e., automobile transportation) over several periods. He argued that both a monopoly and a competitive market would seek to minimize the cost of providing this flow to consumers. The monopoly would, of course, choose an output level that restricted the flow of services to maximize profits, but—assuming constant returns to scale in production—there is no reason that durability per se would be affected by market structure. This result is sometimes referred to as *Swan's independence assumption*. Output decisions can be treated independently from decisions about product durability.

Subsequent research on the Swan result has focused on showing how it can be undermined by different assumptions about the nature of a particular durable good or by relaxing the implicit assumption that all demanders are the same. For example, the result depends critically on how durable goods deteriorate. The simplest type of deterioration is illustrated by a durable good, such as a lightbulb, that provides a constant stream of services until it becomes worthless. With this type of good, Equations 14.24 and 14.26 are identical, so Swan's independence result holds. Even when goods deteriorate smoothly, the independence result continues to hold if a constant flow of services can be maintained by simply replacing what has been used—this requires that new goods and old goods be perfect substitutes and infinitely divisible. Outdoor house paint may, more or less, meet this requirement. On the other hand, most goods clearly do not. It is just not possible to replace a run-down refrigerator with, say, half of a new one. Once such more complex forms of deterioration are considered, Swan's result may not hold because we can no longer fall back on the notion of providing a given flow of services at minimal cost over time. In these more complex cases, however, it is not always the case that a monopoly will produce less durability than will a competitive market—it all depends on the nature of the demand for durability.

## Time inconsistency and heterogeneous demand

Focusing on the service flow from durable goods provides important insights on durability, but it does leave an important question unanswered—when should the monopoly produce the actual durable goods needed to provide the desired service flow? Suppose, for example, that a lightbulb monopoly decides that its profit-maximizing output decision is to supply the services provided by 1 million 60-watt bulbs. If the firm decides to

[8] P. L. Swan, "Durability of Consumption Goods," *American Economic Review* (December 1970): 884–94.

produce 1 million bulbs in the first period, what is it to do in the second period (say, before any of the original bulbs burn out)? Because the monopoly chooses a point on the service demand curve where $P > MC$, it has a clear incentive to produce more bulbs in the second period by cutting price a bit. But consumers can anticipate this, so they may reduce their first-period demand, waiting for a bargain. Hence the monopoly's profit-maximizing plan will unravel. Ronald Coase was the first economist to note this "time inconsistency" that arises when a monopoly produces a durable good.[9] Coase argued that its presence would severely undercut potential monopoly power—in the limit, competitive pricing is the only outcome that can prevail in the durable goods case. Only if the monopoly can succeed in making a credible commitment not to produce more in the second period can it succeed in its plan to achieve monopoly profits on the service flow from durable goods.

Recent modeling of the durable goods question has examined how a monopolist's choices are affected in situations where there are different types of demanders.[10] In such cases, questions about the optimal choice of durability and about credible commitments become even more complicated. The monopolist must not only settle on an optimal scheme for each category of buyers, it must also ensure that the scheme intended for (say) type-1 demanders is not also attractive to type-2 demanders. Studying these sorts of models would take us too far afield, but some illustrations of how such "incentive compatibility constraints" work are provided in the Extensions to this chapter and in Chapter 18.

# PRICE DISCRIMINATION

In some circumstances a monopoly may be able to increase profits by departing from a single-price policy for its output. The possibility of selling identical goods at different prices is called *price discrimination*.[11]

**DEFINITION**

**Price discrimination.** A monopoly engages in *price discrimination* if it is able to sell otherwise identical units of output at different prices.

Examples of price discrimination include senior citizen discounts for restaurant meals (which could instead be viewed as a price premium for younger customers), coffee sold at a lower price per ounce when bought in larger cup sizes, and different (net) tuition charged to different college students after subtracting their more or less generous financial aid awards. A "nonexample" of price discrimination might be higher auto insurance premiums charged to younger drivers. It might be clearer to think of the insurance policies sold to younger and older drivers as being different products to the extent that younger drivers are riskier and result in many more claims having to be paid.

Whether a price discrimination strategy is feasible depends crucially on the inability of buyers of the good to practice arbitrage. In the absence of transactions or information costs, the "law of one price" implies that a homogeneous good must sell everywhere for the same price. Consequently, price discrimination schemes are doomed to failure because demanders who can buy from the monopoly at lower prices will be more attractive sources

[9]R. Coase, "Durability and Monopoly," *Journal of Law and Economics* (April 1972): 143–49.

