1

Monopoly

This chapter discusses various arguments made in favor of and in opposition to monopoly power. We will assume here that the goods produced by the monopolist are given, and that their qualities are known by the consumers. We will also assume that the monopolist charges the same price per unit of good for each good produced. (More specifically, there is no price discrimination at a given point of time. We will, however, consider intertemporal price discrimination.)

The best-known monopoly distortion, that related to pricing strategy, will be tackled in section 1.1. In contrast with the behavior of a competitive firm whose product demand is infinitely elastic by definition (and which takes the price as given), a firm exercising monopoly power over a given market can raise its price above marginal cost without losing all its clients. Such behavior leads to a price that is too high and to a "dead-weight" welfare loss for society (unless the firm is able to "price-discriminate" perfectly, as we shall see in chapter 3).

We shall recall the main aspects of the pricing behavior of a single-product monopolist. We shall then consider a multiproduct monopolist with interrelated production costs of or interrelated demands for his various products. Last, we shall study the intertemporal pricing behavior of a durable-good monopolist.

Other distortions may also exist. On the one hand, both theory and practice suggest that it is more difficult for the owners of a firm to keep control over its costs when the firm has monopoly power on the product market. Thus, a monopolist may produce given outputs at a higher cost than a competitive firm (section 1.2). On the other hand, the monopoly rent may give rise to a contest among several firms to obtain or secure it. This contest may involve socially wasteful expenditures, which partly dissipate the monopoly rent. Therefore, monopoly profit should not always be taken into account in the expression of welfare (section 1.3).

Naturally, the conclusions would hold as well for monopsony power (i.e., monopoly power in the input markets).

1.1 Pricing Behavior

The best-known monopoly distortion results from the monopolist's pricing behavior. To focus on this distortion, we assume that the monopolist's products are given and that their existence and quality are known to consumers. We start by reviewing the distortionary markup by a monopoly producer of a single good. We then study the multiproduct monopolist. Last, we consider the issue of intertemporal pricing by a durable-good monopoly.

1.1.1 A Single-Product Monopolist

1.1.1.1 The Inverse Elasticity Rule

Let $q = D(p)$ be the demand for the good produced by the monopoly, with inverse demand function $p = P(q)$. Let $C(q)$ be the cost of producing q units of this good. Assume that demand is differentiable and decreasing with the price (i.e., $D'(p) < 0$),[1] and that cost is differentiable and increasing with the output. A profit-maximizing monopolist chooses the monopoly price p^{m} so as to

$$\max_{p} \, [p\,D(p) - C(D(p))].$$

The first-order condition for this problem is

$$p^{\mathrm{m}} - C'(D(p^{\mathrm{m}})) = -\frac{D(p^{\mathrm{m}})}{D'(p^{\mathrm{m}})},$$

or

$$\frac{p^{\mathrm{m}} - C'}{p^{\mathrm{m}}} = \frac{1}{\varepsilon}, \tag{1.1}$$

where $\varepsilon = -D'p^{\mathrm{m}}/D$ denotes the demand elasticity at the monopoly price p^{m}. Letting $q^{\mathrm{m}} \equiv D(p^{\mathrm{m}})$ denote the monopoly output, one can rewrite the first-order condition as the equality between marginal revenue and marginal cost:

$$\mathrm{MR}(q^{\mathrm{m}}) \equiv P(q^{\mathrm{m}}) + P'(q^{\mathrm{m}})q^{\mathrm{m}} = C'(q^{\mathrm{m}}).$$

For now, we ignore the second-order condition of the maximization problem. Equation 1.1 indicates that the relative "markup"—the ratio between the profit margin (price minus marginal cost) and the price; also called the *Lerner index*—is inversely proportional to the demand elasticity. The monopoly sells at a price greater than the socially optimal price, which is its marginal cost.[2] The price distortion is larger when consumers, facing a price increase, reduce their demand only slightly. The intuition, of course, is that the monopolist is more wary of the perverse effect of a high price on consumption when consumers react to a price increase by greatly reducing their demand.

If the elasticity of demand is independent of price (the demand function is $q = kp^{-\varepsilon}$, where k is a positive constant), the Lerner index is constant. The monopolist adjusts his price to shocks on the marginal cost by using a constant (relative) markup rule. For instance, if his technology exhibits constant returns to scale so that the marginal cost is equal to the average cost or unit cost and if the elasticity of demand is 2, the monopolist systematically charges twice the unit cost. Thus, if we observe a monopolist using such a "rule of thumb," we should not necessarily conclude that this monopolist's pricing behavior is not (privately) optimal.

More generally, observe that a monopoly always operates in a price region such that the elasticity of demand (from equation 1.1) exceeds 1. Where the elasticity is lower than 1, the monopolist's revenue—and, *a fortiori*, his profit—are decreasing in quantity (i.e., increasing in price).

It is a simple yet a very general property of monopoly pricing that the monopoly price is a nondecreasing function of marginal cost. To see this, consider two alternative cost functions for the monopolist: $C_1(\cdot)$ and $C_2(\cdot)$. Assume that these cost functions are differentiable, and that $C_2'(q) > C_1'(q)$ for all $q > 0$. No other assumption on these cost functions is required. Let p_1^{m} and q_1^{m} denote the monopoly price and quantity when the cost function is $C_1(\cdot)$; p_2^{m} and q_2^{m} are defined similarly. When the cost function is $C_1(\cdot)$, the monopolist prefers charging p_1^{m} rather than any other price. In particular, he could charge price p_2^{m} and sell quantity q_2^{m}. Thus,

$$p_1^{\mathrm{m}}q_1^{\mathrm{m}} - C_1(q_1^{\mathrm{m}}) \geqslant p_2^{\mathrm{m}}q_2^{\mathrm{m}} - C_1(q_2^{\mathrm{m}}). \tag{1.2}$$

Similarly, the monopolist prefers to charge p_2^{m} rather than p_1^{m} when his cost function is $C_2(\cdot)$:

1. See the introduction.

2. See the introduction.

$$p_2^m q_2^m - C_2(q_2^m) \geqslant p_1^m q_1^m - C_2(q_1^m). \tag{1.3}$$

Adding equations 1.2 and 1.3 yields

$$[C_2(q_1^m) - C_2(q_2^m)] - [C_1(q_1^m) - C_1(q_2^m)] \geqslant 0, \tag{1.4}$$

or

$$\int_{q_2^m}^{q_1^m} [C_2'(x) - C_1'(x)]dx \geqslant 0. \tag{1.5}$$

Because $C_2'(x) > C_1'(x)$ for all x, equation 1.5 implies that $q_1^m \geqslant q_2^m$. In other words, the monopoly price is a nondecreasing function of marginal cost.[3]

1.1.1.2 The Dead-Weight Loss

Equation 1.1 provides a quantification of price distortion, but from a normative viewpoint the appropriate measure of distortion is the loss of social welfare. To measure the latter, we compare the total surplus at the monopoly price with that at the competitive (marginal-cost) price. The total surplus is equal to the sum of the consumer surplus and the producer surplus (or profit), or to the difference between total consumer utility and production costs.[4] In figure 1.1 this surplus is represented by the area *DGAD* under marginal-cost pricing and by the area *DEFAD* under monopoly pricing.

