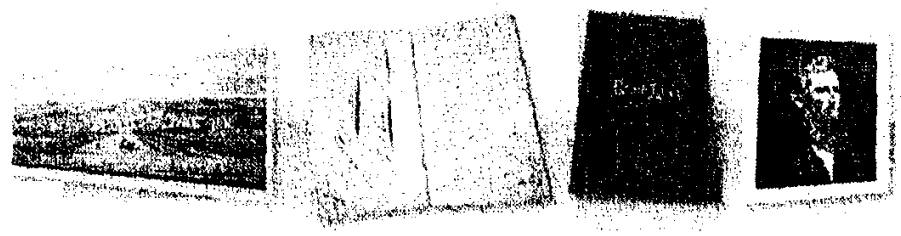

- Trabajo realizado por el equipo de la Biblioteca Digital de la Fundación Universitaria San Pablo-CEU
- Me comprometo a utilizar esta copia privada sin finalidad lucrativa, para fines de investigación y docencia, de acuerdo con el art. 37 del T.R.L.P.I. (Texto Refundido de la Ley de Propiedad Intelectual del 12 abril 1996)

Chapter 3

Competitive Forces

3.1 Introduction

The next step in our analysis of the economic effects of privatization is to consider how competitive and regulatory constraints influence company behavior under public and private ownership. In the previous chapter we examined the discipline on managers of private firms that is sometimes provided by competition in the market for corporate control, and now we turn to the role of competitive forces in product markets. Regulatory constraints will be the subject of chapter 4.

It is important at the outset to underline the obvious fact that *privatization*—the transfer of ownership—and *liberalization*—the opening up of competitive forces—are logically quite distinct concepts. Public ownership does not imply state monopoly, and private ownership does not entail competition. Nevertheless privatization and liberalization are frequently intertwined in policy debate and public perception. The political presentation of privatization policies has emphasized the stimulation of competitive forces, and the privatization of companies such as British Telecom (BT) was indeed accompanied by some important liberalizing measures. Although there is no logical connection between public enterprise and the absence of competition, there are several practical reasons why the two have often gone together.

First, public ownership is one of the main solutions to the problems of market failure that arise in industries where competition is impossible or undesirable, or where major externalities exist. Much of the postwar nationalization in Britain was motivated, at least in part, by the belief that competitive solutions were unsuitable in the markets in question, and there were particular concerns about natural monopoly in industries such as gas, electricity, railways, water, and telecommunications. Although many of these concerns were—and still are—well founded, the danger exists that competitive forces become too neglected, and are assumed to be irrelevant, when perhaps they have a useful role to play. Conditions of demand and technology may change so that yesterday's natural monopoly no longer persists, and the difficulties of efficiently controlling public enterprise may

turn out to be such that competitive forces have significant advantages despite being imperfect.

A second possible link between public enterprise and monopoly has to do with the interests and influence of public sector managers and related civil servants. In the spirit of the "capture" theories of American-style regulation (see section 4.5 below) some would contend that managers of Britain's nationalized industries have successfully resisted the advent of effective competition by their influence on political decision making, institutional inertia, and the support of employees. In contrast, consumers and potential competitors have a relatively ineffective voice. On this view privatization is connected with liberalization insofar as it breaks the anticompetitive institutional blockage. However, as we shall see later, it can equally be argued that managers of companies being privatized—such as BT and British Gas—have been quite successful in limiting the competitive threats that they subsequently face in the private sector.

In any event, it is important to try to understand the costs and benefits of attempted competitive solutions to the market failure problems in the industries in the privatization program, and the aim of this chapter is to present some of the pertinent theories of competition. The recent economics literature on industrial organization has made several important (and sometimes controversial) contributions, and we shall emphasize three themes in particular.

First, and most important, is the role of *potential competition*, which we discuss in section 3.3. The main question here is whether liberalizing conditions of entry into an industry creates entry threats of sufficient power to impel the incumbent firm or firms to behave efficiently and in accordance with consumer preferences. In their theory of contestable markets, Baumol and others examine free entry in its purest form—in which case the answer is in the affirmative. But the economics of strategic entry deterrence and predatory behavior shows that in many circumstances incumbent firms may be able to thwart potential competitors by anticompetitive tactics. The economics of potential competition is therefore highly pertinent to the problems of regulating privatized industries. Do the forces of potential competition operate with such effectiveness as to remove or diminish the need for regulation? Or do policies of liberalization actually require public intervention to ensure that potential competition is effective?

Our second theme is competition as an *incentive mechanism*, which we consider in section 3.4. The problems of monitoring and rewarding managers' effort and efficiency arise under both public and private ownership. The asymmetry of information that exists between the managers of the firm and the government ministry or regulatory agency is at the heart of the matter. The competitive process and the results it delivers (especially in the form of comparisons between rivals' performances) can reveal information in a most economical fashion, and it therefore acts as a natural and useful incentive mechanism. Thus the competitive process provides a spur to *internal* efficiency and the elimination of X-inefficiency, as well as serving as a mechanism conducive to *allocative* efficiency.

Thirdly, in section 3.5, we look specifically at competition issues that arise in connection with *networks* and *vertical relationships*. These have been important in several of the industries where privatization has occurred, of which telecommunications is a good example. If a subscriber in town A wishes to call another in town B, he must use three elements of the telecommunication system—the local networks in A and B, and the long-distance link between the towns. In effect he demands a composite commodity consisting of local and long-distance links. The local networks are monopolized by BT, the dominant firm, but there is scope for competition in the supply of the long-distance link, which BT also offers. Unless constrained from doing so, BT will thwart competition in the long-distance market. This is the crucial *interconnection* issue. It has parallels in the gas and electricity industries, where rival producers rely on the transmission networks of the dominant firms. The economics of vertical relationships therefore bears on U.K. regulatory policy and the wider question of whether vertical separation of a privatized company is appropriate.

Before proceeding to these questions, however, we first outline some relevant welfare economics of actual and potential competition, with particular reference to industries with economies of scale.

3.2 Some Relationships between Competition and Welfare

There is of course a vast literature on the welfare properties of industrial competition, and this is not the place to attempt a survey. Instead we shall briefly discuss two issues of particular relevance to the privatization program. The first is the trade-off between allocative efficiency and scale economies, and the related question of whether free entry can lead to undesirable losses of cost efficiency. The second issue concerns the role of public enterprises in markets where they compete with private firms. We consider how public firms should behave, and what objectives their managers should be given, in those circumstances.

3.2.1 The Trade-Off between Allocative Efficiency and Scale Economies

An argument often advanced for restricting entry into industries with economies of scale is that entry leads to undesirable duplication of fixed costs, and that it is better to have a few (or even just one) large firms than to have more smaller ones. However, the problem is that market power is greater when there are fewer firms, and monopolistic behavior worsens allocative efficiency. The trade-off between allocative efficiency and scale economies is central to many problems in competition policy (e.g. mergers), and precisely the same question—albeit in a much more complex form—arises for example in relation to the licensing of network operators to compete with BT. Two questions can be asked about the trade-off. First, what number of firms maximizes social welfare? Second, does free entry lead to the existence of too few or too many firms at the market equilibrium?

As to the ideal number of firms, the *first-best* outcome when there are economies of scale is to have a single firm operating where price equals marginal cost. This is optimal because the cost of an extra unit of output is equal to consumers' willingness to pay for it. The first-best outcome attains allocative efficiency *and* productive efficiency (no duplication of fixed costs).

However, there are two difficulties with this outcome. First, the firm makes a loss. If lump-sum transfers were feasible, the government could costlessly make good the loss, but otherwise it could not. If such transfers were costly enough, the problem would be to maximize W subject to $\pi \geqslant 0$. The solution to this *second-best* problem would be to have a single firm producing where price equals average cost. The second difficulty is that the firm is unlikely to want to operate where price is equal to marginal (or average) cost, especially if it is privately owned. Regulatory constraint might somehow compel that outcome, but our present concern is with competitive forces unfettered by regulation.

The following simple model illuminates the trade-off between scale economies and allocative efficiency. (For more detailed treatments the reader is referred to Williamson (1968), von Weizsäcker (1980), Perry (1984), Mankiw and Whinston (1986), and Suzumura and Kiyono (1987).)

Suppose that we have an industry consisting of n similar firms each supplying the same homogeneous good. Let $V(Q)$ be consumer utility and $P(Q) = V'(Q)$ be the inverse demand curve for the good, where Q is the sum of the firms' outputs.