[10]For a summary, see M. Waldman, "Durable Goods Theory for Real World Markets," *Journal of Economic Perspectives* (Winter 2003): 131–54.

[11]A monopoly may also be able to sell differentiated products at differential price–cost margins. Here, however, we treat price discrimination only for a monopoly that produces a single homogeneous product. Price discrimination is an issue in other imperfectly competitive markets besides monopoly but is easiest to study in the simple case of a single firm.

of the good—for those who must pay high prices—than is the monopoly itself. Profit-seeking middlemen will destroy any discriminatory pricing scheme. However, when resale is costly or can be prevented entirely, then price discrimination becomes possible.

## First-degree or perfect price discrimination

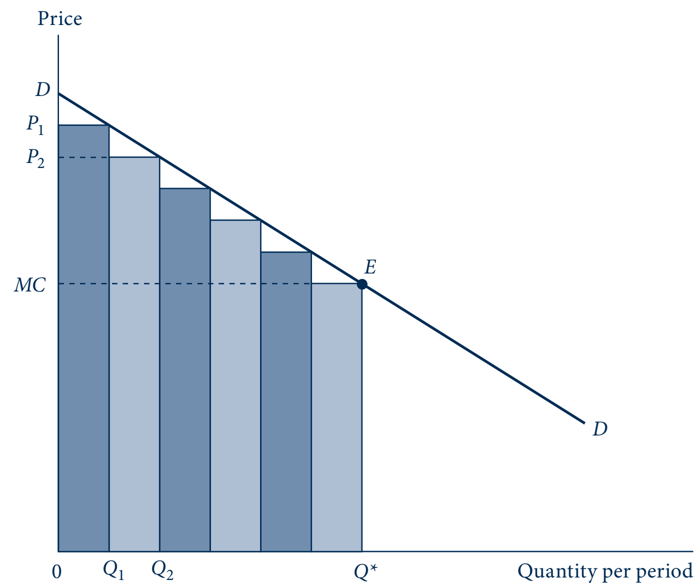
If each buyer can be separately identified by a monopolist, then it may be possible to charge each the maximum price he or she would willingly pay for the good. This strategy of *perfect* (or *first-degree*) price discrimination would then extract all available consumer surplus, leaving demanders as a group indifferent between buying the monopolist's good or doing without it. The strategy is illustrated in Figure 14.4. The figure assumes that buyers are arranged in descending order of willingness to pay. The first buyer is willing to pay up to $P_1$ for $Q_1$ units of output; therefore, the monopolist charges $P_1$ and obtains total revenues of $P_1Q_1$, as indicated by the lightly shaded rectangle. A second buyer is willing to pay up to $P_2$ for $Q_2 - Q_1$ units of output; therefore, the monopolist obtains total revenue of $P_2(Q_2 - Q_1)$ from this buyer. Notice that this strategy cannot succeed unless the second buyer is unable to resell the output he or she buys at $P_2$ to the first buyer (who pays $P_1 > P_2$).

The monopolist will proceed in this way up to the marginal buyer, the last buyer who is willing to pay at least the good's marginal cost (labeled *MC* in Figure 14.4). Hence total quantity produced will be $Q^*$. Total revenues collected will be given by the area $DEQ^*0$. All consumer surplus has been extracted by the monopolist, and there is no deadweight loss in this situation. (Compare Figures 14.3 and 14.4.) Therefore, the allocation of resources under perfect price discrimination is efficient, although it does entail a large transfer from consumer surplus into monopoly profits.

**FIGURE 14.4**

Perfect Price Discrimination

Under perfect price discrimination, the monopoly charges a different price to each buyer. It sells $Q_1$ units at $P_1$, $Q_2 - Q_1$ units at $P_2$, and so forth. In this case the firm will produce $Q^*$, and total revenues will be $DEQ^*0$.



### EXAMPLE 14.4 First-Degree Price Discrimination

Consider again the Frisbee monopolist in Example 14.1. Because there are relatively few high-quality Frisbees sold, the monopolist may find it possible to discriminate perfectly among a few world-class flippers. In this case, it will choose to produce that quantity for which the marginal buyer pays exactly the marginal cost of a Frisbee:

$$P = 100 - \frac{Q}{20} = MC = 0.1Q. \tag{14.27}$$

Hence

$$Q^* = 666$$

and, at the margin, price and marginal cost are given by

$$P = MC = 66.6. \tag{14.28}$$

Now we can compute total revenues by integration:

$$\begin{aligned} R &= \int_0^{Q^*} P(Q)\,dQ = \left(100Q - \frac{Q^2}{40}\right)\Bigg|_{Q=0}^{Q=666} \\ &= 55{,}511. \end{aligned} \tag{14.29}$$

Total costs are

$$C(Q) = 0.05Q^2 + 10{,}000 = 32{,}178; \tag{14.30}$$

total profits are given by

$$\pi = R - C = 23{,}333, \tag{14.31}$$

which represents a substantial increase over the single-price policy examined in Example 14.1 (which yielded 15,000).

