tion of Bayesian equilibrium shows precisely how Nash equilibrium may be extended to this type of situation. Finally, in dynamic games with asymmetric information, the notions of perfect equilibrium and Bayesian equilibrium can be combined to further extend the relevance of Nash equilibrium.

Because most problems of industrial organization can be solved with a handful of basic game-theoretic concepts, it is recommended that the reader develop at least a casual familiarity with game theory. Although most of the arguments in part II can be understood at an intuitive level, the reader will benefit from a formal acquaintance with the concept of Nash equilibrium and its extensions, in the same way that optimization techniques clarify the study of the exercise of monopoly power. The reader may find the Game Theory User's Manual (chapter 11) useful in this respect.

Does noncooperative game theory remain relevant in situations in which firms appear to collude? In industrial organization, as in other fields, collusion and noncooperative behavior are not inconsistent. First, an altruistic party's objective function may embody the objectives of another party. In such a case, the first party's own interest is to make decisions that help the other party. (Here, altruism means cooperative actions taken purely for reason of self-interest.) Second, in the absence of altruism, parties facing conflicts may wish to change the rules of the game they are playing if this game has disastrous consequences for them. Signing a contract is a way of doing so. For instance, duopolists may agree to share the market in order to avoid cutthroat competition. However, signing a contract is formally only a part of a bigger noncooperative game. These two reasons why collusion can emerge from self-interested behavior may have limited relevance in IO. First, firms are rarely thought of as altruistic. Second, signing collusive contracts to prevent competition is often illegal. A third and more important reason is that, in a dynamic context, a firm may want to "pull its punches" because an aggressive action would trigger a rational reaction or retaliation from its opponents. (This will be emphasized in chapter 6 and, to a lesser extent, in chapter 8.) Again, the collusion is only apparent; it results from optimal noncooperative behavior. (This type of collusion is sometimes called *tacit* collusion.)

Reaction Functions: Strategic Complements and Substitutes

Consider a simultaneous-move game between (for simplicity) two firms. Assume that each action belongs to the real line and that the profit functions $\Pi^i(a_i, a_j)$ are twice continuously differentiable in the actions. The (necessary) first-order condition for a Nash equilibrium is that for each firm i

$$\Pi_i^i(a_i^*, a_j^*) = 0, \tag{2}$$

where a subscript denotes a partial derivative (e.g., $\Pi_i^i \equiv \partial\Pi^i/\partial a_i$). The second-order condition is that $a_i = a_i^*$ yields a local maximum:

$$\Pi_{ii}^i(a_i^*, a_j^*) \leqslant 0. \tag{3}$$

Assume that each firm's profit function is strictly concave in its own action everywhere: $\Pi_{ii}^i(a_i, a_j) < 0$ for all (a_i, a_j). Then the second-order condition is satisfied and, furthermore, the first-order condition given in equation 2 is sufficient for a Nash equilibrium. A Nash equilibrium is then given by a system of two equations with two unknowns (equation 2).

Let us define $R_i(a_j)$ as the best action for firm i given that firm j chooses a_j:

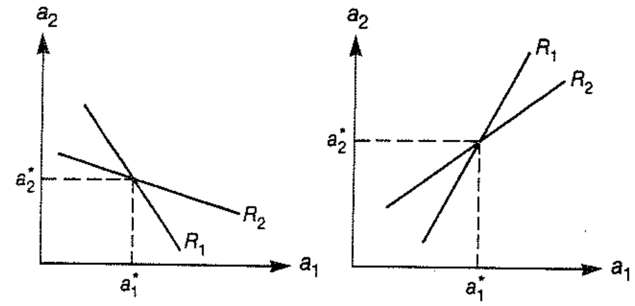
$$\Pi_i^i(R_i(a_j), a_j) = 0. \tag{4}$$

$a_i = R_i(a_j)$ is unique from our strict-concavity assumption,[2] and is called firm i's reaction to a_j. A Nash equilibrium is a pair of actions (a_1^*, a_2^*) such that $a_1^* = R_1(a_2^*)$ and $a_2^* = R_2(a_1^*)$. In such an equilibrium, each firm reacts optimally to the other firm's anticipated action.

A crucial element of part II is the sign of the slope of reaction functions for the various strategic variables we consider. This slope is obtained by differentiating equation 4:

$$R_i'(a_j) = \frac{\Pi_{ij}^i(R_i(a_j), a_j)}{-\Pi_{ii}^i(R_i(a_j), a_j)}. \tag{5}$$

2. We will assume that it exists and is an interior solution. In other words, going to the boundary of the feasible set of actions (e.g., $-\infty$ or $+\infty$) is not optimal for firm i.