The net consumer surplus under monopoly is the area of the "triangle" *CDE* in figure 1.1. The monopolist's profit is equal to the total revenue, $p^m q^m$, minus the integral of the marginal cost—i.e., equal to the area of the "trapezoid" *ACEF*. Thus, the "dead-weight" welfare loss is equal to the area of the "triangle" *EFG*. (These are a proper triangle and trapezoid only if the demand and marginal-cost curves are linear.)

The welfare loss does not necessarily decrease with the elasticity of demand, even though the relative markup does (from equation 1.1). The monopoly situations for which we observe strong price distortions correspond to those in which demand elasticity is low, so that consumers decrease their quantity demanded only slightly in response to a unit price increase. Consequently, in precisely these situations, price changes do not affect quantity consumed very much; rather, they elicit a large

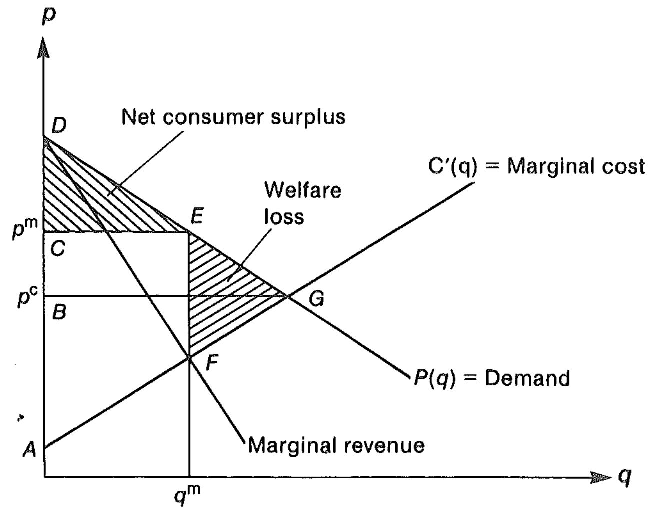

Figure 1.1

monetary transfer from consumers to the firm. Hence, we cannot conclude that the welfare loss is monotonic in the elasticity of demand.

*Exercise 1.1*** In a monopolized industry, the demand function has a constant elasticity: $q = D(p) = p^{-\varepsilon}$ where $\varepsilon > 1$ is the elasticity of demand. Marginal cost is constant and equal to c.

(i) Show that a social planner (or a competitive industry) would yield a total welfare of

$$W^c = c^{1-\varepsilon}/(\varepsilon - 1).$$

(ii) Compute the welfare loss, WL, under monopoly.

(iii) Show that the ratio WL/W^c (relative dead-weight loss) increases with ε, that WL is nonmonotonic in ε, and that the fraction Π^m/W^c of potential consumer surplus that can be captured by the monopolist increases with ε. Discuss the result. (Note that the "size" of the market changes with ε.)

*Exercise 1.2** Suppose that all consumers have unit demand. They buy 0 or 1 unit of the good produced by the monopolist. They are identical, and they have willingness to pay (valuation) $\bar{s}$ for the good. Show that monopoly pricing does not create a welfare loss.

3. This style of proof is familiar from the literature on incentives. Though less familiar in industrial organization, it will be used occasionally in this book.

The monopoly price may be nonunique owing to nonconcavities in the profit function. It is then a correspondence rather than a function. The result then says that *any* optimal price for cost function $C_2(\cdot)$ (weakly) exceeds *any* optimal price for cost function $C_1(\cdot)$.

4. This criterion ignores problems of income distribution; see the introduction.

Of course, the dead-weight welfare loss represents only what can be gained from moving from a monopoly situation to an ideal situation. It thus yields an upper bound on the efficiency gain to be realized by correcting monopoly pricing. The actual efficiency gain must be computed for any policy intervention that does not yield marginal-cost pricing. To put it another way, a high total distortion associated with the monopolization of an industry is a signal that some public intervention might be desirable, but it does not suggest a course of action. An analyst or a government should begin analyzing the causes of monopolization (see chapter 8), as well as the set of potential interventions. The latter will depend crucially on the information available to the social planner concerning industry conditions (cost structure, demand).

Remark The welfare loss can be measured empirically by estimating the demand curve and the marginal-cost curve. For a discussion of methodology and results see Scherer 1980, p. 461. Harberger's (1954) estimate of a total welfare loss not exceeding 0.1 percent of the gross national product implied that economists were wasting their time focusing on the monopoly-pricing problem and gave rise to much controversy about both the data and the methodology.[5] Industrial-organization economists are, in general, mainly interested in industries which are monopolized to at least some extent. An economy-average number understates the typical distortion in those industries because it includes many fairly competitive industries in the sample. The dead-weight loss is only one of the harmful effects of monopoly, as we shall see below. (Including some of the other distortions, such as those associated with rent seeking, some researchers found welfare losses of up to 7 percent of GNP. See Cowling and Mueller 1978 and Jenny and Weber 1983; see Scherer 1980 for a skeptical view of these high estimates.)

1.1.1.3 The Effect of Commodity Taxation

Consider one possible policy prescription for restoring the social optimum in the presence of monopoly. Suppose that the government taxes monopoly output at the rate t. Then the monopolist chooses p to

$$\max_p [p\,D(p+t) - C(D(p+t))],$$

from which it follows that

$$D(p+t) + D'(p+t)(p - C') = 0$$

or

$$[D(p+t) - t\,D'(p+t)] + D'(p+t)(p+t-C') = 0.$$

To restore the social optimum, marginal cost C' must coincide with the price faced by the consumers $(p+t)$ and thus with the marginal utility in terms of money to the consumers. Therefore, we must set

$$t = D(p^c)/D'(p^c) < 0$$

(i.e., $t/p^c = -1/\varepsilon$), where p^c is the competitive price (determined by the intersection of the demand and marginal-cost curves in figure 1.1). Since $t < 0$, we must subsidize the output of the monopolist. We can explain this rather paradoxical result as follows: The problem with monopoly pricing is that it induces consumers to consume too little of the good. In order to achieve an efficient allocation of resources, we induce them to consume more by subsidizing the good.

*Exercise 1.3** A monopolist's marginal cost of supplying a good to consumers is $\tilde{c} = c + t$ (where t is a unit commodity tax). Let $p^m(\tilde{c})$ denote the corresponding monopoly price.

(i) Compute $dp^m/d\tilde{c}$ for the following demand functions: $p = q^{-1/\varepsilon}$, $p = \alpha - \beta q^\delta$, $p = a - b \ln q$.

(ii) Sumner (1981) uses an ingenious approach to estimate the elasticity of demand—and thus the degree of monopoly power—in the American cigarette industry. He notes that in the United States, commodity taxes—and therefore the generalized cost $\tilde{c}$—vary across states. Although data on c are hard to obtain, data on t are readily available. Sumner uses varying levels of taxation across states to estimate the elasticity of demand. Bulow and Pfleiderer (1983) argue that the method has limited applicability. What do you think?