**QUERY:** What is the maximum price any Frisbee buyer pays in this case? Use this to obtain a geometric definition of profits.

## Third-degree price discrimination through market separation

First-degree price discrimination poses a considerable information burden for the monopoly—it must know the demand function for each potential buyer. A less stringent requirement would be to assume the monopoly can separate its buyers into relatively few identifiable markets (such as "rural–urban," "domestic–foreign," or "prime-time–off-prime") and pursue a separate monopoly pricing policy in each market. Knowledge of the price elasticities of demand in these markets is sufficient to pursue such a policy. The monopoly then sets a price in each market according to the inverse elasticity rule. Assuming that marginal cost is the same in all markets, the result is a pricing policy in which

$$P_i\left(1 + \frac{1}{e_i}\right) = P_j\left(1 + \frac{1}{e_j}\right) \tag{14.32}$$

or

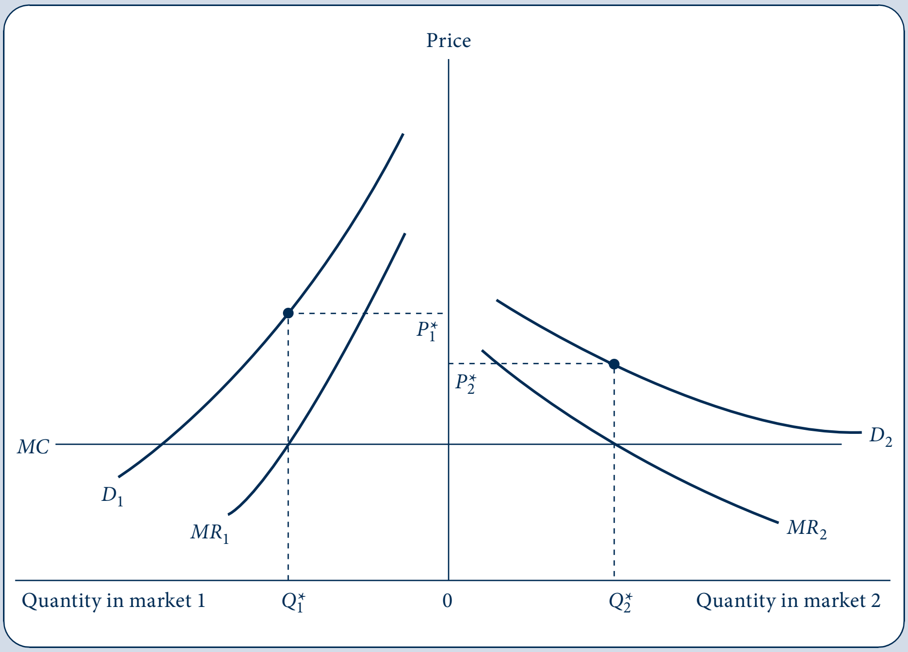
$$\frac{P_i}{P_j} = \frac{(1 + 1/e_j)}{(1 + 1/e_i)}, \tag{14.33}$$

where $P_i$ and $P_j$ are the prices charged in markets $i$ and $j$, which have price elasticities of demand given by $e_i$ and $e_j$. An immediate consequence of this pricing policy is that the

**FIGURE 14.5**

Separated Markets Raise the Possibility of Third-Degree Price Discrimination

If two markets are separate, then a monopolist can maximize profits by selling its product at different prices in the two markets. This would entail choosing that output for which $MC = MR$ in each of the markets. The diagram shows that the market with a less elastic demand curve will be charged the higher price by the price discriminator.




profit-maximizing price will be higher in markets in which demand is less elastic. If, for example, $e_i = -2$ and $e_j = -3$, then Equation 14.33 shows that $P_i/P_j = 4/3$—prices will be one third higher in market $i$, the less elastic market.

Figure 14.5 illustrates this result for two markets that the monopoly can serve at constant marginal cost ($MC$). Demand is less elastic in market 1 than in market 2; thus, the gap between price and marginal revenue is larger in the former market. Profit maximization requires that the firm produce $Q_1^*$ in market 1 and $Q_2^*$ in market 2, resulting in a higher price in the less elastic market. As long as arbitrage between the two markets can be prevented, this price difference can persist. The two-price discriminatory policy is clearly more profitable for the monopoly than a single-price policy would be because the firm can always opt for the latter policy should market conditions warrant.