Strategic substitutes ($\Pi^i_{ij} < 0$) Strategic complements ($\Pi^i_{ij} > 0$)

Figure 2

We thus have $\text{sign}(R_i') = \text{sign}(\Pi^i_{ij})$. Π^i_{ij} is the cross-partial derivative of firm i's profit function, i.e., the derivative of its marginal profit with respect to its opponent's action. The reaction curve is upward sloping if $\Pi^i_{ij} > 0$ and downward sloping if $\Pi^i_{ij} < 0$. Following Bulow, Geanakoplos, and Klemperer,[3] we will also consider the actions of the two firms to be strategic complements if $\Pi^i_{ij} > 0$ and strategic substitutes if $\Pi^i_{ij} < 0$.[4] As we shall see further on, prices are often strategic complements, and capacities are often strategic substitutes.

The construction of the reaction functions in a simultaneous-move game, performed in figure 2, is no more than a technical and illustrative device. By definition of simultaneous choices, a firm chooses its action before observing that of its opponent. Hence, it has no possibility of reacting. Reaction functions depict what a firm would do if it were to learn of a change in its opponent's action (which it does not). Points other than the Nash point on the reaction curves are never observed.

In contrast, reaction functions have real economic content in dynamic (sequential) games. For instance, if firm i chooses a_i first and firm j observes this choice before choosing a_j, firm i can use the function R_j to compute how a change in its behavior affects its opponent's behavior.

3. J. Bulow, J. Geanakoplos, and P. Klemperer, "Multimarket Oligopoly: Strategic Substitutes and Complements," *Journal of Political Economy* 93 (1985): 488–511.

4. This terminology is inspired by demand theory. Two goods are complements for a consumer if a decrease in the price of one good makes the other good more attractive to the consumer. Here, a decrease in a_j induces a decrease in a_i if $\Pi^i_{ij} > 0$, and conversely for substitutes.

8

Entry, Accommodation, and Exit

In the preceding chapter we saw how fixed costs (or, more generally, increasing returns) generate an imperfectly competitive market structure by limiting entry. However, even when fixed costs do restrict entry, positive (supranormal) profits are not ensured. Indeed, in the free-entry equilibrium, the firms make zero profit (up to the integer problem). In order to explain why the profit rate is systematically greater in certain industries than in others, some type of restriction to entry must exist in these industries to prevent other firms from taking advantage of the profitable market situations. Along these lines, Bain (1956) defined as a barrier to entry anything that allows incumbent firms to earn supranormal profits without threat of entry.[1]

Occasionally government restricts entry—for example, by introducing permits, licenses, patents, and taxi medallions. These restrictions may generate above-normal profits.[2] Other examples include the use of certain government purchasing policies or the granting of import licences (in situations that are not already domestically competitive, perhaps because of significant fixed costs) to form domestic monopolies.[3] In this chapter we consider barriers to entry not created by government.

Bain (1956) informally identified four elements of market structure that affect the ability of established firms to prevent supranormal profits (rents) from being eroded by entry:

1. Stigler (1968) offered an alternative definition based on cost asymmetries between incumbents and entrants. Von Weizsäcker's definition (1980a, p. 400) that "a barrier to entry is a cost of producing that must be borne by a firm which seeks to enter an industry but is not borne by firms already in the industry and that implies a distortion in the allocation of resources from the social point of view" is related to Stigler's. For comprehensive treatments of barriers to entry, see Encaoua et al. 1986 and von Weizsäcker 1980b.

2. In New York, a taxi medallion sells for $100,000. That can be interpreted as the present discounted value of the positive profits to be earned in the market, entry into which is legally restricted.

3. Another institutional barrier to entry may well be the lags and costs imposed by regulatory processes. For instance, MCI spent $10 million in regulatory and legal costs and waited seven years to gain permission to construct a microwave system, which cost $2 million and took seven months to complete. The established regulated firm, AT&T, which had a staff of lawyers and economists expert in regulatory matters, skillfully argued there was no need for the new service and that MCI only intended to enter the profitable part of the market, which, AT&T claimed, was used to subsidize some less profitable services ("cream skimming"). For criticisms of the Noerr-Pennington doctrine, which (particularly in the AT&T case) shields businesses from liability for their participation in governmental proceedings, see Brock and Evans 1983 and Brock 1983b. Those authors argue that business' interference in the regulatory process may be pure waste and that, because abuses are unlikely to be caught, this interference (called "regulatory-process predation") should be dealt with severely.