Let $q(n)$ be output per firm when there are n firms in the industry, and let $C(q)$ be the cost function of each firm. We shall suppose for simplicity that

social welfare is equal to the sum of consumer and producer surplus. As a function of the number of firms, social welfare can then be written as

$$W(n) = V[nq(n)] - nC[q(n)]. \tag{3.1}$$

We must now specify how the market equilibrium depends on the number of firms, i.e. how q depends on n. For example, we could assume that there is Cournot behavior at market equilibrium. That is to say, firms operate at an equilibrium in which none has any incentive to alter its behavior given the output decisions of its rivals. But rather than specifying any particular oligopoly solution concept, we shall simply assume that $\mathrm{d}q/\mathrm{d}n < 0$, i.e. output per firm falls as the number of firms increases. This assumption holds true in a wide range of circumstances, including Cournot behavior in normal cost and demand conditions. In the model the assumption implies that the market equilibrium number of firms will not be too small, and tends to be excessive (see Mankiw and Whinston, 1986; Suzumura and Kiyono, 1987). The reason is as follows.

If we were allowed to treat the number of firms n as a continuous variable, then the socially optimal number of firms n^*, given market equilibrium behavior, would be such that

$$\frac{\mathrm{d}W(n^*)}{\mathrm{d}n} = 0. \tag{3.2}$$

In fact the number of firms must be an integer, and there is little chance that n^* is an integer. But if $W(n)$ is single-peaked (when n is regarded as a continuous variable) then the optimal number of firms is one of the two integers closest to n^* as defined by (3.2).

The market equilibrium number of firms when there is free entry is given by the condition that further entry would be unprofitable. Let $\pi(n)$ be profit per firm as a function of n. If n was a continuous variable, the equilibrium number of firms $\hat{n}$ would be given by

$$\pi(\hat{n}) = 0. \tag{3.3}$$

If we take the equilibrium number to be one of the integers neighboring $\hat{n}$, we can now compare n^* and $\hat{n}$ to see whether there will be too much or too little entry at the market equilibrium. Differentiating W with respect to n in (3.1) we obtain

$$\frac{\mathrm{d}W}{\mathrm{d}n} = \left(q + n\frac{\mathrm{d}q}{\mathrm{d}n}\right)V' - C - n\frac{\mathrm{d}C}{\mathrm{d}q}\frac{\mathrm{d}q}{\mathrm{d}n}. \tag{3.4}$$

Recalling that $V' = P$ and rearranging terms we obtain

$$\frac{dW}{dn} = n\left(P - \frac{dC}{dq}\right)\frac{dq}{dn} + (Pq - C). \tag{3.5}$$

The second term on the right-hand side of (3.5) is profit per firm, which is zero at free-entry equilibrium. As regards the first term, imperfect competition implies that $P > dC/dq$ (i.e. price exceeds marginal cost), and the assumption that $dq/dn < 0$ therefore implies that $dW/dn < 0$ at $\hat{n}$. It follows that social welfare (as defined) would be increased at the margin by reducing the number of firms below the free-entry equilibrium level. If n is treated as a continuous variable, the optimal number n^* is less than the equilibrium number $\hat{n}$. Thus there is a tendency for excessive entry in the context of the model.

In other words, if output per firm falls with n, then the entry of a new firm is good for consumers, bad for existing firms, and approximately neutral at equilibrium for the new firm, but the benefit to consumers is outweighed by the detriment to existing firms. The result might be different if social welfare gave more weight to consumer surplus than to producer surplus. For example, if no weight is given to producer interests, all that matters is the sign of dQ/dn, which is the response of industry output to changes in n. Under plausible assumptions this is positive, and there is too little entry at the market equilibrium.

Of course it would be foolish to draw any policy conclusions from preliminary analysis of the kind described above. For one thing, policy measures designed to influence the number of firms in the industry (e.g. licensing of entry) must be considered alongside policies (such as regulation) that are intended to affect firm behavior more directly. Secondly, the *threat* of entry can itself have a major influence on industry conduct—n firms surrounded by unsurmountable entry barriers are likely to behave very differently from n firms facing entry threats from potential competitors. It is partly for this reason that the liberalization of entry conditions has been so central to the debate on privatization in Britain. We will return to the important theme of entry threats and potential competition in section 3.3 below, where we consider a number of nonprice aspects of company behavior, including product differentiation, investment, advertising, and research and development (R&D). Thirdly, there are major informational advantages from having more competitors, and incentives for internal efficiency can be enhanced. These important effects are omitted from analysis above, but we discuss them further in section 3.4 below.

3.2.2 Public Enterprise in Competition

We began this chapter by saying that public ownership and competition are perfectly compatible with each other. Indeed, we believe that it would often be advantageous to confront public enterprises with competitive forces even where privatization is not taking place. Legislation in Britain in 1981–1983 opened up the (legal) possibility of some competition with public enterprises in the telecommunications, gas, and electricity industries, and although privatization has since occurred in two of those three sectors, it did not necessarily have to follow liberalization.

There are several reasons why competitive forces might improve industry performance when public enterprise exists. First, internal efficiency might be enhanced by the disciplining effect of competitive threats upon the managers of public firms. It has been said that the greatest of all monopoly profits is the quiet life, and if that is so, there is a case against public monopoly. This case is all the stronger if mechanisms for regulating public enterprise—whether they are explicit or implicit—have serious weaknesses, as both empirical and theoretical evidence suggests. In particular, competitive forces break the "monopoly of information" about industry conditions that may otherwise exist, and which tends to hamper the effectiveness of the regulatory regime. Note that competition need not be an alternative to regulation—it can instead be a useful supplement to it.

A second advantage of opening up the possibility of competition with a dominant public firm is that it creates opportunities for innovation. Rival firms might have the incentive to introduce new products or processes into parts of the public firm's market which the public firm had little incentive to introduce or perhaps had not even thought of. This in turn acts as a spur to the public firm to be more innovative itself, which takes us back to the point in the previous paragraph.

These points will be amplified when we come to consider entry threats and the role of competition as an incentive mechanism. But first we briefly consider some of the questions addressed in the (relatively small) literature on competition in "mixed markets"—i.e. markets in which private and public firms compete.

The simplest case to study is a homogeneous goods duopoly with a welfare-maximizing public firm and a profit-maximizing private firm. If it is assumed that marginal cost is an increasing function of output, the Nash equilibrium in quantities involves the public firm's operating where its marginal cost equals the market price and the private firm's operating where its marginal revenue equals marginal cost. However, if the public firm is a Stackelberg leader (i.e. has first move), it can obtain a better

outcome because it can choose any point on the private firm's reaction function (instead of just the Nash outcome). Welfare is enhanced by the public firm's producing less than its Nash output in order to induce a beneficial expansion in the output of the private firm (given that it has a downward-sloping reaction function).

Beato and Mas-Colell (1984) reverse this move order by giving the public firm the second move. Since the public firm will choose to operate where price equals marginal cost, this is equivalent to the public firm's being committed in advance to a marginal cost (MC) pricing rule. The private firm is then the Stackelberg leader and will optimize given the reaction function (i.e. the MC pricing rule) of its public competitor. Beato and Mas-Colell find that the outcome implied by the MC pricing rule compares favorably with the result obtained when the public firm is the Stackelberg leader. Harris and Wiens (1980) also give second move to the public firm but allow it to commit itself to any strategy in advance. With such powers of commitment, it is best for the public firm to promise to behave in such a way that the competitive price comes about. Then the private firm is induced to act so that price equals marginal cost. However, unlike the MC pricing rule, this commitment by the public firm lacks credibility.

De Fraja and Delbono (1986) consider what objective should be given to the managers of a public firm faced by a number of profit-maximizing competitors. They find that, if the number of firms is large enough, it may be better for the public firm to have profit maximization, rather than welfare maximization, as its objective. Even if there is only one private competitor, Nash equilibrium social welfare is increased if the public firm's managers give some weight to profit in their calculations. The reason is that the Nash equilibrium is then shifted towards the outcome that results when the public firm is the Stackelberg leader (see above).

Finally, Cremer *et al.* (1987) ask whether social welfare would be improved by changing the ownership—or, more precisely, the objective—of some firms in an industry of n firms producing a homogeneous good under increasing returns to scale. They show in their framework that it is often optimal to nationalize a single firm, but that complete nationalization may be desirable.