The welfare consequences of third-degree price discrimination are, in principle, ambiguous. Relative to a single-price policy, the discriminating policy requires raising the price in the less elastic market and reducing it in the more elastic one. Hence the changes have an offsetting effect on total allocational losses. A more complete analysis suggests the intuitively plausible conclusion that the multiple-price policy will be allocationally superior to a single-price policy only in situations in which total output is increased through discrimination. Example 14.5 illustrates a simple case of linear demand curves in which a single-price policy does result in greater allocational losses.[12]

[12]For a detailed discussion, see R. Schmalensee, "Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination," *American Economic Review* (March 1981): 242–47. See also Problem 14.13.

## EXAMPLE 14.5 Third-Degree Price Discrimination

Suppose that a monopoly producer of widgets has a constant marginal cost of $c = 6$ and sells its products in two separated markets whose inverse demand functions are

$$P_1 = 24 - Q_1 \quad \text{and} \quad P_2 = 12 - 0.5Q_2. \tag{14.34}$$

Notice that consumers in market 1 are more eager to buy than are consumers in market 2 in the sense that the former are willing to pay more for any given quantity. Using the results for linear demand curves from Example 14.2 shows that the profit-maximizing price–quantity combinations in these two markets are:

$$P_1^* = \frac{24+6}{2} = 15, \quad Q_1^* = 9, \quad P_2^* = \frac{12+6}{2} = 9, \quad Q_2^* = 6. \tag{14.35}$$

With this pricing strategy, profits are $\pi = (15 - 6) \cdot 9 + (9 - 6) \cdot 6 = 81 + 18 = 99$. We can compute the deadweight losses in the two markets by recognizing that the competitive output (with $P = MC = 6$) in market 1 is 18 and in market 2 is 12:

$$\begin{aligned} DW &= DW_1 + DW_2 \\ &= 0.5(P_1^* - 6)(18 - 9) + 0.5(P_2^* - 6)(12 - 6) \\ &= 40.5 + 9 = 49.5. \end{aligned} \tag{14.36}$$

**A single-price policy.** In this case, constraining the monopoly to charge a single price would reduce welfare. Under a single-price policy, the monopoly would simply cease serving market 2 because it can maximize profits by charging a price of 15, and at that price no widgets will be bought in market 2 (because the maximum willingness to pay is 12). Therefore, total deadweight loss in this situation is increased from its level in Equation 14.36 because total potential consumer surplus in market 2 is now lost:

$$DW = DW_1 + DW_2 = 40.5 + 0.5(12 - 6)(12 - 0) = 40.5 + 36 = 76.5. \tag{14.37}$$

This illustrates a situation where third-degree price discrimination is welfare improving over a single-price policy—when the discriminatory policy permits "smaller" markets to be served. Whether such a situation is common is an important policy question (e.g., consider the case of U.S. pharmaceutical manufacturers charging higher prices at home than abroad).

**QUERY:** Suppose these markets were no longer separated. How would you construct the market demand in this situation? Would the monopolist's profit-maximizing single price still be 15?

# SECOND-DEGREE PRICE DISCRIMINATION THROUGH PRICE SCHEDULES

The examples of price discrimination examined in the previous section require the monopoly to separate demanders into a number of categories and then choose a profit-maximizing price for each such category. An alternative approach would be for the monopoly to choose a (possibly rather complex) price schedule that provides incentives for demanders to separate themselves depending on how much they wish to buy. Such schemes include quantity discounts, minimum purchase requirements or "cover" charges, and tie-in sales. These plans would be adopted by a monopoly if they yielded greater profits than would a single-price policy, after accounting for any possible costs of implementing the price schedule. Because the schedules will result in demanders paying different prices for identical goods, this form of (second-degree) price discrimination is feasible only when there are no arbitrage possibilities. Here we look at one simple case. The Extensions to this chapter and portions of Chapter 18 look at other aspects of second-degree price discrimination.