Economies of scale (e.g., fixed costs) Bain argued that if the minimum efficient scale is a significant proportion of the industry demand, the market can sustain only a small number of firms that make supranormal profits without inviting entry. This argument is examined in section 8.1, where we examine natural monopoly or oligopoly situations and the theory of contestability. See also section 8.6.1.

Absolute cost advantages The established firms may own superior production techniques, learned through experience (learning by doing) or through research and development (patented or secret innovations). They may have accumulated capital that reduces their cost of production. They may also have foreclosed the entrants' access to crucial inputs through contracts with suppliers. In sections 8.2 and 8.6.1 we consider the accumulation of capital by incumbents. Section 4.6.2 examined the market-foreclosure doctrine. R&D activity is studied in chapter 10.

Product-differentiation advantages Incumbents may have patented product innovations (which, of course, can be seen as a cost advantage relative to the product), or they may have cornered the right niches in the product space, or they may enjoy consumer loyalty. (The niche argument is examined in section 8.6.2.)

Capital requirements According to this controversial element of entry barriers, entrants may have trouble finding financing for their investments because of the risk to the creditors. One argument is that banks are less eager to lend to entrants because they are less well known than incumbents; another (which will be examined in section 9.7) is that entrants may be prevented from growing as incumbents inflict losses on them in the product market in order to reduce their ability to find financing for new investments.

Bain also suggested three kinds of behavior by incumbents in the face of an entry threat:

Blockaded entry The incumbents compete as if there were no threat of entry. Even so, the market is not attractive enough to entrants.

Deterred entry Entry cannot be blockaded, but the incumbents modify their behavior to successfully thwart entry.

Accommodated entry The incumbents find it (individually) more profitable to let the entrant(s) enter than to erect costly barriers to entry.

Bain's suggestions obviously begged for further analysis. The most famous model of barriers to entry is the "limit pricing model" (Bain 1956; Sylos-Labini 1962; Modigliani 1958), the basic idea of which is that, under some circumstances, incumbent firms may sustain a price so low that it discourages entry. This story remained controversial until Spence (1977), Dixit (1979, 1980), and Milgrom and Roberts (1982) clarified its underlying aspects.[4] Very roughly, the Spence-Dixit reconsideration (section 8.2 below) offers to regard the Stackelberg model of sequential quantity competition as one of sequential capacity choices. That is, although product-market competition (if any) determines the market price in the short run, in the longer run firms compete through the accumulation of capacity. (See chapter 5 for the reinterpretation of quantities as capacities.) An incumbency advantage (the possibility of early capital accumulation) leads the incumbent firm to accumulate a large capacity (and therefore to charge a low price) in order to deter or limit entry. The Milgrom-Roberts reconsideration of the limit-pricing story (studied in chapter 9) is based on the asymmetry of information between the incumbent and the entrant. In their model, the incumbent charges a low price not because he has a large productive capacity (capacity constraints play no role there) but because he tries to convey the information that either the demand or his own marginal cost is low, thus signaling a low profitability of entry to the potential entrant(s). These two models have fairly distinct positive and normative implications.

Erecting barriers to entry is only one aspect of strategic competition. Inducing exit of rivals is another. And even if neither entry nor exit is at stake (the "accommodation" case), firms battle for market shares. Chapter 6 examined examples of such battles, in which firms repeatedly compete in price. Firms also compete in non-price aspects

4. Part of the controversy is due to the fact that the timing of the underlying game and the strategic instruments were not completely described (for instance, the "Sylos-Labini postulate" holds that potential entrants expect established firms to maintain the same output if entry occurs, yet the story is named "limit-pricing"), nor was the commitment value of either quantity or price carefully examined.

(capacities, technology, R&D, advertising, product differentiation, etc.). Chapters 5 and 7 offered examples of non-price competition, but there we focused on once-and-for-all (static) situations in which firms choose their non-price variables simultaneously; the important possibility of influencing rivals' subsequent non-price behavior was ignored. This chapter examines strategic interaction in a dynamic context.

There are a variety of business strategies available to a firm, depending on whether it wants to deter entry, to induce exit, or (if those goals are too costly) to do battle with its rivals. As we will see, optimal strategies also depend on whether reaction curves are sloping upward (strategic complements) or downward (strategic substitutes). Section 8.3 offers a taxomony of relevant business strategies, all of them meant to soften the rivals' behavior. Section 8.4 applies these strategies to a number of strategic situations.

The excellent surveys of Gilbert (1986, 1987), Kreps and Spence (1984), Shapiro (1986), and Wilson (1984) address some of the points raised in this chapter. Much of the material of this chapter and the following one is derived from Fudenberg and Tirole 1986 (see also Fudenberg and Tirole 1984). Section 8.1 draws from Fudenberg and Tirole 1987.