The literature on competition in mixed markets is not extensive. although we believe that the topic is important. The approaches reported above do not consider the effects of competition upon incentives for internal efficiency (see section 3.4), but they do show how public enterprise can have a positive role in improving allocative efficiency in imperfectly competitive markets. The results of the analysis are sensitive to

assumptions about move order and objectives, but the simple MC pricing rule emerged as an attractive guideline in markets where competitors are few. However, it can be desirable for public enterprises to attach more weight to profits where competition is stronger.

3.3 Potential Competition and Entry Threats

Privatization of firms in competitive markets raises few problems for industrial policy because competition between firms in the market effectively regulates company behavior and provides reasonably good incentives for internal and allocative efficiency. But the mechanism of competition between existing firms cannot be relied upon if the incumbent firm being privatized has a preponderant market share and if it is not being split into units that will subsequently compete with each other. It is then only the entry and growth of new rivals—or at any rate the threat of it—that can provide competitive disciplines for the incumbent firm. If these do not exist, regulation of the incumbent's pricing behavior will be needed to check monopolistic abuse. Regulation in the form of policies to prevent anticompetitive behavior may be needed to make entry threats properly effective.

In this section we examine the theory of contestable markets and recent theories of strategic entry deterrence and predatory behavior. Two of the main questions that we have in mind were stated in the introduction to this chapter. Do the forces of potential competition operate with such effectiveness as to remove or diminish the need for regulation? Or do policies of liberalization actually require regulation to ensure that potential competition is effective?

3.3.1 The Theory of Contestable Markets

In a contestable market the threat of potential competition is at its most potent. The incumbent firm or firms are compelled to meet consumers' wishes with maximum efficiency, for otherwise new entrants will simply take their business away. The theory of contestable markets proposed by Baumol and his colleagues was developed in a series of articles beginning in the late 1970s and was expounded in the book by Baumol *et al.* (1982) and in Baumol's (1982) presidential address to the American Economic Association. The theory has provoked great academic controversy and has featured prominently in public debate on regulation and antitrust. An evaluation of the theory is essential for an understanding of the role of competitive forces in constraining the behavior of apparently dominant

firms, and hence it is at the heart of much debate on privatization policy.

One of the main contributions of contestability theory has been to illuminate the economics of multiproduct industries and to introduce important cost concepts (such as economies of scope and trans-ray convexity). However, we shall begin by focusing on its contribution to the theory of potential competition in the context of an industry supplying a single product. Our plan is first to define what is meant by "contestability," then to infer what properties must hold in a contestable market, and finally to discuss the implications of the theory for practical policy making in the light of the criticisms that have been made of it. More extensive discussions of contestability theory are provided by Baumol and Willig (1986), Brock (1983), Schwartz (1986), Sharkey (1982), Shepherd (1984), and Spence (1983).

A contestable market is one in which existing firms are vulnerable to hit-and-run entry. All firms—actual and potential—have access to the same production methods, and hence their *cost functions are identical.* Moreover, entry involves *no sunk costs*: a firm can enter the market without making irrecoverable expenditures, and so there are no barriers to exit. Therefore

> "Even a very transient profit opportunity need not be neglected by a potential entrant, for he can go in, and, before prices change, collect his gains and then depart without cost, should the climate grow hostile." (Baumol, 1982, p. 4.)

The distinction here between sunk costs and fixed costs should be noted. Fixed costs are costs that do not vary with the level of output; they are not always sunk. For example, a firm that leases a photocopier for a short period of time incurs a fixed cost but not sunk costs. It is perfectly possible to have a contestable market with high fixed costs, so long as there are no sunk costs, and indeed a natural monopoly market can in principle be contestable and vulnerable to hit-and-run entry.

If equilibrium exists in a contestable market, the following properties hold.

(i) Price P equals average cost AC. That is to say, all firms make zero profits. If $P >$ AC there would be an incentive for a new firm to enter the market and displace an incumbent firm by slightly shading its price. If $P <$ AC a firm would do better to leave the market, which can be done costlessly in view of the assumption that sunk costs are absent. It follows that $P =$ AC at equilibrium.

(ii) There is no inefficiency in production. If an existing firm was producing its output at more than minimum cost, rivals would find it profitable to enter the market, take the other firm's business, and produce more efficiently.

(iii) Price is at least as great as marginal cost. If some firm was operating where $P <$ MC, a new firm could profitably enter the market producing slightly less than the other firm and take that firm's business by shading price infinitesimally. This strategy would be profitable because the marginal revenue lost by reducing output, which is approximately equal to P, is less than the saving of MC.

(iv) If there are two or more firms at equilibrium in the market, then price is no greater than MC. If one of the existing firms was operating where $P >$ MC, a new firm could come in and produce slightly more than that firm. The entrant's marginal revenue MR ($= P$) would exceed his MC, and so he would make more profit than the firm displaced. The entrant could (and would) also take some of the business of other incumbent firms by an arbitrarily small price reduction. This argument does not go through if initially there is only one firm in the market, because then the only way for a firm to increase sales volume is to lower price. Figure 3.1 shows how $P >$ MC is possible at equilibrium in a market characterized by natural monopoly cost conditions. (Note that (iii) and (iv) together imply that $P =$ MC if there are two or more firms at equilibrium.)

(v) The number and configuration of firms at equilibrium is such that industry output is produced at minimum cost. Therefore there is productive efficiency at the industry level as well as the level of the firm (see

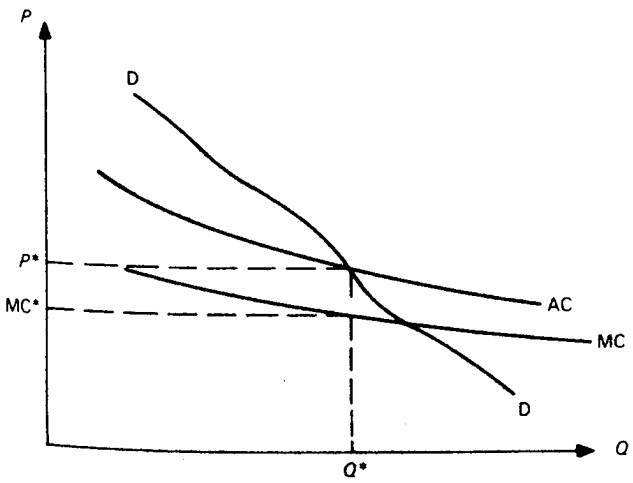

Figure 3.1 Equilibrium with price greater than marginal cost

(ii)). If this property did not hold, a new entrant producing at minimum average cost could profitably enter the industry by taking business from firms producing at higher cost levels. This result about productive efficiency at the industry level implies that market structure is endogenous in a contestable market.

Results (i) through (v) above were stated in the context of a single-product industry in order to show as clearly as possible why they hold. The arguments generalize naturally to multiproduct industries, where (i) through (v) also hold. We shall state two more results about equilibrium in contestable multiproduct industries.

(vi) There will be no cross-subsidies between products. This follows from the fact that price must be at least as great as marginal cost for all products in the industry. Otherwise a new firm could profitably enter producing slightly less of the product in question than some existing firm. This profit opportunity is ruled out only if cross-subsidy is absent.

(vii) Under certain assumptions about cost conditions (trans-ray convexity and declining ray average costs), a natural monopolist deters entry by charging Ramsey prices. Ramsey prices maximize welfare subject to a minimum profit constraint. (The basic reason for this result can be outlined as follows. If firm B wants to take the business of firm A, it will seek to maximize its profits subject to the constraint of offering consumers a more attractive deal than firm A. The dual of this constrained optimization problem is to maximize consumer wellbeing subject to a minimum profit constraint. That is essentially the same as the Ramsey optimization problem.)

By now it will be evident that the properties that hold at equilibrium in a contestable market are highly desirable by the yardsticks of traditional welfare economics. There is internal efficiency, because firms—individually and collectively—are compelled by the entry threat to produce at minimum cost, and there is allocative efficiency because $P = \text{MC}$, except possibly in a natural monopoly (in which case pricing is optimal subject to the nonnegative profit constraint).