## Two-part tariffs

One form of pricing schedule that has been extensively studied is a linear two-part tariff, under which demanders must pay a fixed fee for the right to consume a good and a uniform price for each unit consumed. The prototype case, first studied by Walter Oi, is an amusement park (perhaps Disneyland) that sets a basic entry fee coupled with a stated marginal price for each amusement used.[13] Mathematically, this scheme can be represented by the tariff any demander must pay to purchase $q$ units of a good:

$$T(q) = a + pq, \tag{14.38}$$

where $a$ is the fixed fee and $p$ is the marginal price to be paid. The monopolist's goal then is to choose $a$ and $p$ to maximize profits, given the demand for this product. Because the average price paid by any demander is given by

$$\overline{p} = \frac{T}{q} = \frac{a}{q} + p, \tag{14.39}$$

this tariff is feasible only when those who pay low average prices (those for whom $q$ is large) cannot resell the good to those who must pay high average prices (those for whom $q$ is small).

One approach described by Oi for establishing the parameters of this linear tariff would be for the firm to set the marginal price, $p$, equal to $MC$ and then set $a$ to extract the maximum consumer surplus from a given set of buyers. One might imagine buyers being arrayed according to willingness to pay. The choice of $p = MC$ would then maximize consumer surplus for this group, and $a$ could be set equal to the surplus enjoyed by the least eager buyer. He or she would then be indifferent about buying the good, but all other buyers would experience net gains from the purchase.

This feasible tariff might not be the most profitable, however. Consider the effects on profits of a small increase in $p$ above $MC$. This would result in no net change in the profits earned from the least willing buyer. Quantity demanded would drop slightly at the margin where $p = MC$, and some of what had previously been consumer surplus (and therefore part of the fixed fee, $a$) would be converted into variable profits because now $p > MC$. For all other demanders, profits would be increased by the price increase. Although each will pay a bit less in fixed charges, profits per unit bought will increase to a greater extent.[14] In some cases it is possible to make an explicit calculation of the optimal two-part tariff. Example 14.6 provides an illustration. More generally, however, optimal schedules will depend on a variety of contingencies. Some of the possibilities are examined in the Extensions to this chapter.

[13]W. Y. Oi, "A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly," *Quarterly Journal of Economics* (February 1971): 77–90. Interestingly, the Disney empire once used a two-part tariff but abandoned it because the costs of administering the payment schemes for individual rides became too high. Like other amusement parks, Disney moved to a single-admissions price policy (which still provided them with ample opportunities for price discrimination, especially with the multiple parks at Disney World).

[14]This follows because $q_i(MC) > q_1(MC)$, where $q_i(MC)$ is the quantity demanded when $p = MC$ for all except the least willing buyer (person 1). Hence the gain in profits from an increase in price above $MC$, $\Delta p q_i(MC)$, exceeds the loss in profits from a smaller fixed fee, $\Delta p q_1(MC)$.

**EXAMPLE 14.6 Two-Part Tariffs**

To illustrate the mathematics of two-part tariffs, let's return to the demand equations introduced in Example 14.5 but now assume that they apply to two specific demanders:

$$\begin{aligned} q_1 &= 24 - p_1, \\ q_2 &= 24 - 2p_2, \end{aligned} \tag{14.40}$$

where now the $p$'s refer to the marginal prices faced by these two buyers.[15]

**An Oi tariff.** Implementing the two-part tariff suggested by Oi would require the monopolist to set $p_1 = p_2 = MC = 6$. Hence in this case, $q_1 = 18$ and $q_2 = 12$. With this marginal price, demander 2 (the less eager of the two) obtains consumer surplus of 36 $[= 0.5 \cdot (12 - 6) \cdot 12]$. That is the maximal entry fee that might be charged without causing this person to leave the market. Consequently, the two-part tariff in this case would be $T(q) = 36 + 6q$. If the monopolist opted for this pricing scheme, its profits would be

$$\begin{aligned} \pi = R - C &= T(q_1) + T(q_2) - AC(q_1 + q_2) \\ &= 72 + 6 \cdot 30 - 6 \cdot 30 = 72. \end{aligned} \tag{14.41}$$

These fall short of those obtained in Example 14.5.

**The optimal tariff.** The optimal two-part tariff in this situation can be computed by noting that total profits with such a tariff are $\pi = 2a + (p - MC)(q_1 + q_2)$. Here the entry fee, $a$, must equal the consumer surplus obtained by person 2. Inserting the specific parameters of this problem yields

$$\begin{aligned} \pi &= 0.5 \cdot 2q_2(12 - p) + (p - 6)(q_1 + q_2) \\ &= (24 - 2p)(12 - p) + (p - 6)(48 - 3p) \\ &= 18p - p^2. \end{aligned} \tag{14.42}$$

Hence maximum profits are obtained when $p = 9$ and $a = 0.5(24 - 2p)(12 - p) = 9$. Therefore, the optimal tariff is $T(q) = 9 + 9q$. With this tariff, $q_1 = 15$ and $q_2 = 6$, and the monopolist's profits are 81 $[= 2(9) + (9 - 6) \cdot (15 + 6)]$. The monopolist might opt for this pricing scheme if it were under political pressure to have a uniform pricing policy and to agree not to price demander 2 "out of the market." The two-part tariff permits a degree of differential pricing $(\bar{p}_1 = 9.60, \bar{p}_2 = 9.75)$ but appears "fair" because all buyers face the same schedule.