8.1 Fixed Costs: Natural Monopoly and Contestability

This section addresses the role of fixed costs as a barrier to entry. Recall Bain's argument that under increasing returns to scale, only a finite number of firms are viable, and these firms make positive (supranormal) profits without triggering entry—for instance, if potential entrants know that a duopoly yields negative profits, an established firm can quietly enjoy a monopoly profit without worrying about the threat of entry. This conclusion was challenged by Baumol, Panzar, and Willig (1982), who argued that having one or a limited number of firms does not mean there is no competition and that potential competition (the threat of entry) may serve to discipline established firms.[5]

8.1.1 Fixed Costs versus Sunk Costs

In a one-period (i.e., timeless) view of the world, a fixed cost is easily defined as a cost that a firm must incur in order to produce and that is independent of the number of units of output. For instance, a firm may incur cost $C(q) = f + cq$ for $q > 0$ and cost $C(q) = 0$ for $q = 0$. (Fixed costs are instances of increasing returns to scale. See the chapter on the theory of the firm for the notions of subadditivity and natural monopoly.) The timeless model of production is, of course, an abstraction. Once time is introduced, one must carefully define the notion of production period. To see this, suppose (with Weitzman [1983]) that a firm produces output $q > 0$ per period in two consecutive periods at cost $2(f + cq)$, where f is the per-period fixed cost. Absent entry and exit costs, it would be cheaper to produce output $2q$ in the first period and 0 in the second. This would cost $f + 2q$ and save f. (We ignore interest and storage costs, assuming that the lag between the periods is short; we also ignore uncertainty about future demand, which may lead firms to wait to produce future supply.) More generally, dividing the production period by 2 and doubling the production intensity saves on fixed costs, so that all production should take place over a very short interval of time and fixed costs should be negligible relative to variable costs. To avoid this extreme conclusion, it is important to realize that fixed costs are always sunk to some extent. The presence of market imperfections prevents instantaneous rental of capital or hiring of labor. Or the firm may need to buy up front specific investment that has no intrinsic value to other firms (and therefore has no value on a second-hand market) and cannot be allocated to another use within the firm.

We will define fixed costs as costs that are independent of the scale of production and are locked in (committed, sunk) for some short length of time, which defines the "period." For example, suppose that deciding to produce a positive quantity requires a firm to immobilize machines, capital, land, legal, public relations, and advertising services, and general staff for one month. The firm cannot get away with incurring half of the relevant fixed costs and doubling its production rate during fifteen days, stop

5. See Baumol et al. 1982 for further references. See also Brock 1983a, Spence 1983, Baumol et al. 1986, and Schwartz 1986.

production, and save the remaining half during the second fortnight (and possibly resume production thereafter). Thus, one can envision a discrete-time model in which a firm incurs a cost of $f + cq$ in each period if it produces at that date and zero otherwise. The real time length of each period indicates the length of time over which the cost is incurred.[6]

The distinction between "fixed costs" and "sunk costs" is one of degree, not one of nature. Fixed costs are sunk only in the short run. (Of course, there is the question of how short the short run is, and how the length of commitment to investments compares with the time scale of product competition, e.g., of price changes. We will come back to this issue when discussing the contestability theory.) Sunk costs are those investment costs that produce a stream of benefits over a long horizon but can never be recouped. A machine will be labeled a fixed cost if the firm rents it for a month (or can sell it without capital loss a month after its purchase) and a sunk cost if the firm is stuck with it.

The notions of fixed and sunk costs are idealizations for several reasons. First, there is clearly a continuum of degrees of commitment between these two polar cases of short and eternal commitment. Second, both notions assume that the investment cost cannot be recouped at all during the commitment period (whatever it is). In practice, a machine would have some value lower than its original value on the second-hand market. Also, leasing and labor contracts can be breached at some penalty cost. Thus, commitment is not quite an all-or-nothing notion. What we really mean by period of commitment is a period of time over which the cost of being freed from the commitment within the period is sufficiently high that it does not pay to be freed. For simplicity, we will content ourselves with assuming that investment costs are completely sunk for the whole period. Third, and a related point, our notion of commitment is largely a purely technological one (though filtered through the existing set of input-market institutions). In practice, the date at which a firm resells its assets or modifies its rental or labor contracts may also depend on how well the firm is doing in the product market and on strategic considerations in this market.

8.1.2 Contestability

Following Baumol et al. 1982, let us consider a homogeneous-good industry with n firms. All firms have the same technology, and producing output q costs $C(q)$ with $C(0) = 0$. We split the set of firms into two groups: m "incumbents" (without loss of generality, we can assume that the incumbents are firms $i = 1, \ldots, m$) and $n - m \geq 0$ "potential entrants."