The theory has been proposed as a competitive benchmark that is in a sense more general than the traditional textbook model of perfect competition. Instead of a huge number of very small firms being the hallmark of competition, it is entry threats that make for competition in a contestable market, even though only one or two firms might exist at equilibrium. Indeed market structure is determined endogenously by the

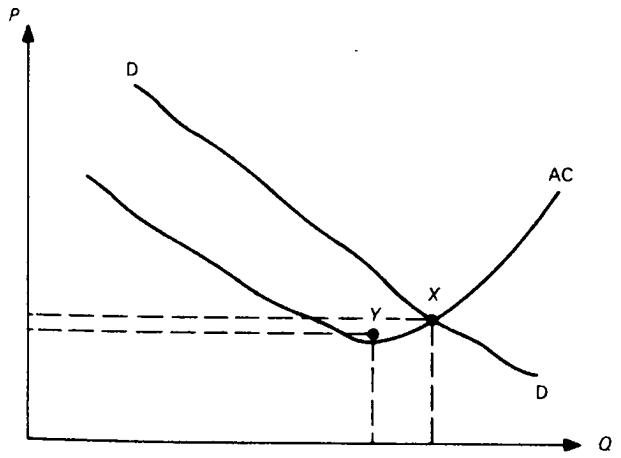

Figure 3.2 An example of nonsustainable natural monopoly

fundamental parameters of technology and demand. It is sunk costs (as opposed to fixed costs) that are of central importance—a point which has implications for policy.

The desirable properties stated above rest on the assumption that equilibrium in a contestable market exists. Unfortunately it need not do so, as the following examples show. Figure 3.2 depicts a natural monopoly industry—cost and demand conditions are such that single-firm production is most efficient. However, the price–output combination X is not an equilibrium, because it is vulnerable to entry by a firm operating at point Y. The entrant can undercut the firm and still make a profit. This is an example of the *nonsustainability* of natural monopoly. A generalization of this example is an industry with textbook U-shaped average cost curves. If P^* is the minimum average cost and if q^* is the associated level of output, equilibrium exists in the contestable market only if demand at price P^*, denoted by $D(P^*)$, is an integer multiple of q^*. The chance of this happening is effectively zero. This problem can be avoided by supposing that cost curves are flat-bottomed, but that is perhaps a rather artificial step to take.

A second important reason for nonexistence can arise in multiproduct industries (see Faulhaber, 1975). Suppose that there are three products in the industry and that there is demand for one unit of each product. Let $C(k)$ be the cost of producing k of the products ($k = 1, 2,$ or 3). It is quite possible that

$$\frac{3}{2}C(2) < C(3) < C(2) + C(1), \tag{3.6}$$

for example if $C(1) = 6$, $C(2) = 8$, and $C(3) = 13$. There is a natural monopoly because the most efficient way to produce the three products is for one firm to produce all of them. But that state of affairs is vulnerable to entry by a firm producing two out of the three products. A three-product firm could cover its costs provided that the sum of the prices of the products was no less than 13. In that case the sum of the prices of the two most expensive products could not be less than 8½. But then a new firm supplying only two products could undercut the first firm and still make a profit.

This multiproduct example of the nonsustainability of natural monopoly is sometimes associated with "cream skimming" and "destructive competition." Cream skimming occurs when a rival takes a profitable component of the incumbent's business, leaving the incumbent with a loss on the rest. The (alleged) undesirability of cream skimming is sometimes offered as a justification for restrictions on entry.

Although the nonsustainability of natural monopoly is a theoretical possibility, we would urge that arguments for restrictions on entry based on grounds of cream skimming or destructive competition should be treated with the greatest caution. First, there is a broad class of natural monopoly cost conditions that imply sustainability (see Sharkey (1982) for a full discussion). Secondly, natural monopoly conditions are not as widespread as is often claimed. Thirdly, the vested interests of incumbent firms are such that they should bear a heavy burden of proof when arguing for restrictions on entry. Fourthly, the sustainability problem—and existence problems in general—are of little or no consequence in industries that are not contestable.

At this point it is appropriate to address some of the main criticisms (apart from the existence problems just discussed) that have been leveled against the theory. We shall focus on claims that the theory has minimal relevance to real-world industries and that it is therefore devoid of practical policy implications. If the theory can withstand such criticisms, it has very important consequences, for it suggests that opening industries to potential competition (even industries that are highly concentrated in terms of current market shares) does away with the need for regulatory policies because the force of potential competition itself produces desirable results.

The first major criticism is that the theory assumes an unnatural sequence of events when entry occurs. In particular, it assumes that the entrant can establish itself, undercut the existing firm(s) on price, and take as much of their business as it likes, all before the existing firms respond

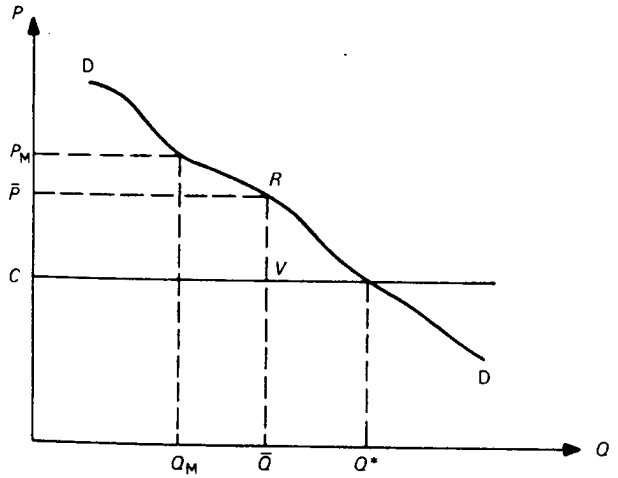

Figure 3.3 Nonrobustness: illustrative market conditions

by lowering their own prices. The criticism is that this implicit assumption—that entry can occur more quickly than price changes by existing firms—is precisely the *reverse* of what it is realistic to suppose.

The second criticism is that the assumption of zero sunk costs is *nonrobust* (see Schwartz and Reynolds, 1983; Dasgupta and Stiglitz, 1985). Proponents of contestability theory can reply to the objection that the assumption of zero sunk costs is never strictly *true* by saying that no theory can be more than an approximation to the truth. But the claim of nonrobustness is more damaging, because it implies that the theory is not useful even as an approximation to reality.

Both criticisms can be illustrated with the aid of figure 3.3. The diagram concerns a market for a homogeneous product. The demand curve is labeled DD, and the constant level of unit costs is denoted by C. Social welfare would be maximized when price $P = C$ and output is Q^*. Suppose the entry into the industry involves sunk cost S, which can be regarded as an entry fee. Suppose that firm A is currently in the market and that firm B is contemplating entry. Let T_A be the time that it would take for firm A to respond to B's decision to enter, and let T_B be the time that it would take B to set up at full scale. Thus, if $T_A > T_B$, then B can set up before A responds to B's entry. But if $T_A < T_B$, then A responds to B's entry before B has set up at full scale. Suppose that, once A has responded to B's entry, competition between them rapidly drives price down to the cost level C. Note that this does not involve firm A's reducing its price below the level of costs. Define T as follows:

$$T = \begin{cases} T_A - T_B \text{ if } T_A > T_B \\ 0 \text{ otherwise.} \end{cases} \quad (3.7)$$

Thus, T is the amount of time that B would have at full scale before A's response drove prices down to level C.

Suppose that A is currently charging price $\bar{P}$. Should B enter the market? If he entered and charged a price just below $\bar{P}$, he would earn a profit flow of almost

$$\pi(\bar{P}) = (\bar{P} - C)D(\bar{P}), \quad (3.8)$$

i.e. the rectangle $\bar{P}RVC$. This profit flow would last for T units of time, at which point profits would be driven down to zero. Entry is worthwhile if this profit flow covers the sunk cost S, i.e. if and only if

$$T\pi(\bar{P}) > S. \quad (3.9)$$

Now, if sunk costs are zero ($S = 0$) and the entrant can set up before the incumbent firm responds ($T > 0$), then any price P in excess of C is vulnerable to entry. This is the case of a contestable market.

If, however, the incumbent can react faster than (or just as fast as) the entrant can establish himself at full scale, then $T = 0$. This is usually the realistic case. Now, if sunk costs are positive—no matter how tiny—then $S > 0$, and from (3.9) it follows that no price would attract the entry of firm B. Therefore A could charge the monopoly price P_M without inducing entry. In this case, even though sunk costs might be miniscule, the threat of entry is nonexistent, and the incumbent can behave as a monopolist with impunity.

The remaining case is that in which $T > 0$ and $S > 0$. The incumbent will deter entry if he charges a price P such that

$$\pi(P) \leqslant S/T. \quad (3.10)$$

Even if S is rather small, this constraint on the behavior of the incumbent firm will be very weak if T is small. It may well be that the monopoly price P_M satisfies (3.10). S must be infinitesimal if the entry threat is to promote anything remotely like competitive behavior in this model.