**QUERY:** Suppose a monopolist could choose a different entry fee for each demander. What pricing policy would be followed?

# REGULATION OF MONOPOLY

The regulation of natural monopolies is an important subject in applied economic analysis. The utility, communications, and transportation industries are highly regulated in most countries, and devising regulatory procedures that induce these industries to operate in a desirable way is an important practical problem. Here we will examine a few aspects of the regulation of monopolies that relate to pricing policies.

## Marginal cost pricing and the natural monopoly dilemma

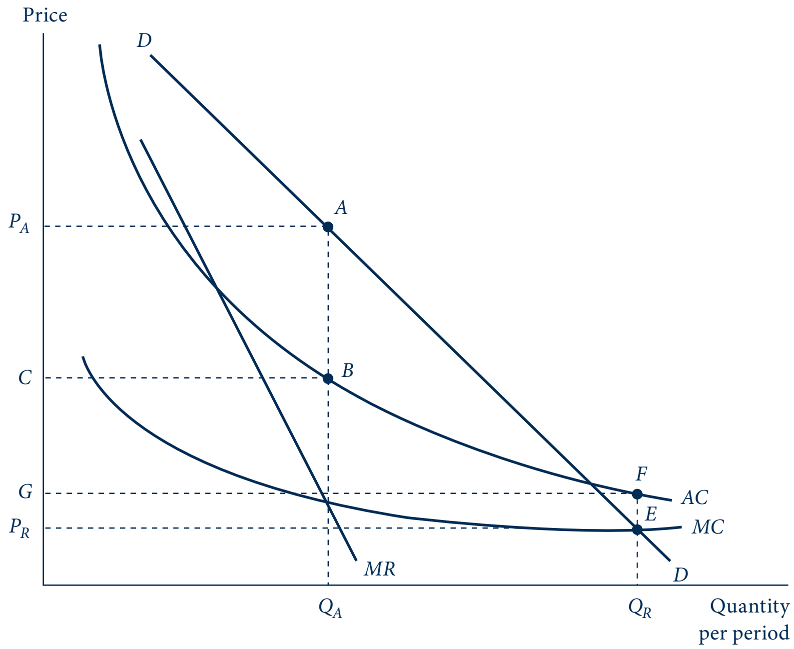
Many economists believe it is important for the prices charged by regulated monopolies to reflect marginal costs of production accurately. In this way the deadweight loss may be

[15] The theory of utility maximization that underlies these demand curves is that the quantity demanded is determined by the marginal price paid, whereas the entry fee $a$ determines whether $q = 0$ might instead be optimal.

**FIGURE 14.6**

Price Regulation for a Decreasing Cost Monopoly

Because natural monopolies exhibit decreasing average costs, marginal costs decrease below average costs. Consequently, enforcing a policy of marginal cost pricing will entail operating at a loss. A price of $P_R$, for example, will achieve the goal of marginal cost pricing but will necessitate an operating loss of $GFEP_R$.




minimized. The principal problem raised by an enforced policy of marginal cost pricing is that it will require natural monopolies to operate at a loss. Natural monopolies, by definition, exhibit decreasing average costs over a broad range of output levels. The cost curves for such a firm might look like those shown in Figure 14.6. In the absence of regulation, the monopoly would produce output level $Q_A$ and receive a price of $P_A$ for its product. Profits in this situation are given by the rectangle $P_AABC$. A regulatory agency might instead set a price of $P_R$ for the monopoly. At this price, $Q_R$ is demanded, and the marginal cost of producing this output level is also $P_R$. Consequently, marginal cost pricing has been achieved. Unfortunately, because of the negative slope of the firm's average cost curve, the price $P_R$ (= marginal cost) decreases below average costs. With this regulated price, the monopoly must operate at a loss of $GFEP_R$. Because no firm can operate indefinitely at a loss, this poses a dilemma for the regulatory agency: Either it must abandon its goal of marginal cost pricing, or the government must subsidize the monopoly forever.