5. In particular, Harberger assumed unit demand elasticities, which creates a downward bias in the estimation of welfare losses. Furthermore, Bergson (1973) showed that Harberger's partial-equilibrium approach can be a major source of bias on either side. Another drawback of Harberger's approach is the identification of the competitive profit rate with the *mean* (cross-sectional) profit rate, which incorporates monopoly profits.

Despite the simplicity of the result, the subsidy solution has only a few advocates. Its critics point out that the concept of total surplus accords equal weight to consumer surplus and to the monopoly profit of the firm's shareholders, so that a pure transfer from consumers to shareholders has no reported social cost. The implementation of such a policy raises further problems. It is difficult for the government to estimate demand elasticity and to determine the marginal cost of the monopolist. Of course, it is in the firm's self-interest for the state to err in granting too large a subsidy.[6] Faced with this situation, the firm will seek to "inflate" the subsidy by its actions and in its dealings with the government. To use such a subsidy policy in a discriminating way, the government will most likely need to obtain some information about demand and cost directly, and not through the monopolist. Demand information can be obtained through sampling, although this technique is potentially expensive and may be hard to implement if the monopolist supplies only a few large customers. Cost information is even harder to extract, because the monopolist is, for obvious reasons, reluctant to release accurate estimates of its cost structure.[7] Alternatively, the government can offer the monopolist incentives to reveal its cost structure. For instance, it can reward (in a lump-sum way) the monopolist when the latter charges low prices. The government may thus induce the monopolist to charge a low price when he has a low marginal cost. This type of policy tends to reduce the dead-weight loss.

By considering "sophisticated" incentive schemes, we are moving away from industrial organization proper into regulation—a realm in which a subsidy policy is not optimal any longer, as there exist alternative regulatory schemes that yield lower welfare losses.[8] Why stay away from regulation? First, it is a large field that can hardly be treated in a concise manner; second, its theoretical foundations require some familiarity with the theory of incentives, which would require further developments. The point here is simply that the government's incomplete information about market conditions creates difficulties for intervention. For a correct treatment of the matter, informational asymmetries should be explicitly introduced into the model; then the efficiency of various types of intervention (including commodity taxation) should be analyzed.

1.1.1.4 Second-Order Conditions

Let us return briefly to second-order conditions, which require concavity or quasi-concavity of the objective function. It happens that the profit function of the monopolist is not always concave even if his cost function is convex. The problem is that the revenue function may not be concave—that is, marginal revenue may not be decreasing everywhere. The second derivative of the revenue function $R(p) = p\,D(p)$ is

$$R''(p) = 2\,D'(p) + p\,D''(p).$$

Our assumption that demand is downward sloping ensures that the first term in $R''(p)$ is negative. The second term is nonpositive if demand is linear or, more generally, concave. If demand is convex, the revenue function—and thus the profit function—may not be concave.[9]

*Exercise 1.4** Assume that demand has constant elasticity ε:

$$q = D(p) = p^{-\varepsilon}.$$

Suppose that the cost function is convex. Show that the monopolist's profit function is quasi-concave if $\varepsilon > 1$.

1.1.2 Multiproduct Monopoly

Consider now the case of a multiproduct firm which has monopoly power over all the goods it manufactures. It

6. The "envelope theorem" supports this. Where Π denotes the monopolist's profit, $d\Pi/dt = (p - C')D' = -D$ using the first-order condition.

7. The monopolist may not know the exact cost structure himself, of course. What matters for the argument, however, is simply that the monopolist has private information about the technology.

8. For an analysis of optimal price regulation under asymmetric information about the technology, see the pioneering papers by Baron and Myerson (1982) and Sappington (1982) on the single-product monopolist and that by Sappington (1983) on the multiproduct monopolist. For an analysis of optimal price and cost regulation under asymmetric information about the technology and under moral hazard, see Laffont and Tirole 1986. For surveys of this line of research and further topics (*ex ante* and *ex post* competition, dynamics, etc.), see Baron 1986, Besanko and Sappington 1987, Caillaud et al. 1988, and Sappington and Stiglitz 1987.

9. If the objective function is not concave, the achievement of the social optimum by a subsidy policy becomes still more difficult. (The monopolist's reaction function—the determination of the price, p, depending on the tax, t, imposed—is discontinuous. See Guesnerie and Laffont 1978 for a discussion of this point.)

produces goods $i = 1, \ldots, n$, charges prices $p = (p_1, \ldots, p_n)$, and sells quantities $q = (q_1, \ldots, q_n)$, where $q_i = D_i(p)$ is the demand for good i. The cost of producing the output vector is $C(q_1, \ldots, q_n)$.

In subsection 1.1.1 we analyzed the case of a single-product monopoly or, equivalently, that of a multiproduct monopoly for which demands are independent: $q_i = D_i(p_i)$ (the demand for good i depends only on the price of good i) *and* total cost can be decomposed in n subcosts:

$$C(q_1, \ldots, q_n) = \sum_{i=1}^{n} C_i(q_i)$$

(cost separability). The pricing problem can then be decomposed into n subsidiary pricing problems. Equation 1.1 tells us that the monopolist imposes a higher markup on those goods with a lower elasticity of demand. We will derive a straightforward implication of this result in chapter 3, where we will reinterpret a manufacturer selling the same good in several distinct markets as a multiproduct monopolist. This result represents the simplest form of "Ramsey pricing," which depicts how markups should vary with the elasticities of demand.[10]

More generally, the multiproduct monopolist maximizes

$$\sum_{i=1}^{n} p_i D_i(p) - C(D_1(p), \ldots, D_n(p)).$$

This results in the following formula, which generalizes the equality between marginal revenue and marginal cost:

$$\left(D_i + p_i \frac{\partial D_i}{\partial p_i}\right) + \sum_{j \neq i} p_j \frac{\partial D_j}{\partial p_i} = \sum_j \frac{\partial C}{\partial q_j} \frac{\partial D_j}{\partial p_i} \quad \text{for all } i. \tag{1.6}$$

To analyze this formula, we will consider the two polar cases. (The second-order conditions for equation 1.6 will not be discussed here.) We will phrase the results in terms of the biases that would result if the firm were operated by n independent divisions, each producing one good and maximizing profit on that good.

1.1.2.1 Dependent Demands, Separable Costs

Let us assume that the total cost can be split into n costs:

$$C(q_1, \ldots, q_n) = \sum_{i=1}^{n} C_i(q_i).$$

Then, after some algebraic manipulation, equation 1.6 becomes

$$\frac{p_i - C_i'}{p_i} = \frac{1}{\varepsilon_{ii}} - \sum_{j \neq i} \frac{(p_j - C_j')D_j \varepsilon_{ij}}{R_i \varepsilon_{ii}},$$

where $\varepsilon_{ii} \equiv -(\partial D_i/\partial p_i)(p_i/D_i)$ is the own elasticity of demand (which we will assume to be positive), $\varepsilon_{ij} \equiv -(\partial D_j/\partial p_i)(p_i/D_j)$ is the cross-elasticity of demand for good j with respect to the price of good i, and $R_i \equiv p_i D_i$ is the revenue associated with good i.