In short, any departure from the strict assumptions of contestability theory ($S = 0$, $T > 0$) leads to an outcome that may be radically different from the outcome predicted by the theory.

Contestability theory is important insofar as it provides a valuable framework for thinking about a number of problems of industrial organization: it has led to several important lines of research on multiproduct industries; it has introduced important concepts and analytical tools; it allows for the endogenous determination of market structure; it highlights the importance of sunk costs; and it underlines the beneficial role that potential competition can play in improving industrial performance.

However, we have grave doubts about the empirical applicability of contestability theory, especially to major industries of the kind being privatized in Britain. The entrenched dominant positions of companies like BT and British Gas, and the inevitably sunk nature of many costs, are such that entry threats *on their own* cannot remotely be expected to compel the dominant firm to behave benignly. Entry threats are usually very desirable, but they will be impotent unless supplemented by effective competition policies that strike at anticompetitive behavior against potential entrants by incumbent firms. Moreover, there are a number of markets and submarkets where the threat of entry simply does not exist and cannot be made to exist. Direct regulation of the dominant firm is then required.

3.3.2 Strategic Entry Deterrence

In this section we consider in more detail how an incumbent firm with market power might act so as to stifle potential competition. This is a central question for privatization policies in the U.K., because major enterprises such as BT and British Gas have entered the private sector with positions of great dominance, and it is not clear to what extent measures to liberalize entry into their industries will promote truly effective competition. This depends partly upon the nature of policies to guard against anticompetitive behavior designed to thwart potential entrants. A theoretical perspective on this issue is provided by the recent literature on *strategic entry deterrence* (see Vickers (1985a) for a survey). In this section we outline some of the main elements of that literature, and we consider the problem of designing policies against predatory behavior. We postpone until later in the chapter our discussion of competition and entry deterrence in networks and vertically related markets.

The key to strategic entry deterrence by an incumbent dominant firm is to make credible the threat of responding to entry—for example by very aggressive pricing—in such a way that the rival would regret having entered the market. The threat is credible if and only if it is believed to be in the interests of the incumbent to carry it out when entry actually occurs. In order to assess the danger of strategic entry deterrence thwarting potential competition and to consider appropriate policy measures, we must focus precisely on this question of credibility. One school of thought

(often associated with the University of Chicago) has it that predatory pricing is irrational, and therefore unlikely to occur, because it hurts the incumbent as much as the entrant. In that case threats of predatory pricing are not credible. But others have recently contended that predatory and entry-deterring behavior of various kinds is entirely possible, and that policy measures are vital if dominant firms are not to choke off the threat of potential competition.

For simplicity, we shall focus on the case where one incumbent firm is seeking to deter the entry of one potential rival. The entry decision of the rival depends upon his beliefs as to the likely profitability of being in the market. Entry will occur if and only if the expected profits exceed the expected costs of entry. How can the incumbent influence those beliefs in such a way as to deter entry? We shall address this question in two steps. First, we shall suppose that each firm is fully informed about the behavior, opportunities, and motivation of the other. Secondly, we shall relax these assumptions about the information available to the firms, and examine the roles of signaling and reputation in entry deterrence.

The best known traditional analysis of entry deterrence is the Bain–Sylos limit pricing model (see Hay and Morris, 1979, pp. 185–190). In that analysis it is assumed to be common belief that the incumbent would not change its pre-entry output level in the event of entry. Although it makes for tractability, that assumption is not credible, because the incumbent almost certainly would wish to change its behavior when faced with a rival. More generally, as Friedman (1979) has observed, in a game of complete and perfect information, and with no intertemporal interdependences of cost or demand conditions, the incumbent's output level before the entry decision ought to make no difference to the rival's assessment of the profitability of entry. Once entry occurs, a new game begins, and the parameters of that game are independent of previous behavior. This suggests that flexible instruments such as price or quantity are less likely to be the means of entry deterrence than instruments that have a more lasting effect upon cost or demand conditions. The key is for the incumbent to *commit* himself to a course of conduct that would be detrimental to an entrant. A large amount of literature, from which but a few items will be mentioned, is devoted to this theme. Figure 3.4 sketches a simple schema.

First, the incumbent chooses the level of some strategic variable K. Numerous interpretations can be given to K, but for the moment regard it as the incumbent's capacity level. If the rival chooses not to enter, he gets zero and the incumbent gets $I^0(K)$, as shown at the foot of the right-hand branch. Note that the incumbent's payoff depends on K even if entry does

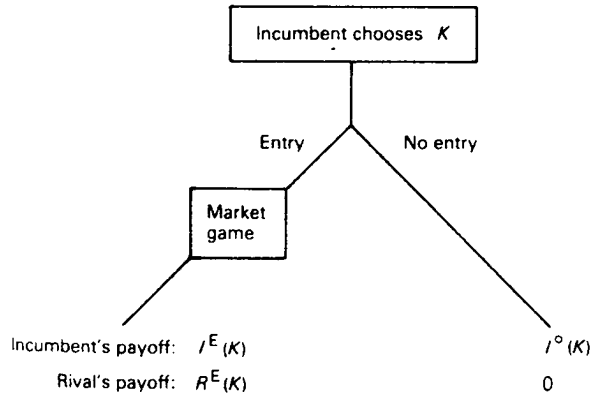

Figure 3.4 Strategic entry deterrence

not occur. If entry does take place, a duopoly exists. Without examining the details of the duopolist's interactions, let us assume that the upshot of the "market game" between them is that the incumbent gets $I^E(K)$ and the rival gets $R^E(K)$ in the event of entry. The schema is broad enough to allow for the rival also to choose some strategic variable after the incumbent's choice of K. This would be included in the black box of the market game.

The incumbent's choice of K deters entry if $R^E(K) < 0$. It may be that entry is deterred even by the level of K that would have been chosen by a pure monopolist facing no threat of entry. Then entry is said to be blockaded. Or it may be that the incumbent does better to permit entry than to deter it. But here our concern is with the remaining case, in which strategic entry deterrence is optimal for the incumbent.

What are the likely instruments of entry deterrence? It is useful to distinguish between those that affect costs (the incumbent's and/or the rival's) and those that affect demand. As to the former, Dixit (1980) showed how the incumbent's choice of *capacity* could deter entry, but in Dixit's model excess (in the sense of idle) capacity is not observed. More generally, K can be interpreted as the incumbent's level of *capital input*. Strategic entry deterrence commonly implies overcapitalization, in the sense that the output eventually produced by the incumbent could have been produced more efficiently with a lower level of capital and a correspondingly higher level of variable factors of production (see Spence, 1977). The same holds when K is interpreted as the incumbent's cost-reducing R&D expenditure. In all these examples, the incumbent's commitment of high K promises that he will supply a high output level, or charge a low price, in the market game.

The choice of K is therefore offputting to the rival, not because of its direct effects, but because of its indirect influence upon the outcome of the market game.

In criticism of the Bain–Sylos analysis, it was stated above that the incumbent's pre-entry output would not affect the rival's entry decision if there were no intertemporal interdependences of cost or demand. But such interdependences do hold if the *experience curve* effect operates—i.e. if a firm's cost level is a declining function of its cumulative output. In that case it is possible for the incumbent's choice of output to deter entry strategically (see Spence, 1981; Fudenberg and Tirole, 1983).

As well as the incumbent lowering his own costs, there may be ways for him to *raise the rival's cost* (see Salop and Scheffman, 1983). For example, by setting high wage rates in the industry, the incumbent increases his own costs and those of an entrant. The direct effect of this upon the incumbent is unfavorable, but if the indirect effect is to deter the rival's entry then the ploy may well be beneficial to him in overall terms.

Another way in which entry might be deterred is for the incumbent to deny the rival access to technology that would allow him to compete. Gilbert and Newbery (1982) examine *pre-emptive patenting*—the acquisition of a patent by an incumbent firm with the purpose of denying the patent, and hence an entry opportunity, to a potential rival. An important factor here is that the incumbent's incentive to win the patent is likely to exceed the rival's incentive, even if the patent is for a technology inferior to that already enjoyed by the incumbent. This is because the incumbent's monopoly persists if he denies entry to the rival, whereas competition, which is less profitable than monopoly, occurs in the event of entry. This result has an important bearing on the question of the persistence of monopoly, for the advantage of the incumbent arises from his *strategic position*; intrinsically the incumbent may be no different from the rival. The result also offers an explanation of the phenomenon of sleeping (i.e. unused) patents, because the pre-emption result does not depend upon the patent's being for a technology that is superior to the incumbent's existing technology.