An industry configuration is a set of outputs $\{q_1, \ldots, q_m\}$ for the incumbents and a price p charged by all incumbents (the potential entrants stay out of the market).

The industry configuration is *feasible* if the market clears (i.e., if total output is equal to total demand at price p: $\sum_{i=1}^{m} q_i = D(p)$) and if firms make non-negative profits (for any incumbent firm, $pq_i \geq C(q_i)$). It is *sustainable* if no entrant can make a profit taking the incumbents' price as given (there do not exist a price $p^e \leq p$ and an output $q^e \leq D(p^e)$ such that $p^e q^e > C(q^e)$).

A *perfectly contestable market* is one in which any equilibrium industry configuration must be sustainable.

These definitions extend straightforwardly to multiproduct technologies; it suffices to allow outputs and prices to be multidimensional vectors. Indeed, the theory of contestability has been partly motivated by multiproduct technologies, and some of its interesting developments are related to the issue of "cross-subsidization." (See footnote 7 below.)

Here we will content ourselves with an exposition of the single-product case.

To illustrate the concept of sustainability, let us consider our standard example of increasing-returns technology:

$$C(q) = f + cq.$$

Let

$$\tilde{\Pi}^m \equiv \max_q \{[P(q) - c]q\}$$

denote the monopoly profit gross of the fixed cost. Assume that a monopoly is viable: $\tilde{\Pi}^m > f$. Figure 8.1 depicts the unique sustainable configuration in this industry. There exists only one incumbent in the industry, charging price p^c and supplying output q^c. The other firms stay out. The contestable price-output pair $\{p^c, q^c\}$ is obtained

6. See page 363 of Baumol et al. 1986 for a more complete discussion of this point.

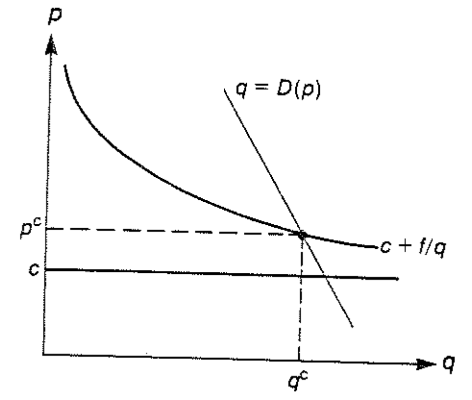


Figure 8.1

from the intersection of the average-cost curve and the demand curve:

$$(p^c - c)D(p^c) = f.$$

A firm that charges $p < p^c$ and produces a positive quantity loses money, because its price is below the average cost. (This also shows that the contestable price is smaller than the monopoly price p^m.) Conversely, a price above p^c is not sustainable, because an entrant can undercut this price and still make a strictly positive profit.

In this example, the theory of contestability predicts the following conclusions:

(1) There is a unique operating firm in the industry (technological efficiency).

(2) This firm makes zero profit.

(3) Average-cost pricing prevails. Furthermore, the allocation is constrained efficient, in the sense that it is socially efficient, given the constraint that a social planner does not use subsidies.[7]

Thus, the mere "threat of entry" has an effect on the market behavior of the incumbent firm (conclusion 2 and first part of conclusion 3). The second part of conclusion 3 is not surprising. The fixed cost is not duplicated in the sustainable outcome. Thus, only the market price matters in the assessment of efficiency. Clearly, the first-best outcome is obtained when the incumbent charges the marginal cost; however, in the absence of a subsidy, the firm would lose f and would not be willing to operate. Short of the first-best outcome, a social planner prefers the lowest price that allows the firm to make a non-negative profit, i.e., p^c.[8]

This set of conclusions is striking. It has long been argued that an industry subject to non-negligible increasing returns could not behave competitively and therefore should be nationalized, or at least carefully regulated. If, however, such an industry behaves like a perfectly contestable market, it comes as close to marginal-cost pricing as is consistent with viable firms (if subsidies are prohibited). In the absence of actual competition, potential competition is very effective in disciplining the incumbent firms. Hence, the unregulated organization of industries with increasing returns to scale should be less of a problem than would appear at first glance. Clearly,

7. In the multiproduct case, Baumol et al. (1982) show that a sustainable allocation, if it exists, satisfies the following conditions: (a) Industry cost minimization holds (a generalization of conclusion 1). (b) Firms make no profit (conclusion 2). (c) The revenue made by a firm on a subset of products is at least as big as the cost savings that would result from not producing these products (keeping the outputs of the other products as given). (d) The price of a product exceeds its marginal production cost for any firm that produces it. They are equal if more than one firm supplies the good. (e) Under some assumptions (see Baumol et al. 1977), Ramsey prices and outputs—i.e., those that are welfare optimal subject to the constraint that the firm earns a profit equal to the maximum profit permitted by barriers to entry—are sustainable.