First, consider the case of goods that are *substitutes*, i.e., for all j different from i, $\partial D_j/\partial p_i > 0$ or $\varepsilon_{ij} < 0$. In this case, the Lerner index for each good i exceeds the inverse of the own elasticity of demand. This can be explained simply: An increase in the price of good i raises the demand for good j. So, if the firm is decomposed into n divisions, each producing and marketing its own good and maximizing its own revenue ($R_i - C_i$), each division charges too low a price from the point of view of the aggregate firm. The divisions are *de facto* competitors because of the substitutability between their goods. Hence, they must be given incentives to raise their own price (eliminate the externalities between them).

Second, for *complements* ($\partial D_j/\partial p_i < 0$ for all j different from i), the inverse of the own elasticity of demand exceeds the Lerner index for each good. This can easily be understood: A decrease in the price of good i raises the demand for good j. An interesting phenomenon that may arise with complements is that one or several of the goods may be sold below marginal cost (so their Lerner index may be negative), so as to raise the demand for other goods sufficiently. This possibility will be demonstrated in chapter 3.

10. See Ramsey 1927 and Robinson 1933. The result is due to Robinson; the link with Ramsey's contribution was made later. The traditional Ramsey context is that of a multiproduct firm whose objective is the maximization of social welfare rather than profit. Boiteux (1956) constructed a general-equilibrium model in which the social planner, having authority over some public firms, maximizes a social-welfare function subject to the constraint that these firms make non-negative profits. The general Ramsey formula naturally depends on the cross-elasticities of demand and on the elasticities of supply. On this topic see also Baumol and Bradford 1970, Sheshinski 1986, and Brown and Sibley 1986.

These models contain no endogenous explanation of the budget constraint for the public sector.

*Exercise 1.5** A firm has monopoly power on the production of nuts (good 1) and bolts (good 2). Nuts and bolts are perfect complements. Thus, demand depends only on the total price: $D_i(p_1, p_2) = D(p_1 + p_2)$ for all i. Show that equation 1.6 boils down to the monopoly-pricing formula in a single composite market.

An Application: Intertemporal Pricing and Goodwill

Consider a monopoly producer of a single good. This good is sold in two consecutive periods: $t = 1, 2$. At date 1 the demand is $q_1 = D_1(p_1)$ and the production cost is $C_1(q_1)$; at date 2 the demand is $q_2 = D_2(p_2, p_1)$ and the production cost is $C_2(q_2)$. There is a goodwill effect, in that a lower first-period price raises first-period demand and raises second-period demand as well: $\partial D_2/\partial p_1 < 0$.[11] The monopolist's profit is thus

$$p_1 D_1(p_1) - C_1(D_1(p_1))$$
$$+ \delta(p_2 D_2(p_2, p_1) - C_2(D_2(p_2, p_1))),$$

where δ is the discount factor. Letting $\tilde{D}_2 \equiv \delta D_2$ and $\tilde{C}_2 \equiv \delta C_2$, we can rewrite this profit as that of a multiproduct monopoly with interdependent demands. The two economic goods are the single good at the two different dates. From our previous analysis, we can conclude the following: As $\partial D_1/\partial p_2 = 0$, the monopolist charges the monopoly price in the second period conditional on the goodwill accumulated in the first (in other words, the second-period Lerner index is equal to the inverse of the second-period elasticity of demand). In the first period, however, the monopolist charges a price under the static monopoly price, i.e., under the price that maximizes $p_1 D_1(p_1) - C_1(D_1(p_1))$. This is very natural, because the monopolist realizes that a lower price today raises demand tomorrow. He then takes a dynamic perspective by sacrificing some short-run profits to raise future profits.

1.1.2.2 Independent Demands, Dependent Costs

Let us now assume that the demand for good i depends on its price only: $q_i = D_i(p_i)$. Designing a taxomony for dependent costs is a bit more complex than designing one for dependent demands. Indeed, although in the dependent-demand case one can easily envision a set of divisions, each in charge of one product, it may be rather unnatural to separate total cost into several components. Yet there are some cases in which such a decomposition may be reasonable. The application below, also drawn from an intertemporal problem, illustrates this. Before turning to the application, however, the reader would do well to tackle the following exercise.

*Exercise 1.6*** A power plant (or a hotel, or an airline) faces two types of demand: off-peak ($q_1 = D_1(p_1)$) and peak ($q_2 = D_2(p_2)$), where $D_1(p) = \lambda\, D_2(p)$ with $\lambda < 1$. (For simplicity, the demands are independent.) The marginal cost of production is c (as long as capacity is not satiated). The marginal cost of investing one unit of capacity is γ. The same capacity serves peak and off-peak demands.[12]

(i) Show that if off-peak demand is small relative to peak demand (where "small" is to be defined), the monopolist equates marginal revenues to c and $(c + \gamma)$ respectively.

(ii) Treat the case in which off-peak demand is not small. Solve the case in which demands have constant elasticity.

Application: Learning by Doing

In some industries, cost reductions are achieved over time simply because of learning. Through repetition of its activity, the firm gains proficiency. Learning by doing is especially apparent in industrial activity. For example, in the 1920s the commander of Wright-Patterson Air Force Base noted that the number of direct labor hours required to assemble a plane decreased as the total number of

11. The "reduced-form" demand function is not quite satisfactory. The rational foundations of goodwill must be based on the analysis of consumer behavior; see chapter 2.

The astute reader may have noticed that the formalization of the demand curve in period 1 implicitly assumes that the monopolist chooses the two prices sequentially, i.e., does not commit himself about p_2 in period 1. Otherwise, with rational consumers living for two periods, the announcement of a low p_2 in period 1 encourages the consumers to try the good, because they will enjoy a high surplus if they like the good. In this case, D_1 decreases with p_2.

The reader may also feel that calling $D_1(p_1)$ a "first-period demand function" is a bit misleading, because rational consumers consider the possibility of repeat purchase when deciding whether to buy in the first period. Maybe the best way of thinking about this model at the current stage is as follows: One can envision two different groups of consumers at the two dates. The goodwill effect stems from word of mouth between the two generations. The more consumers there are at date 1, the more the generation-2 consumers learn about the characteristics or the existence of the product.