Turning from the cost side to the demand side, there are further ways in which an incumbent firm can make the prospect of entry unattractive for a rival firm. In some circumstances, strategic *advertising* deters entry, although in Schmalensee's (1983) exploration of advertising and entry deterrence it emerged that low advertising was the way to deter entry. The reason was that high advertising would cause the incumbent to have a higher price in the market game. For the rival, the favorable latter effect would outweigh the disadvantage of high advertising and make entry more attractive. *Brand proliferation*—the introduction of numerous new products—can also serve to deter entry (see Schmalensee, 1978). To use the locational analogy common in the analysis of product differentiation, brand proliferation fills up product space in such a way that there are no remaining slots or niches for profitable entry.

Product differentiation was one of Bain's three sources of barriers to entry. Although advertising is often regarded as a measure of product differentiation, Bain did not see it as the heart of the problem. Schmalensee (1982) shows how buyers' uncertainty about the quality of new brands can give established (or pioneering) brands an advantage in a differentiated market. The new brand would have to be priced substantially below the existing brand to induce consumers to experiment with it; part of the cost of the experiment is the loss of surplus currently being enjoyed on the existing brand. (Farrell (1986) develops this point in his analysis of moral hazard as a barrier to entry.) This is again an example where the incumbent has strategic advantage solely because of already being in the market.

To summarize so far, there are numerous ways in which an incumbent firm can influence cost and/or demand conditions by strategic investments in such a way as to discourage entry into his market. Bulow *et al.* (1985) and Fudenberg and Tirole (1984) have examined in general terms various types of entry deterrence. Returning to the schema above, we can ask whether overinvestment or underinvestment in the strategic variable K deters entry. The answer depends of course upon the specification of the market game. In some instances (such as excess capacity deterring entry), the incumbent deters entry by being "large." Fudenberg and Tirole call this the "top dog effect." In other instances, the incumbent deters entry by being "small," and thereby promises an aggressive response in the event of entry. This is the "lean and hungry look." It may be better for the incumbent to accommodate rather than deter entry. Then he will act strategically to influence the nature of entry, by being either "large" (a "fat cat") or "small" (the "puppy dog ploy"). The aim of all these strategic moves is to cause the rival to choose to act more favorably for the incumbent than he would otherwise do.

So far we have supposed that firms are fully informed about each other's opportunities and motivation, but we now turn to the second step in the analysis of strategic entry deterrence by relaxing this assumption. Milgrom and Roberts (1982a) have shown how limit pricing can be used to deter entry when the potential entrant is uncertain as to the cost level of the incumbent firm. The rival's expectations concerning that cost level—and

hence his entry decision—are influenced by the price charged by the incumbent before the entry decision is made. Therefore the pre-entry price can act as a *signal* of the incumbent's efficiency. An incumbent with low costs would like to signal that fact, because the potential rival would then be more reluctant to enter his market. By the same token, an inefficient incumbent would like to masquerade as a low-cost firm in order to make entry less likely. This incentive to signal in an uncertain environment means that the incumbent's pricing can be used as an instrument of entry deterrence. Note that consumers benefit from this kind of limit pricing insofar as it tends to lower price. Other interesting issues arise when there is uncertainty about demand as well as the incumbent's cost level. Then a low pre-entry price might signal either low demand or low costs (see Matthews and Mirman, 1983).

Milgrom and Roberts (1982b) and Kreps and Wilson (1982) have explored another context in which uncertainty about the incumbent plays a role in entry deterrence, namely in connection with *predatory pricing* (see Vickers (1985b) for a survey). They examine a game due to Selten (1978) in which an incumbent firm—a chain store—is threatened by entry in each of a number of towns. Intuitively, we would expect that the incumbent would fight entry if challenged in a town early in the sequence, in order to deter later entrants. However, this is not so if the entrants have complete information about the opportunities and motivation of the incumbent, because in that case he would never fight entry. This is because it is common knowledge that he would not fight in the last town, and so he would have no reason to fight in the last but one and so on. But if the assumption of complete information is very slightly relaxed—so that there is a possibility that the incumbent is somehow committed to fighting—then even an uncommitted incumbent would (rationally) fight entry in early towns to keep up the *reputation* of possibly being a committed fighter. This reputation effect is very powerful, in the sense that a very small amount of incomplete information can make it rational to fight on a large number of occasions.

Another, quite different, context in which predatory pricing is rational has been examined by Benoit (1985) in a paper which also covers the case of imperfect information. He supposes that the entrant's financial resources are such that it would go into bankruptcy after a predatory pricing war of some finite length (say T periods). In other words, its "war chest" is not unlimited. The incumbent's war chest is assumed to be larger than that of the entrant. Assume now that the incumbent would prefer (i) *one* period of predatory fighting followed by perpetual dominance thereafter to (ii) duopolistic coexistence forever. It now follows that it would be rational for the incumbent to engage in predatory pricing rather than to accommodate entry, no matter how large is T. Given the incumbent's preference for (i) over (ii), it would clearly fight to drive out the entrant if $T = 1$. Now, if $T = 2$, both firms know that the entrant would leave the market voluntarily if it were forced into a position where it could survive only one more period before going bankrupt, because it has already been established that the incumbent would force the entrant out otherwise. So the incumbent would find it rational to fight if $T = 2$. The argument can be extended inductively for all T. If the entrant would exit voluntarily at $T = t$, then the incumbent would rationally fight at $T = t+1$, in which case the entrant would exit voluntarily at $T = t+1$ and so on. Here the logic of credible threats works powerfully against entry.

It is now time to draw together some conclusions on strategic entry deterrence. First, there are many circumstances in which a dominant incumbent can act so as to make entry appear an unattractive prospect. Secondly, nonprice behavior is important in this regard, particularly as it affects future cost and demand conditions. Thirdly, it is far too hopeful to suppose that predatory pricing is irrational and that threats of it would therefore be disregarded. In short, there is often ample scope for a dominant firm to undermine "freedom" of entry, and it follows that policies are required to make that freedom effective. In this section we have suggested some of the problems which those policies must address, but it would be foolish to advance general policy prescriptions without taking account of the individual circumstances of particular industries. We will therefore return to the application of the economic principles of strategic entry deterrence and predatory behavior, outlined in this section, later in the book.

3.4 Competition as an Incentive Mechanism

One of the main virtues of competition emphasized by proponents of privatization and liberalization is its role as a mechanism that stimulates *internal* efficiency. Indeed, despite the emphasis of textbook competition theory on *allocative* efficiency in a static environment, it has long been recognized that much of the effect of competition on welfare is due to its role as an incentive system and discovery mechanism in a world of imperfect information (see for example the work of Hayek (1945) and Leibenstein (1966) on X-inefficiency). If competition promotes internal as well as allocative efficiency, then it is doubly beneficial in terms of social welfare.

We considered in the previous chapter whether or not competition in the market for corporate control stimulates internal efficiency by reducing managerial slack, and now we turn to product market competition. Recent work, notably that of Hart (1983), shows how product market competition can influence the nature of incentive contracts between the shareholders and managers of a firm. Suppose that the unit costs of a firm depend on managerial effort and upon some exogenous cost characteristic of the firm. If there is a high degree of correlation between the exogenous cost characteristics of firms in the industry, then competition between them reveals information about managers' effort rates. Being better informed about managerial effort, shareholders can make managerial reward more sensitive to effort, and slack and X-inefficiency can therefore be diminished. Reward can be based, a least in part, upon *relative* performance, and so managers are effectively in competition with each other.

Willig (1985) has performed comparative statics analysis of a principal–agent model to illuminate the effect of product market structure on managerial incentives and behavior. As above shareholders can observe the level of unit costs c, but cannot infer managerial effort from their knowledge of c. The asymmetry of information means that shareholders must reckon with an incentive compatibility constraint when designing an incentive mechanism for managers. That is to say, they must allow for the fact that the managers will behave in a self-interested manner given the incentives that they face. This constraint prevents the attainment of the first-best outcome, and X-inefficiency (in a precise sense) results. Managers gain from this situation, relative to the first-best outcome. Willig does not explicitly examine competition between firms. Instead he introduces a parameter b that is intended to capture the nature of product market competition in a "reduced-form" way. Thus gross profit $V(c, b)$ depends on the parameter b as well as on the level of unit costs. In the comparative statics analysis, the effect of b upon effort and welfare depends on $\partial V/\partial b$, which measures how profit varies with b, and on the cross-partial derivative $\partial^2 V/\partial c \partial b$, which measures how the sensitivity of V to c varies with b. If competition increases that sensitivity without greatly reducing profit, then competition does improve internal efficiency, because managerial reward can be linked more closely to effort.