The intuition for condition c (the no-cross-subsidization result) is that if a set of products were not viable, an entrant could come in with the same production as an incumbent except that it would drop these products and thus make money. Condition d is a generalization of Bertrand competition.

8. To be rigorous, we must check that the social planner could not do better by forcing the firm to randomize between different prices. To see that price randomization lowers welfare, it suffices to show that the aggregate welfare $W(p)$ is concave in p. If this is the case, then from Jensen's inequality $E\,W(p) \leqslant W(Ep)$, and welfare is higher under the deterministic price Ep than under the random price p (where E denotes expectation over the price). If, furthermore, the profit function $\Pi(p)$ is concave in p, the firm makes non-negative profits under the deterministic price Ep if it makes non-negative profit under the random price p (since $\Pi(Ep) \geqslant E\,\Pi(p) \geqslant 0$), so the firm's non-negative profit constraint is harder to satisfy with a random price than with a deterministic one. For our purpose, let us assume that

$$D'(p) + (p - c)D''(p) \leqslant 0.$$

Then

$$\Pi''(p) \equiv 2\,D'(p) + (p - c)D''(p) < 0.$$

Also,

$$\begin{aligned} W''(p) &= [S(p) + \Pi(p)]'' \\ &= [-D(p) + D(p) + (p - c)D'(p)]' \\ &= D'(p) + (p - c)D''(p) \leqslant 0 \end{aligned}$$

(where S denotes the net consumer surplus). Thus, both Π and W are concave. For a much more general result on the undesirability of random prices, see Samuelson 1972.

such a theory, if applicable, has strong implications for the deregulation of the airlines and similar industries.

Baumol et al. (1982) show that, for different demand and cost functions, natural monopolies may not be sustainable. That is, there may not exist a price-output pair $\{p^c, q^c\}$ such that firms make non-negative profits, the market clears, and the allocation cannot be upset by profitable entry at price-output pair $\{p^e, q^e\}$ such that $p^e \leqslant p^c$ and $q^e \leqslant D(p^e)$. That is, constrained efficient market structures may not be sustainable against entry.

*Exercise 8.1*** In a one-good industry, consider a U-shaped average-cost curve. Suppose that the demand curve intersects the average-cost curve slightly to the right of the most efficient scale (i.e., the average-cost-minimizing output). Using a diagram, show that there exists no sustainable allocation.

The natural question is this: Which situation is depicted by the contestability axioms—in particular, the sustainability axiom? One would want to describe (at least in a stylized way) competition in a natural-monopoly industry, and to compare its outcome with the contestable one.

One game that yields the contestable outcome is the following: Suppose that firms first choose prices simultaneously and then choose outputs. (Picking an output involves deciding whether to enter—i.e., whether to choose a strictly positive output.) This two-stage game is the reverse of the two-stage game described in chapter 5, in which firms chose quantities before prices. Suppose that all potential firms choose price p^c. Then one of them chooses output q^c and the others stay out (produce nothing). This is clearly an equilibrium. All firms make zero profit. If a firm were to undercut p^c, it could not supply the market profitably.[9] As Baumol et al. rightly note, the theory of perfectly contestable markets can thus be seen as a generalization of Bertrand competition to markets with increasing returns to scale.[10]

The preceding game portrays the vision of an industry in which prices adjust more slowly than decisions about quantities or entry. Prices are considered rigid at the time firms choose their quantities. As prices are generally thought of as amenable to relatively quick adjustment, the technology thus involves a fixed cost in the sense of subsection 8.1.1. This vision is implicit in the slightly more sophisticated "hit-and-run entry" story offered by the proponents of contestability. Suppose that the incumbent's price is rigid for a length of time τ, and that entry and exit are costless. If the incumbent's price exceeds p^c, an entrant can enter, undercut p^c slightly (thereby conquering the incumbent's entire market share), and exit the industry before τ units of time having elapsed—i.e., before the incumbent can respond by lowering his price. The entrant (who, by assumption, incurs no entry or exit cost) thus makes a positive profit. Thus, only price p^c is "sustainable."