12. Optimal pricing by a firm producing several goods from the same capacity was first studied by Boiteux (1949).

aircraft assembled increased. More recently, learning by doing has been observed in the manufacture of semiconductors and computers.[13]

Consider a single-good monopolist producing at dates $t = 1, 2$. At date t, the demand is $q_t = D_t(p_t)$ (demand can be time dependent). The total cost is $C_1(q_1)$ at date 1 and $C_2(q_2, q_1)$ at date 2, where $\partial C_2/\partial q_1 < 0$. We are thus assuming that a higher production at the beginning lowers the production cost later—i.e., that "practice makes perfect." The monopolist's profit is then

$$p_1 D_1(p_1) - C_1(D_1(p_1)) + \delta(p_2 D_2(p_2) - C_2(D_2(p_2), D_1(p_1))).$$

The maximization of this profit with respect to p_1 and p_2 (i.e., equation 1.6) leads to equality between the marginal revenue and the marginal cost (with respect to current output) in the second period. However, in the first period the marginal revenue is lower than the marginal cost. Thus, the monopolist charges less than the one-period (myopic) monopoly price (the price that maximizes $p_1 D_1(p_1) - C_1(D_1(p_1))$) in the first period; this policy enables him to sell more, which increases production and learning.[14] Put differently, the firm would underproduce in the first period if it were run by two consecutive managers maximizing short-term profit. Exercise 1.7 shows in a slightly more general model that one can obtain a further result if the demand is stationary and costs decrease with experience: that the firm's output grows over time. This result would be very natural if the firms behaved myopically. The decrease in marginal cost due to learning by doing leads to an output expansion. However, a nonmyopic firm also desires to produce much in the first period in order to learn. The result shows that the second effect is dominated by the first.

*Exercise 1.7****[15] The monopoly producer of a single good has a constant unit cost $c(\omega(t))$ at time t, where $\omega(t)$ is the firm's "experience" at that date. (Assume $c > 0$, $c' < 0$, and $\lim_{t\to\infty} c(t) > 0$.) Time is continuous and runs from zero to infinity. Experience accumulates with production: $d\omega(t)/dt = q(t)$, where $q(t)$ is production at date t. (Those who have done the empirical work have assumed, as we do, that production exhibits constant instantaneous returns to scale and that the appropriate measure of experience is cumulative output.) Let $R(q)$ denote the revenue function as a function of quantity (supposing demand is invariant). Assume $R' > 0$ and $R'' < 0$. Let r denote the interest rate. The monopolist's objective function is

$$\int_0^\infty [R(q(t)) - c(\omega(t))q(t)]e^{-rt}dt.$$

(i) Show that at each instant the monopolist sets marginal revenue equal to the average (discounted) unit cost in the future:

$$A(t) = \int_t^\infty c(\omega(s))re^{-r(s-t)}ds.$$

Hint: Consider the current cost and the future savings from changing $q(t)$ slightly.

(ii) Show that output increases over time.

1.1.3 A Durable-Good Monopolist

As was noted above, in the case of a product that gives rise to goodwill the firm ought to take a dynamic perspective and sacrifice some current profits to enhance future profits. Repeat purchases (which will be studied more generally in chapter 2) are an instance of a dynamic link between the periods: Customers are more likely to buy tomorrow if they do so today. Here we investigate another kind of intertemporal link on the demand side—one that is associated with the durability of goods. We now assume that the lifetime of the good exceeds the basic "period" (i.e. length of time between price revisions). In contrast with the goodwill paradigm for nondurable goods, a customer who buys a durable good today is unlikely to buy the same good tomorrow. Thus, the goods offered by the monopolist at two different dates are substitutes rather than complements. (Intertem-

13. One of the very first theoretical analyses of this phenomenon is Arrow 1961.

14. Learning by doing can also be viewed, to some extent, as a form of dynamic increasing returns to scale (see Scherer 1980, chapter 4). In particular, it is easily seen that a competitive equilibrium cannot exist under learning by doing if instantaneous costs exhibit constant returns to scale (see Fudenberg and Tirole 1983). For an existence theorem with convex instantaneous production costs, see Rasmussen 1986.

15. This exercise is drawn from Fudenberg and Tirole 1983.

poral pricing by a durable-good monopolist is studied in much detail in the supplementary section; only the broad issues will be mentioned here.)

As we have seen, a durable-good monopolist creates his own competition. By selling today, he reduces demand tomorrow. As we will see, to sell to the residual demand, the monopolist lowers the price tomorrow. But consumers ought to expect a price decrease and hold back on their purchases today. These rational expectations hurt the monopolist.

Suppose that there are seven consumers. These consumers have "willingnesses to pay" or "valuations" $v = 1, 2, \ldots, 7$, respectively; v represents the present discounted value of the flow of services from the date of purchase on. Each consumer can derive utility from only one unit of the durable good. Assume further that there is no cost to produce the good and that the good is infinitely durable. Time is discrete: $t = 1, 2, \ldots$. The discount factor between periods is δ.

Assume first that the monopolist makes a once-and-for-all offer in the first period. (This thought experiment is meant to describe what happens in the absence of intertemporal effects.) The monopolist then charges the monopoly price, $p^m = 4$, and sells to consumers with valuation 4 to 7. (The monopoly profit is equal to 16.) Now consider the multiperiod model. Suppose that the monopolist charges 4 in period 1, and that consumers with valuations exceeding 4 accept. At the beginning of period 2, the monopolist is left with a residual demand, composed of the consumers with valuations 1 through 3. The monopolist is then tempted to charge a lower second-period price. For instance, if the second period is the last period at which the monopolist sells,[16] he charges the monopoly price corresponding to the residual demand, i.e., 2. Now, consider what happens when the consumers realize in period 1 that the monopolist will have, *ex post*, an incentive to lower the price in period 2. Consumers with high valuations may still accept paying 4 because they are eager to get the good.[17] However, the consumer with valuation 4, for instance, does not buy, because he would get a zero surplus whereas by waiting he could get a positive surplus. Thus, the expectation of future price cuts reduces the demand in period 1.

To solve for the equilibrium, one must find a sequence of prices and consumers' expectations such that the expectations are rational given the firm's behavior and such that the firm's behavior is optimal given the consumers' expectations. The supplementary section explains how to do this. The equilibrium takes the form of a decreasing price sequence. Thus, the monopolist price-discriminates over time: He first charges a high price and sells only to the consumers who are most eager to buy the good. He then cuts his price to reach a slightly less eager clientele, and so on. This type of intertemporal discrimination behavior is often encountered in practice. For instance, books are often introduced in hardcover and then published in paperback form a few months or years later. It is well known that the production-cost difference between a hardcover and a paperback is fairly small. Thus, most of the price differential can be explained by the intertemporal-discrimination model. Another example is the first-run movie feature that is shown later on television, as a home video, on airlines, or at second-run moviehouses.

The flexibility that the monopolist has to adjust his price over time actually hurts him. Indeed, it can be shown that he would be better off if he could *ex ante commit* himself not to haggle, i.e., not to lower the price once high-valuation consumers have bought. ("No haggling" is actually optimal for the monopolist when he can commit. The fixed price is then, of course, the monopoly price.) This is explained by the fact that consumers wait for the day when the monopolist will cut his price. Here, price discrimination is *involuntary*—the firm would, *ex ante*, prefer not to be able to discriminate. Further, it can be shown that the profit loss for the monopolist under noncommitment becomes very high when his price adjustments are frequent; in fact, a conjecture due to Coase (and proved by other researchers; see supplementary section) states that when price adjustments become more and more frequent the monopolist's profit converges to zero. All trade takes place almost instantaneously, at prices

16. This may occur if the monopolist has "outside opportunities" or a fixed cost of production and/or marketing, which induce him to leave the market.