The effect of competition upon incentives and efficiency has only recently been the subject of precise theoretical analysis, and much remains to be done. There are close parallels between this work and the recent literature on regulation under asymmetric information (see sections 4.3 and 4.6.2 below and the references cited therein). Both are branches of the principal–agent literature that examines the optimal design of incentive contracts. Although this literature on incentive theory is technically quite complex, it succeeds in illuminating some basic points. Imperfect information constrains the design of incentive contracts, and inefficiencies result. Information therefore has value insofar as it makes possible incentive schemes that are more sensitive to effort. Competition can act as a source of this kind of information, because rewards can be based on performance comparisons. Thus there is a theoretical underpinning to the idea that competition promotes efficiency within the firm, as well as efficient behavior in product markets.

3.5 Interconnection, Networks, and Vertical Integration

Having looked at some general aspects of competition theory pertinent to the privatization program, we now address the specific question of competition in networks and vertically related markets. In many industries involved in the U.K. privatization program, several distinct economic activities are required to supply the final product to consumers. Gas must be discovered, extracted, and distributed to the user's appliance. Electricity must be generated and transmitted through the grid. A long-distance phone call in effect consists of a trunk link and local links. The supply of a private branch exchange (PBX) telephone exchange involves both manufacture and distribution. A long-distance coach journey can be thought of as consisting of facilities to embark and disembark as well as the conveyance of the passenger from one town to another.

In such industries it is often the case that effective competition is more feasible in some activities than it is in others. For example, in the current state of technology, competition in long-distance telecommunications is possible, whereas in local networks typically there is natural monopoly and inevitable market power. In the energy industries the prospect of effective competition in the distribution and transmission of energy is remote, but there is no reason in principle why it should not exist in the production of gas or in the generation of electricity.

The question for policy in these circumstances is how to promote and maintain effective competition in activities where it is feasible in the face of monopoly in the related activities. In other words, can the problems of dominance be confined to those activities in which competition cannot exist, or can the dominant firm thwart competition throughout the industry? We shall pursue this question by examining an industry in which

output is produced by combining two inputs, A and B. The inputs could be thought of as long-distance and local telecommunications links respectively, or as gas production and distribution. We shall suppose that competition is infeasible in relation to input B, perhaps because natural monopoly conditions of supply prevail, but that competition can be effective in relation to input A.

We shall attempt to answer two questions.

(i) Will the dominant firm in activity B (the "B-firm") find it profitable to thwart competition in activity A?

(ii) If so, what can public policy do to safeguard competition in activity A?

In regard to (ii) we shall focus on two policy measures in particular—*vertical separation* (i.e. banning the B-firm from the A-sector), and rulings on *interconnection* (i.e. stipulating aspects of the relationship between the B-firm and the A-sector).

3.5.1 The Dominant Firm's Incentive to Thwart Competition

To answer question (i) let us begin by supposing that the B-firm can exclude competitors from the A-sector, for example by charging huge prices for the B input. This would happen if BT were allowed to charge Mercury excessive amounts for the local elements of the long-distance calls of Mercury's subscribers, or if British Gas could charge enormous fees for distributing gas produced by rival firms. Suppose that the B-firm is allowed to operate also in the A-sector. Would it wish to exercise its power to exclude rivals there?

If the B-firm's rivals in the A-sector had no cost advantage and if there were nondecreasing returns to scale, then the answer would certainly be positive. Whatever the level of total industry profits without exclusion, the B-firm could achieve at least that level of profit all for itself by excluding rivals in the A-sector and replicating their behavior. Moreover, the B-firm would probably want to change industry behavior, perhaps by restricting industry output, in order to boost industry profits yet further. Excluding rivals in these circumstances therefore allows a double gain—industry profits increase, and the B-firm obtains all of them rather than just part.

We have been speaking as though the inequalities in the previous paragraph were strict, as in general they are, but there is a famous special case in which they are not. This case, in which the B-firm is indifferent between excluding rivals in the A-sector and not doing so, is as follows. Inputs A and B are combined in fixed proportions, there are constant unit costs, and competition in the A-sector is so severe that price is driven down to the level of cost. It then turns out that the B-firm can extract all the monopoly profit that is to be had—for example by buying the A input (by assumption at cost) from the competitive A-sector and combining it with the B input, or by selling the B input to the A-sector with the full monopoly mark-up.

However, as soon as we relax the assumptions of the special case, there is a strictly positive incentive for the B-firm to exclude rivals from the A-sector. If the A and B inputs are combined in *variable* (as opposed to fixed) proportions, then the B-firm's market power is diminished by the ability of rivals in the A-sector to substitute away from the B input, at least to some extent. Alternatively, if *imperfect competition* exists in the A-sector, so that rivals there obtain positive profits, then the B-firm can appropriate those profits (and perhaps more besides) by excluding those rivals.

These rather straightforward arguments do not go through if the B-firm is at a cost disadvantage in the A-sector, or if returns to scale there are decreasing. The reason is that the average cost of producing the industry's output might increase if competitors were excluded from the A-sector, and industry profit might fall as a result. But this will not necessarily happen, because the B-firm will be able to exert greater market power which will tend to increase industry profits. In addition, the B-firm obtains all—rather than part—of industry profit in the event of exclusion. There is therefore a wide range of circumstances in which exclusion is profitable for the B-firm even if cost conditions do not satisfy the assumptions initially made. The example below concerning interconnection will illustrate this point.

3.5.2 Policy Measures to Safeguard Competition

So far we have shown that the B-firm generally has an incentive to thwart competitors in the A-sector by refusing to supply them the B input at reasonable prices. The possible exception is when the B-firm's rivals in the A-sector have a great efficiency advantage there. We now ask whether policy measures exist that can safeguard competition and improve social welfare. We focus on two such measures:

(i) vertical separation—not allowing the B-firm to own any firm that operates in the A-sector;

(ii) interconnect rulings—to stipulate terms of the relationship between the B-firm and the A-sector.

Telecommunications policies in the United States and in Britain illustrate (i) and (ii) respectively. In America when AT&T was broken up

after 1982 the long-distance division was separate from the local operating companies. Therefore the local operating companies, being under separate ownership, had less incentive to thwart competition to AT&T in the long-distance market (although some interconnection questions still arose). In contrast, the U.K. Government kept BT intact when privatization occurred. The ability of Mercury—BT's long-distance rival—to compete in the long-distance sector depends on a ruling by Oftel on interconnection which stipulates the terms on which BT must allow Mercury access to its local network.

Although a policy of vertical separation allows competition in the A-sector, it does not necessarily enhance social welfare by itself. The reason is the "double wedge problem" (see Waterson (1984, chapter 5) for a fuller discussion of this and other problems in vertically related markets). Suppose that A and B are combined in fixed proportions to produce the final product. Unless there is *perfect* competition in the A-sector, so that price is driven down to the level of costs, vertical separation has the effect of *raising* the price charged to the final consumer. If the B firm is "downstream," this happens because the B-firm will operate where marginal revenue is equal to marginal cost plus the mark-up of firms in the A-sector. The latter mark-up is absent if the rivals to the B-firm are excluded. Then marginal revenue is equal to the B-firm's overall marginal cost of supplying the final product and, unless the B-firm is relatively inefficient in the A-sector, price is lower with vertical integration. A similar argument operates if the B-firm is "upstream." Unless the A-sector is perfectly competitive, its demand for the B input will be described by a condition that implies that net price exceeds their unit costs. The demand for B is therefore reduced, and it follows that overall output is lower, and the price faced by consumers is higher, than if B were a vertically integrated firm.

In sum, under fixed proportions, with exogenous common unit costs in the A-sector and with no constraint on the market power of the B-firm, vertical separation has undesirable consequences because it exacerbates the harmful effects of the B-firm's monopoly power. This result can be reversed if we relax any of the stated assumptions. First, the assumption of fixed proportions is important, because under variable proportions the market power of the B-firm is weakened by virtue of the ability of rivals in the A-sector to substitute away from input B so as to cut costs. Secondly, if the B-firm has higher costs than other firms in activity A then vertical integration has an obvious drawback. Thirdly, the same is true if competition in activity A has the effect of stimulating internal efficiency in

that activity, since costs are then endogenous (see section 3.4). Finally, this entire analysis has supposed that the B-firm is unregulated. If, however, its market power is constrained by regulation, the merits of vertical separation are much stronger. Market power is held in check where it exists, and the advantages of competition are gained where possible. This issue will be taken further in relation to the industries discussed in later chapters, and with the aid of the theoretical perspective on regulation offered in chapter 4.