This interpretation of contestability has come under attack on the grounds that prices seem to adjust more rapidly than decisions about quantities or entry. Price adjustment does seem faster in the railroad industry, where entry and expansion entail a long-process of buying up parcels of land (generally requiring powers of eminent domain), engineering and building the railroad, and so on. It may be even faster in the airline industry, where opening a new route is a relatively fast process.[11,12]

---

9. To prove that this is the unique equilibrium, consider the highest price $\bar{p} > p^c$ charged in equilibrium by any firm. Show that this price has probability 1 of being *strictly* higher than the lowest price charged by the other firms. Conclude that this firm makes zero profit, which in turn implies that the lowest price charged by the other firms is p^c with probability 1.

10. See Grossman 1981 for an alternative approach to contestability in a one-good industry. Grossman assumes that firms announce supply curves rather than prices.

11. However, Bailey and Panzar (1981) argue that the theory of contestable markets is relevant to city-pair airline markets. There are returns to scale in this industry, but fixed costs are not sunk. (The aircraft can be recovered at little cost. Sunk costs, such as those for runways, towers, and ground facilities, are incurred by municipalities.) Bailey and Panzar offer some evidence that monopolists (almost 70% of routes are served by a single carrier) behave more or less competitively on their long-haul routes immediately after deregulation. In contrast, Bailey et al. (1985) and other find that fares are higher when concentration is higher when all else is equal (but the relation, although statistically strong, is not economically large).

12. Brock and Scheinkman (1983) study "quantity sustainability." They say that a price-quantity allocation $(\bar{p}, \bar{q})$ is quantity sustainable if any production plan by an entrant q^e makes negative profit at the market-clearing price for quantity $\bar{q} + q^e$. That is, the entrant assumes that the established firm's output remains fixed after entry. Brock and Scheinkman show that under some assumptions price sustainability implies quantity sustainability, and that in the single-product case the allocation (p^c, q^c) at which the demand curve intersects the average-cost curve is quantity sustainable (it is not necessarily price sustainable—see exercise 8.1).

Perry (1984) considers price strategies but departs from the uniform-pricing assumption made by Baumol et al. (1982). The incumbent announces a price-quantity schedule: He stands ready to supply q_1 units of the good at price p_1, then q_2 more units at price $p_2 > p_1$ (so that his total supply at price p_2 is $q_1 + q_2$), and so on. The entrant reacts by announcing a price-quantity schedule himself. Sustainability is easier to obtain than under uniform pricing, because

If one takes the view that prices generally adjust more rapidly than capacities, the incumbent's price is unlikely to be locked in when the entrant finishes assembling his production facilities. That is, entry ought to induce the incumbent to reduce his price fairly quickly to adjust to competitive pressure. If the incumbent's price reacts quickly to entry (where "quickly" is relative to the time scale of the entrant's investment), hit-and-run entry is not profitable, as there is no scope for two price-competing firms in a natural monopoly.

An alternative way of thinking about contestability is to envision short-run capacity commitments rather than price rigidities. In this view, prices adjust "instantaneously." (This, of course, is not realistic; it is a metaphor for the idea that prices adjust quickly relative to the time scale of the capacity game.) That is, at any point in time, each firm chooses its price so as to maximize its profit, given the current vector of capacities.

An old intuition in industrial organization states that if the incumbent is committed to his capacity only in the short run, he and the potential entrant are almost on equal footing, so that barriers to entry (and the incumbent's profit) are low. Indeed, in a model where firms are stuck with their capacity choices for a short period of time, it can be shown that there exists an equilibrium in which only the incumbent produces; this apparent monopolist accumulates and constantly renews (approximately) capacity q^c and makes (almost) no profit. If the incumbent's equilibrium capacity were lower (allowing positive profits), an entrant could come in and, because the incumbent's capacity commitment is short, would incur duopoly losses for a short time before the incumbent would exit. The entrant would then take over the market and enjoy incumbency. Thus, the prospect of high steady-state profits together with the brevity of the fight to kick out the incumbent would encourage entry. This approach to contestability is developed in more detail in the supplementary section.

8.1.3 War of Attrition

Another popular approach to natural monopoly is the war of attrition. Like the short-run capacity-commitmen approach sketched in the preceding paragraph, it assume that price adjustments take place more quickly than quan tity adjustments.

The war of attrition was introduced in theoretical biol ogy, by Maynard Smith (1974), to explain animals' fight for prey. Two animals fighting for prey may resembl two firms fighting for control of an increasing-return industry. Fighting is costly to the animals; at the ver least, they forgo the opportunity of other activities an become exhausted. Similarly, duopoly competition ma be costly because it generates negative profits. In bot cases, the object of the fight is to induce the rival to giv up. The winning animal keeps the prey; the winning firm obtains monopoly power. The loser is left wishing it ha never entered the fight. (For such a fight to take place, it outcome cannot be deterministic. Each player must hav at least some chance of winning in order to be willing t participate.) In a war of attrition, each player waits an suffers for a while. If at some point in time his rival ha not yet quit, a player gives up.