17. In order for the consumer to accept, v must satisfy $v - 4 \geqslant \delta(v - 2)$ or $v \geqslant (4 - 2\delta)/(1 - \delta)$. Such v's exist if the discount factor is not too close to unity—i.e., if consumers are impatient.

close to marginal cost. This result may be extreme, but it illustrates the issue well.

The supplementary section describes these points in a more formal way. It also discusses credibility of commitment and how, in practice, the monopolist can escape the Coase problem to some extent. That section is preceded by an example in which the monopolist produces a good that is recycled. The example is constructed in such a way that the modeling of the buyers' expectations is irrelevant. Thus, it forms a simple introduction to the durable-good problem. It also serves as a background for a brief discussion of monopoly power in the aluminum market.

1.1.4 Learning the Demand Curve

Throughout this chapter—and most of the book—we assume that the monopolist knows his demand curve perfectly. One way of justifying this is to assume that the monopolist conducts market surveys. But such surveys are costly and imperfect, and they always leave some residue of uncertainty about the demand curve. A complementary way of learning demand is to experiment by changing prices over time, which usually allows a better estimation of the demand curve than keeping one's price constant.

There is a small literature on optimal intertemporal pricing by a monopolist in a Bayesian setting.[18] There are few general conclusions about the price path to be followed by the monopolist, which obviously need not be monotonically increasing or decreasing over time. One thing is certain: When setting his price at a given date, the monopolist should not maximize expected current profit given his current (posterior) beliefs about the demand curve. Rather, he should also take into account the value of information thus obtained for future pricing. Aghion, Bolton, and Jullien (1988) and Lazear (1986) have studied models of a stable (nonstochastic) demand curve. Aghion et al. ask whether the monopolist eventually learns his demand curve and therefore charges the full-information monopoly price in the long run. The answer is intuitive. Suppose that it is initially known that the profit function is concave and continuous, but that its exact shape is unknown. Then the monopolist will not stop experimenting before reaching the monopoly price. For assume that he keeps his price constant from some period on. By charging a price slightly different from this price, he learns the slope of the profit function at this price, and he does not affect his expected current profit much. But learning the slope is very valuable for the future, and is therefore desirable. The trick is that, by altering his price by an arbitrarily small amount, the monopolist can make his experimentation costs arbitrarily small and still learn very useful information about the gradient of his objective function. Aghion et al. show that nonconcavities or discontinuities in the profit function may prevent the monopolist from learning his true monopoly price even if the demand curve is deterministic. (See Rothschild 1974 and McLennan 1984 for further results of finite experimentation in models that allow stochastic demands.) For instance, in the case of a nonconcave profit function the previous local-experimentation reasoning shows that the monopolist will eventually reach a local maximum of the profit function. To reach a global maximum, however, would require nonlocal experimentation (large changes in price), which may prove too costly if the discount factor is not sufficiently high. Thus, the monopolist may well settle for incomplete learning, even in the long run. Lazear (1986) looks at a simple case of learning and obtains a few interesting comparative-statics results. For instance, he shows how thin markets (such as that for a mansion) are likely to exhibit a fairly rigid intertemporal pricing pattern, whereas thicker markets (such as that for a very ordinary condominium) will yield larger price changes, the idea being that in a thick market the seller learns more about the demand curve from observing current demand. Similarly, markets with a very diffuse prior probability distribution on the demand curve will also exhibit large price changes.

1.1.5 Inventories

It is assumed throughout most of the book that, in each period, sales originate from current production. In practice, inventories may allow firms to separate production from sales. A sizable and interesting literature treats the

18. The papers usually abstract from other intertemporal pricing considerations mentioned in this chapter (intertemporal substitution by consumers for durable goods, inventories, goodwill, learning by doing) to focus on the learning aspect.

dynamics of quantity and price adjustments, when a firm faces shocks and can smooth its price path and its production path through inventory holdings. For instance, Blinder (1982) analyzes how a monopolist's production, inventories, and price adjust to demand shocks depending on whether these shocks are transitory or permanent. He assumes that in each period the marginal cost of production is increasing with output. Because of the cost convexity, the monopolist prefers a deterministic production to a random one with the same mean. In an intertemporal context, this means that he prefers a stable production to a fluctuating one. Thus, he would like to smooth demand shocks over time; this is exactly what inventories allow him to do. Consider first a transitory (single-period) upward shock in demand. In Blinder's model, in the absence of inventories, the price and the output adjust upward. They still do so in the presence of inventories, but to a lesser extent. The firm can reduce its inventory temporarily and replenish it later. The effect of a single-period increase in demand can thus be spread at the production stage over several periods. A permanent shock in demand cannot be smoothed as much. A high demand today implies a high demand in the future. That is, the marginal cost of production will be high tomorrow as well. Thus, production (as well as price) reacts more to a permanent shock than to a transitory shock.[19]

Another common theme in the literature on inventory behavior is the asymmetric price response to upward and downward shocks. In particular, Reagan (1982; see also Reagan and Weitzman 1982 for the competitive case) assumes that the monopolist can sell only from existing inventories. That is, there is a lag between the use of inputs and the availability of outputs. Current inventories act as a capacity constraint on sales in each period. When demand is high, output is determined entirely by inventories and the firm's price adjusts so as to clear the market (i.e., to satisfy demand). In contrast, when demand is low, the inventory constraint is not binding (sales are lower than inventories). The firm reacts both by choosing a low price and by reducing production. Because of this possibility of quantity adjustment for low demand, but not for high demand, the monopolist's price tends to react more to upward shocks in demand than to downward shocks, as Reagan showed.[20]

1.2 Cost Distortions

In section 1.1 the emphasis was on the distortion on the demand side associated with a monopolist's pricing behavior. Monopoly power can also have perverse effects on the supply side. In particular, for given goods produced by the monopolist and given quantities of those goods to be supplied to the consumers, a monopolist may produce at a higher cost than would a competitive firm. In particular, it has often been suggested that firms in a monopoly situation tend to pay little attention to cost-cutting strategies, engage in slack, and so forth. Hicks (1935), for instance, noted that "the best of all monopoly profits is a quiet life." Machlup (1967) suggested that managerial slack can exist only if product markets are not perfectly competitive. These ideas may seem paradoxical; after all, the monopoly power is on the output side, and it is not easy to figure out why output distortions should have any effect on the cost of producing a given amount of a good.

To investigate this question, we must go back to the concept of cost function—more precisely, to the delegation problem. As was discussed in the chapter on the theory of the firm, a firm's shareholders, who wish to maximize profits, may have a hard time monitoring and controlling the activities of the firm's employees (executives, workers). The latter naturally seek objectives other than profit maximization, and unless the shareholders perfectly observe the technological environment and the employees' behavior (which is highly unrealistic) the firm is likely to engage in "X-inefficiency" (Leibenstein 1966). Indeed, we know from the chapter on the theory of the firm that, whatever the market structure, the firm will generally be able to engage in such inefficiency (i.e., that Machlup's suggestion holds only in very special cases). The question here is how this inefficiency is affected by market power on the product market.