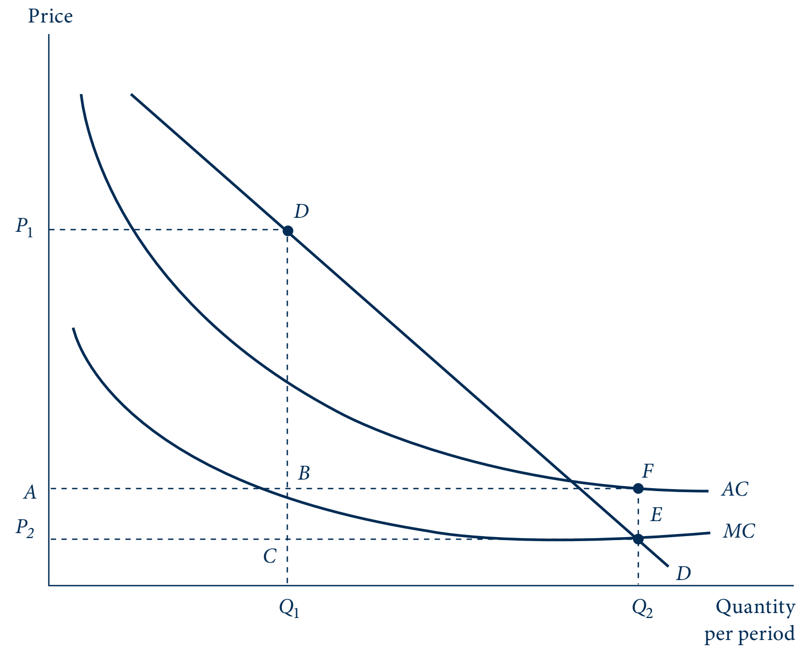
## Two-tier pricing systems

One way out of the marginal cost pricing dilemma is the implementation of a multiprice system. Under such a system the monopoly is permitted to charge some users a high price while maintaining a low price for marginal users. In this way the demanders paying the high price in effect subsidize the losses of the low-price customers. Such a pricing scheme is shown in Figure 14.7. Here the regulatory commission has decided that some users will pay a relatively high price, $P_1$. At this price, $Q_1$ is demanded. Other users

**FIGURE 14.7**

Two-Tier Pricing Schedule

By charging a high price ($P_1$) to some users and a low price ($P_2$) to others, it may be possible for a regulatory commission to (1) enforce marginal cost pricing and (2) create a situation where the profits from one class of user ($P_1DBA$) subsidize the losses of the other class ($BFEC$).




(presumably those who would not buy the good at the $P_1$ price) are offered a lower price, $P_2$. This lower price generates additional demand of $Q_2 - Q_1$. Consequently, a total output of $Q_2$ is produced at an average cost of $A$. With this pricing system, the profits on the sales to high-price demanders (given by the rectangle $P_1DBA$) balance the losses incurred on the low-priced sales ($BFEC$). Furthermore, for the "marginal user," the marginal cost pricing rule is being followed: It is the "intramarginal" user who subsidizes the firm so it does not operate at a loss. Although in practice it may not be so simple to establish pricing schemes that maintain marginal cost pricing and cover operating costs, many regulatory commissions do use price schedules that intentionally discriminate against some users (e.g., businesses) to the advantage of others (consumers).

## Rate of return regulation

Another approach followed in many regulatory situations is to permit the monopoly to charge a price above marginal cost that is sufficient to earn a "fair" rate of return on investment. Much analytical effort is then devoted to defining the "fair" rate concept and to developing ways in which it might be measured. From an economic point of view, some of the most interesting questions about this procedure concern how the regulatory activity affects the firm's input choices. If, for example, the rate of return allowed to firms exceeds what owners might obtain on investment under competitive circumstances, there will be an incentive to use relatively more capital input than would truly minimize costs. And if regulators delay in making rate decisions, this may give firms cost-minimizing incentives that would not otherwise exist. We will now briefly examine a formal model of such possibilities.[16]

## A formal model

Suppose a regulated utility has a production function of the form

$$q = f(k, l). \tag{14.43}$$

This firm's actual rate of return on capital is then defined as

$$s = \frac{pf(k, l) - wl}{k}, \tag{14.44}$$

where $p$ is the price of the firm's output (which depends on $q$) and $w$ is the wage rate for labor input. If $s$ is constrained by regulation to be equal to (say) $\bar{s}$, then the firm's problem is to maximize profits

$$\pi = pf(k, l) - wl - vk \tag{14.45}$$

subject to this regulatory constraint. The Lagrangian for this problem is

$$\mathcal{L} = pf(k, l) - wl - vk + \lambda[wl + \bar{s}k - pf(k, l)]. \tag{14.46}$$