We now consider rulings on terms of *interconnection* as a second policy measure to combat the B-firm's desire to exclude competitors from the A-sector. More generally we refer to governmental regulation of the terms, notably the price, on which the A-sector can obtain input B from the dominant firm. We use the term "interconnection" because that is the word used in telecommunications economics, where this is central (see section 8.4.1 below). The same issue arises for example in relation to the licensing of innovations, a subject of antitrust cases (for an outline of the theory of licensing see Shapiro (1985)).

The question is to choose the optimal price at which the dominant firm must make input B available to the A-sector. We have already established that the B-firm would set that price at a prohibitively high level to exclude rivals if it had the discretion to do so. We shall use the following simple model to show the determinants of the optimal interconnection price. For simplicity we suppose that the B-firm has only one rival in the A-sector. This corresponds to the situation in the U.K. telecommunications industry: Mercury is BT's only rival in long-distance networks, and BT dominates local networks. Let a and b respectively be the unit costs of producing the A and B inputs, which we assume to be combined in a fixed one-to-one proportion. We will suppose that the cost of combining the inputs to produce the final product is negligible, so that $c = a+b$ is the marginal cost of supplying that product. (Alternatively, we could interpret either a or b as including the cost of combining the inputs.) Let i denote the "interconnection" charge, i.e. the amount paid to the B-firm by its rival for each unit of input B. Therefore the unit cost of supplying the final product is c for the B-firm and $a+i$ for its rival. Let X and Y respectively be the amounts supplied by those two firms, let $Q = X+Y$ be the industry output, and let $P(Q)$ be the inverse demand function. The profit of the dominant firm is then given by

$$\begin{aligned}\pi^D &= XP(Q) + iY - bQ - aX \\ &= [P(Q) - c]X + (i - b)Y,\end{aligned} \tag{3.11}$$

and the profit of the rival is

$$\pi^R = [P(Q) - (a+i)]Y. \tag{3.12}$$

If there is no government intervention, the dominant firm will set i at a prohibitively high level to exclude the rival and will produce where the marginal revenue equals c. In order to say what happens when i does not deter the rival, we need to specify the nature of the duopolistic interaction between the two firms. To this end we introduce a "conjectural variation" term λ as a summary statistic of the degree of rivalry. This term is often interpreted as each firm's expectation of how the other firm's output will change as a result of a change in its own level of output. (However, we are not committed to this interpretation: λ can simply be regarded as a device that saves repeating algebra, because $\lambda = 0$ corresponds to Cournot behavior, $\lambda = -1$ corresponds to Bertrand behavior, and so on).

From (3.11) and (3.12) we now have the first-order conditions

$$(1+\lambda)XP' + P - c + \lambda(i-b) = 0 \tag{3.13}$$

and

$$(1+\lambda)YP' + P - (a+i) = 0. \tag{3.14}$$

Adding (3.13) and (3.14) we obtain

$$(1+\lambda)QP' + 2P = (1+\lambda)c + (1-\lambda)(a+i) \tag{3.15}$$

Now let

$$\eta(Q) = -P/QP' \tag{3.16}$$

be the elasticity of demand. Then (3.15) can be written

$$P[2 - (1+\lambda)/\eta] = (1+\lambda)c + (1-\lambda)(a+i). \tag{3.17}$$

Optimality requires that price P equals marginal cost c. Substituting $P = c$ into (3.17) implies that the optimal interconnection charge is

$$i^* = b - \frac{(1+\lambda)c}{(1-\lambda)\eta}. \tag{3.18}$$

The terms on the right-hand side of (3.18) are b, the unit cost of the B input, and an adjustment term involving λ and η. For $-1 < \lambda < 1$, optimality requires i^* to be *lower than* b. This is in order to offset the duopoly mark-up arising from the imperfect competition between the two firms. This adjustment factor is greater when competition between firms is less intense (i.e. λ is larger) and when the elasticity of demand is lower. Therefore

marginal cost pricing ($i = b$) is *sub*optimal unless there is Bertrand, or price-taking, behavior in activity A. In this simple model, an interconnection charge below the level of cost is better. The dominant firm loses and the rival gains from such a policy.

The basic model is very rudimentary, and it can be extended in several directions, for example to the cases of several rivals, cost differences, nonconstant returns to scale, nonlinear pricing, or other social welfare objectives. As it stands the model does not include investment behavior, but it is important for interconnection policy to specify who pays what for investment, since there are externalities between the dominant firm and its rivals. This issue was part of Oftel's ruling on interconnection, which we shall consider in chapter 8.

A simple way of developing the model, which is relevant to the telecommunications industry, is to add a separate market for B. This is appropriate if A and B are long-distance and local links respectively, because there is a demand for local calls (B only) as well as for long-distance calls (which require A plus B). If price discrimination is disallowed, so that the dominant firm is constrained to charge the rival no more for B than it charges consumers in the separate B market, and if the dominant firm is otherwise unregulated, then it has an incentive to raise its price in the B market above the pure monopoly level. This injures the rival in the A-plus-B market and benefits the dominant firm there. This gain outweighs the loss from diverging from pure monopoly behavior in the B market, at any rate for small changes. A ban on price discrimination is therefore liable to introduce distortions elsewhere, and a ruling on interconnection is a superior policy to adopt.

The analysis becomes somewhat more complex if the dominant firm is also subject to regulatory constraint. BT must keep the price of a basket of its services below a given limit. If that constraint binds, the price of its local calls (B only) is inversely related to the price of its long-distance calls (A plus B). It follows that the dominant firm's incentive to price aggressively in the market where there is some competition is stronger, because revenues can be recouped in the other market. The profit function for the dominant firm stated in (3.11) needs extending to cover the other completely monopolized market, with the price in that market inversely related to P. Thus the regulatory mechanism may have the curious effect of making the dominant firm behave more aggressively towards rivals, because the pricing formula allows it to restore revenues elsewhere. We shall discuss further this possible disadvantage of constraining the price level of a wide *basket* of services in chapter 8 on the telecommunications industry.

Another important feature of that industry (among others) is the presence of *network externalities*. The demand for telecommunications services by any one customer depends upon who else is a customer of the same network, since he can call—and be called by—only those people. Network externalities are important for the nature of competition and welfare (see Katz and Shapiro, 1985, 1986; Farrell and Saloner, 1985). In regard to competition, it is clear that a rival would be at a grave disadvantage with respect to the dominant firm if it were denied access to its network of established subscribers. The interconnection question therefore has a further dimension. As to the welfare implications of network externalities, there is possibly a case for some subsidy to new subscribers to the network, since they provide a positive externality to existing subscribers. Space does not permit a fuller treatment of network economics here, but the issue will recur in later chapters.

3.6 Concluding Remarks

Theories about competitive forces of course account for a large portion of microeconomics, and it is therefore impossible in one chapter to say a great deal about them. Nevertheless it was essential for us to outline some selected aspects of those theories in view of the importance attached to competition in the British debate on privatization. To conclude, and to point forward to later chapters, we shall briefly recap on why we selected the topics described in this chapter. Section 3.2.1 looked at the trade-off that sometimes exists between competition and scale economies. This trade-off is at the heart of the question of how many rivals should be licensed to compete with a dominant enterprise such as BT. It also bears on the issue of desirability of unrestricted free entry. Section 3.2.2 discussed public enterprises as competitors, an important topic given that there is no necessary link between private ownership and competition.

Section 3.3 concerned the forces of potential competition, whose virtues have been widely acclaimed by proponents of deregulation and liberalization. Although we too would generally advocate freedom of entry, we argued that contestability theory was not suitable for the analysis of the major industries considered in this book, and we urged the need for policy measures to contain the danger that a powerful incumbent firm could render freedom of entry ineffective by predatory tactics. Section 3.4 briefly described some ways in which the competitive process can act as a mechanism with good incentives for internal efficiency. Such considerations have featured prominently in the debate on privatization and liberalization. Finally, in section 3.5 we examined some aspects of competition in vertically related markets, including the question of interconnection which arises not only in telecommunications but also in such industries as gas and electricity. In the following chapters we shall often have cause to refer back to the analysis of competitive forces described above. But first we must look at the economics of regulatory mechanisms for constraining market power.