As we saw in that preliminary chapter, shareholders

19. See also the discussion of Blinder in Schutte 1983.

20. For other references on this topic, see Phlips 1980, 1983 and Amihud and Mendelson 1983.

can use the performances of firms with related technologies (or demands) as a yardstick to control the performance of their firm. For instance, the shareholders may be suspicious of their firm's claim that it is facing adverse exogenous conditions when other firms known to face similar supply or demand conditions are doing well. In such a case, the managers' use of the excuse that "times are hard" to conceal slack and justify low profits is not as credible as when there is no other firm with which to compare their firm. This "tournament" idea—basing the incentive structure of one's firm on a comparison with the performance of related firms—does not rely on the existence of competition in the product market; one could, *a priori*, compare the performances of two power plants generating power in two independent regions. But the same type of argument can be made when firms compete on the product market. Thus, it seems natural to base the rewards of Ford managers on the performance of General Motors. It can even be argued that, because the exogenous conditions facing two firms are more likely to be correlated when these two firms are in the same product market, yardstick competition will *in practice* be more useful in industries with several competitors than in product-market monopolies.

This yardstick competition, when it is applicable, may explain why a competitive firm's managers are better controlled by the shareholders than a monopoly's managers.[21] However, Hicks' statement is only half-true: Although a monopoly's managers may engage in more slack (the "quiet life"), they may not benefit from it, because the slack is anticipated. In other words, their "participation" constraint may nevertheless be binding. A lower effort, say, is then offset by lower rewards.

1.3 Rent-Seeking Behavior

Section 1.1 described how monopoly pricing lowers consumer surplus and raises a firm's profit relative to a competitive behavior. The decrease in surplus exceeds the increase in profit by an amount equal to the dead-weight loss. Section 1.2 discussed how, for a given output, a monopoly position may inflate costs. These extra costs add to the dead-weight loss. This section discusses a third distortion associated with monopoly: the wasteful expenses incurred to secure or maintain a monopoly position.

Consider the rent associated with monopoly pricing. Abstracting from the control problem discussed in section 1.2 (so that the cost function can be defined independent of the monitoring technology), one can see that this rent is equal to the monopoly profit represented by the trapezoid *CEFAC* in figure 1.1. It is clear that the existence of this potential rent may lead to rent-seeking behavior. Firms will tend to spend money and exert effort to acquire the monopoly position; once installed in that position, they will tend to keep on spending money and exerting effort to maintain it.

A firm may incur both strategic and administrative expenses to obtain or keep a monopoly position. An example of a strategic expense is the research-and-development cost of obtaining a patent, which secures a monopoly position for the patented product (see chapter 10). Other examples are the accumulating of various forms of capital and the erecting of barriers to entry (chapter 8). Among the administrative expenses are the costs of lobbying and advertising campaigns aimed at influencing the public and its elected representatives ("Our firm is at the service of the consumer") and of legal defense against charges of antitrust violations.

Posner (1975) analyzes an extreme case of rent-seeking behavior in a contest between firms to become a monopolist and concludes that all monopoly rents should be counted in the costs of monopoly. In other words, the actual dead-weight loss is represented in figure 1.1 by the area *CEGFAC*. The two main axioms leading to this conclusion are the following.

(1) *rent dissipation:* The total expenditure by firms to obtain the rent is equal to the amount of the rent.

21. These reflections also imply that the public sector may be more inefficient than the private sector and, at the same time, not introduce any unnecessary inefficiency. The reason is that in many countries the public sector encompasses many industries that are "natural monopolies." Because of the existence of large fixed costs, say, an industry cannot be competitive and is nationalized or regulated (e.g., railroads, the postal service, electricity companies, the telecommunications industry). Hence, the public sector forms a biased sample in terms of product-market power, which naturally leads to more slack. However, the public sector may not be more inefficient *per se*, because many of its firms would have engaged in X-inefficiency anyway had they remained private.

(2) *socially wasteful dissipation:* This expenditure has no socially valuable by-products.[22]

Axiom 1 is the zero-profit free-entry condition. The idea is that entry (or increases in rent-seeking expenses) occurs until the expected rent—i.e., the probability of obtaining the rent times the amount of the rent—equals the rent-seeking cost for each firm. In equilibrium, for instance, ten firms may spend $1 each to have a 10 percent chance of obtaining $10 in rent, in which case the total cost is equal to the rent.

The plausibility of axiom 1 depends on the way the contest is organized. One cannot *a priori* measure rent dissipation without going into the microfoundations of the particular situation.[23] Axiom 1 may not be satisfied for many reasons (see Fisher 1985). First, monopolies can be obtained through luck rather than through foresight. An extreme and somewhat contrived case is that of the patenting of a fortuitous invention. Second, and more important, the contenders may not begin on equal footing; one firm may already have patents, access to particular mineral resources, private information about technology or demand, or incumbency advantages,[24] which will make it the most powerful candidate for the monopoly position. Because a firm's competitors may be less willing to spend money to obtain the monopoly position, it may be able to keep some of the rent. Consider the case in which firms must bid for the privilege of becoming a franchise monopoly. If all the firms are symmetric, the highest bid equals the (common) monopoly rent. With asymmetric bidders, however, the firm with the highest potential rent is able to keep some of the surplus. Third, even with symmetric or almost-symmetric firms, the rent need not be dissipated.[25]

Axiom 2 says that the expenses are socially wasteful. This may be the case when a regulated monopoly position (e.g., the allocation of import franchises) is allocated on the basis of lobbying influence.[26] However, if the same monopoly allocation is allocated through an auction, the expenses are received by the government and thus are not wasteful (in the symmetric case, axiom 1 is satisfied but axiom 2 is violated). There are also intermediate cases in which the expenses are somewhat wasteful. For instance, when air-travel prices and entry on routes were regulated in the United States, airlines competed for customers (the "rent") by offering lavish services. This type of rent-seeking behavior was not entirely wasteful, because customers enjoyed the services. However, the same consumers would have happily traded some of these services for price reductions corresponding to the cuts in services.

An interesting case is that of monopoly rents that are partially transferred to input suppliers. For example, a monopoly rent of 10 may be split into 5 for the owners of the firm and 5 for the workers if the union's bargaining power enables it to appropriate half of the pie. If this represents a simple transfer from the owners to the workers (the labor supply is not altered by the redistribution), the "dissipation" of monopoly profit involves no social loss; the recorded profit (equal to 5) simply underestimates the monopoly rent (equal to 10). However, if the labor supply is affected by the redistribution (for instance, if the existing workers respond to a higher wage by increasing the labor supply), some distortion in allocation is introduced as well.