Notice that if $\lambda = 0$, regulation is ineffective and the monopoly behaves like any profit-maximizing firm. If $\lambda = 1$, Equation 14.46 reduces to

$$\mathcal{L} = (\bar{s} - v)k, \tag{14.47}$$

which, assuming $\bar{s} > v$ (which it must be if the firm is not to earn less than the prevailing rate of return on capital elsewhere), means this monopoly will hire infinite amounts of capital—an implausible result. Hence $0 < \lambda < 1$. The first-order conditions for a maximum are

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial l} &= pf_l - w + \lambda(w - pf_1) = 0,\\
\frac{\partial \mathcal{L}}{\partial k} &= pf_k - v + \lambda(\bar{s} - pf_k) = 0,\\
\frac{\partial \mathcal{L}}{\partial \lambda} &= wl - +\bar{s}k - pf(k, l) = 0.
\end{aligned} \tag{14.48}$$

The first of these conditions implies that the regulated monopoly will hire additional labor input up to the point at which $pf_l = w$—a result that holds for any profit-maximizing firm. For capital input, however, the second condition implies that

$$(1 - \lambda)pf_k = v - \lambda\bar{s} \tag{14.49}$$

or

$$pf_k = \frac{v - \lambda\bar{s}}{1 - \lambda} = v - \frac{\lambda(\bar{s} - v)}{1 - \lambda}. \tag{14.50}$$

Because $\bar{s} > v$ and $\lambda < 1$, Equation 14.50 implies

$$pf_k < v. \tag{14.51}$$

[16]This model is based on H. Averch and L. L. Johnson, "Behavior of the Firm under Regulatory Constraint," *American Economic Review* (December 1962): 1052–69.

The firm will hire more capital (and achieve a lower marginal productivity of capital) than it would under unregulated conditions. Therefore, "overcapitalization" may be a regulatory-induced misallocation of resources for some utilities. Although we shall not do so here, it is possible to examine other regulatory questions using this general analytical framework.

## DYNAMIC VIEWS OF MONOPOLY

The static view that monopolistic practices distort the allocation of resources provides the principal economic rationale for favoring antimonopoly policies. Not all economists believe that the static analysis should be definitive, however. Some authors, most notably J. A. Schumpeter, have stressed the beneficial role that monopoly profits can play in the process of economic development.[17] These authors place considerable emphasis on innovation and the ability of particular types of firms to achieve technical advances. In this context the profits that monopolistic firms earn provide funds that can be invested in research and development. Whereas perfectly competitive firms must be content with a normal return on invested capital, monopolies have "surplus" funds with which to undertake the risky process of research. More important, perhaps, the possibility of attaining a monopolistic position—or the desire to maintain such a position—provides an important incentive to keep one step ahead of potential competitors. Innovations in new products and cost-saving production techniques may be integrally related to the possibility of monopolization. Without such a monopolistic position, the full benefits of innovation could not be obtained by the innovating firm.

Schumpeter stresses the point that the monopolization of a market may make it less costly for a firm to plan its activities. Being the only source of supply for a product eliminates many of the contingencies that a firm in a competitive market must face. For example, a monopoly may not have to spend as much on selling expenses (e.g., advertising, brand identification, and visiting retailers) as would be the case in a more competitive industry. Similarly, a monopoly may know more about the specific demand curve for its product and may more readily adapt to changing demand conditions. Of course, whether any of these purported benefits of monopolies outweigh their allocational and distributional disadvantages is an empirical question. Issues of innovation and cost savings cannot be answered by recourse to a priori arguments; detailed investigation of real-world markets is a necessity.

## SUMMARY

In this chapter we have examined models of markets in which there is only a single monopoly supplier. Unlike the competitive case investigated in Part 4, monopoly firms do not exhibit price-taking behavior. Instead, the monopolist can choose the price–quantity combination on the market demand curve that is most profitable. A number of consequences then follow from this market power.

- The most profitable level of output for the monopolist is the one for which marginal revenue is equal to marginal cost. At this output level, price will exceed marginal cost. The profitability of the monopolist will depend on the relationship between price and average cost.
- Relative to perfect competition, monopoly involves a loss of consumer surplus for demanders. Some of this is transferred into monopoly profits, whereas some of the loss in consumer supply represents a deadweight loss of overall economic welfare.

[17]See, for example, J. A. Schumpeter, *Capitalism, Socialism and Democracy*, 3rd ed. (New York: Harper & Row, 1950), especially chap. 8.