The bottom line is that rent-seeking behaviors certainly waste some of the monopoly profit. That the monopoly profit may be part of the welfare loss associated with monopoly is a well-taken point. However, we should refrain from drawing any general conclusion about which fraction of the monopoly profit should be counted as a welfare loss. Only a careful description of the rent-seeking game can allow us to give an order of magnitude for this

22. A further assumption is that the inputs used to obtain the rent cannot be bid up (their supply is perfectly elastic). An example where they might be bid up is the case considered below of firms vying for favors from civil servants to obtain a regulated monopoly position. The rent, instead of being dissipated, may then be transferred to the civil servants (through bribes in extreme cases). But, as Krueger (1974) notes, becoming a civil servant in charge of the attribution of these rents may lead to rent-seeking behavior at that stage. A seminal paper on rent-seeking activities is Tullock 1967. A useful discussion can be found in Varian 1987.

23. See the discussion of patent races in chapter 10 of this book.

24. On incumbency advantages and Posner's approach, see Rogerson 1982.

25. See the discussion of preemption games in chapter 8.

26. The analysis here is very vague. What is needed is an equilibrium model in which lobbying activities have influence. Incomplete information ought to be the key to building such a model that would explain why lobbying occurs (information, collusion with decision makers, and so on) and whether lobbying expenses are socially wasteful.

fraction. As the rent-seeking games vary considerably in practice, we are obliged to analyze the issue case by case.

1.4 Concluding Remarks

Monopoly power results in high prices and a dead-weight welfare loss. There may also exist other, more subtle distortions, such as X-inefficiency and dissipation of the monopoly profit. (The next chapter considers a further distortion associated with product selection.)

Although pricing distortions are relatively well understood, cost distortions and rent-seeking behaviors have not yet been mastered by economists. To extend sections 1.2 and 1.3 at a theoretical level and to develop empirical methodologies to measure such distortions are two challenges posed by this chapter.

Some mitigating factors balance these perverse effects of monopoly power to an extent.

First, under increasing returns to scale, production by a single firm is technologically more efficient. Indeed, one of the most often-heard arguments in defense of the monopolization of an industry is that it prevents a wasteful duplication of fixed costs. Williamson (1968) questions the refusal of the U.S. courts to recognize a defense of economies of scale in horizontal-merger cases under the Clayton Act.[27] He argues that, under reasonable assumptions about the elasticity of demand, only a small reduction in fixed costs is necessary to offset the dead-weight loss created by the price increase in the case of a merger.

Second, as Joseph Schumpeter suggested, monopoly may be a necessary condition for a decent amount of research and development. In particular, innovation may require the assignment of monopoly property rights (patents).[28]

One cannot express a view on the merits of monopoly without considering its alternatives (e.g., competition, regulated monopoly) and the ways in which these alternatives may be fostered or obstructed (e.g., subsidies, antitrust proceedings, regulation). The relevance of the various arguments for and against monopoly eventually depends on the relative efficiency of all arrangements[29] and on the information possessed by antitrust, regulatory, and other governmental authorities who promote them. This chapter, like most of this book, is more satisfactory at the positive level (How do firms behave on the product market?) than at the normative one (How should the government correct distortions?). Another challenge offered by this chapter is to develop the normative side.

27. Note, however, that economies of scale are considered in merger cases under the present Department of Justice guidelines.

28. We shall return to this argument in chapter 10.

29. For instance, chapter 6 is mainly concerned with the question of whether the pricing distortion is eliminated by competition.

1.5 Supplementary Section: Durable Goods and Limits on Monopoly Power

In this section, we examine how a durable-good monopolist creates his own future competition. The central theme is that his monopoly power can be eroded by the existence of this nurtured competition. We start with a case of a good that has a short lifetime, after which it can be recycled by a competitive industry. The purchasers of the good dispose of it at the end of the lifetime, and this allows us to ignore the purchasers' expectations about future prices. Although extreme, this case offers a simple and instructive introduction to the subject. In the second example (an "intertemporal price-discrimination problem"), we consider a good that does not depreciate, and we focus on the role of consumers' expectations. This example shows how consumers who anticipate a price decrease restrict their purchases.

1.5.1 Recycling

Consider the case of a monopolist producing a good that is recycled by a competitive industry. As a motivation for this case, recall the famous 1945 U.S. Supreme Court case concerning the Aluminum Company of America (Alcoa). Alcoa had about 90 percent of the primary aluminum market. It was considered a monopoly, and was prohibited from expanding (in that the court ordered that the aluminum plants built by the government during the war not be sold to Alcoa), which led rapidly to a more competitive market in primary aluminum.[30] Some economists opposed the court's decision on the grounds that there already existed an approximately competitive industry, independent from Alcoa, that recycled the aluminum Alcoa produced. If this secondary market for aluminum was taken into account, Alcoa's market share was only 64 percent. In fact, the price charged by Alcoa seemed moderate for a monopolist. Some even suggested that Alcoa's price was close to its marginal cost. Let us examine this argument using a simple model.[31]

Consider discrete time periods labeled $1, 2, \ldots, t$. Suppose that there is a demand function in each period: $q_t = D(p_t)$. This demand corresponds to the consumption demand for aluminum (primary or secondary). Let $p_t = P(q_t)$ be the inverse demand function. The aluminum consumed in period t is either lost or recycled by a competitive industry. Let $x_{t+1} \in [0, 1]$ be the fraction of the aluminum that is recycled. The recycling cost is $C(x_{t+1})$, where C is a convex, increasing function (i.e., the recycling technology exhibits decreasing returns). Moreover, assume that $C(0) = 0$, that $C'(0) = 0$, and that $C(1) = +\infty$ (it is impossible to recoup the entire input). If p_{t+1} is the price of aluminum (primary and secondary) in period $t + 1$, the recycled fraction x_{t+1} is

$$p_{t+1} = C'(x_{t+1})$$

(the competitive recycling industry recycles until its marginal cost equals the price of aluminum). We can then write x_{t+1} as an increasing function of p_{t+1}:

$$x_{t+1} = x(p_{t+1}).$$

Remark We are implicitly assuming that the profits from recycling (which are positive because the recycling cost function is convex) accrue to the recycling industry. In other words, the buyers of aluminum at date t dispose of their used aluminum at date $t + 1$. This assumption allows us to write a per-period demand function $p_t = P(q_t)$. As will be seen below, if the consumers are able to reuse the good or resell it, their demand at date t depends on the price they expect at date $t + 1$, say. The anticipations about future prices must then be modeled. One way of justifying this assumption is to envision a recycling industry composed of a large number of recycling firms so that none of them has any power on the (primary plus secondary) aluminum market (i.e., they are price takers). Each of these firms, however, has a local monopoly power in its geographically delineated input market. Thus, they can charge the monopoly price to obtain the scrapped aluminum; i.e., if the aluminum cannot be used without being recycled, it is obtained for free by the recycling

30. The Supreme Court did not actually hear the Alcoa case. Too many of the justices had conflicts—because the case had taken so long to get through the court system, a majority of the justices had served in the Justice Department while the case was in progress. A special three-judge Appeals Court panel was established to make the final resolution of the case.

31. The following discussion is based on Martin 1982. See also Gaskins 1974 and Swan 1980.