The simplest example of a war of attrition is the follow ing: Suppose time is continuous from 0 to $+\infty$. The rat of interest is r. There are two firms, with identical cos functions $C(q) = f + cq$ if $q > 0$ and $C(0) = 0$, per unit c time. Price adjustments are instantaneous. If the two firm are in the market at time t, price equals marginal cost (Bertrand competition) and each firm loses f per unit c time. If only one firm is in the market, the price is equa to the monopoly price, p^m, and the firm makes instanta neous profit $\tilde{\Pi}^m - f > 0$; the other firm makes zero profi Both firms are in the market at date 0. At each instan each firm decides whether to exit (conditional on th other firm's still being in the market at that date). Exit costless. For simplicity, assume that a firm that drops ou never returns (however, the equilibrium we describe belo

profitably undercutting an incumbent is more difficult. The incumbent can sell just enough units at low prices so that the entrant's residual demand curve is moved to the left of his average-cost curve; in a sense, the incumbent is able to commit to a certain output through low prices on these units but can still make money through high prices on the marginal units. Perry shows that the incumbent generally makes a strictly positive profit, and that the existence of sustainable price-quantity strategy may not even require the natural-monopo assumption (which assumption is necessary but not sufficient for the existen of a sustainable allocation under uniform pricing).

is still an equilibrium if costless reentry is allowed). Because the market is profitable for a monopoly, the remaining firm stays in forever after its rival has dropped out.

We now construct a symmetric equilibrium in which, at any instant, each firm is indifferent between dropping out and staying. For a firm to be indifferent, the expected profits from the two actions must be the same. Because dropping out at t means zero profits *from that date on*, each firm's expected present discounted value of profits from any date on must equal zero. If both firms are still in the market at date t, each firm drops out with probability $x\,dt$ between t and $t + dt$, where $x \equiv rf/(\tilde{\Pi}^m - f)$. To see that these strategies form an equilibrium, suppose that at date t both firms are still in the market. If firm 1 drops out, it obtains 0 from t on. If firm 1 stays in until time $t + dt$, it incurs duopoly loss $f\,dt$. However, with probability $x\,dt$, firm 2 drops out during this short interval of time. Firm 1 then becomes a monopoly and, from then on, earns total (discounted) profits $(\tilde{\Pi}^m - f)/r$. If firm 2 is still in at date $t + dt$, firm 1 is willing to drop out and thus make 0 from that date on. Firm 1 is indifferent between dropping out at date t and staying until $t + dt$ if

$$0 = -f\,dt + (x\,dt)[(\tilde{\Pi}^m - f)/r] + 0.$$

The industry outcome is stochastic. Each firm drops out according to a Poisson process with parameter x.[13]

This equilibrium is consistent with free reentry because the value of being in is 0, so there is no reason to reenter once one has left. The equilibrium is not unique,[14]; however, if we depart from our perfect-information assumption and allow uncertainty about rival's fixed (opportunity) costs (see chapter 9), and if the support of this uncertainty is sufficiently large, then the symmetric equilibrium is also the unique equilibrium.

The war of attrition yields the following conclusions:

(1′) There are two firms in the industry for a (random) length of time (technological inefficiency); then one exits.

(2′) Firms earn no *ex ante* rents, but may have *ex post* profits.

(3′) The price is first competitive and then equal to the monopoly price. The allocation is not constrained efficient, and welfare is lower than under contestability.

The second part of conclusion 3′ results from the fact that the contestable allocation is optimal, subject to the no-subsidy constraint. The following exercise checks this result for a simple specification of demand.

*Exercise 8.2*** All the firms in an industry have the same production cost: $C(q) = f = \frac{3}{16}$. (The marginal cost is 0.) The demand is $D(p) = 1 - p$.

(i) Is this a "natural monopoly"?

(ii) Compute the contestable allocation. Calculate the welfare level.

(iii) Derive the symmetric equilibrium of the infinite-horizon, continuous-time war of attrition between two firms. Calculate the expected intertemporal welfare, and compare it with the welfare level from question ii. (Hint: For a Poisson process with parameter y, the probability that no arrival has occurred by time t is e^{-yt}.)

Figure 8.2 illustrates the difference in price dynamics in the contestability and war-of-attrition theories.

It may be instructive to look at the natural monopoly issue from the viewpoint of the literature on rent seeking. As was noted in chapter 1, Posner argued that the prospect of monopoly profits creates a contest to appropriate those profits. All monopoly profits must be added to the usual dead-weight-loss triangle if two postulates

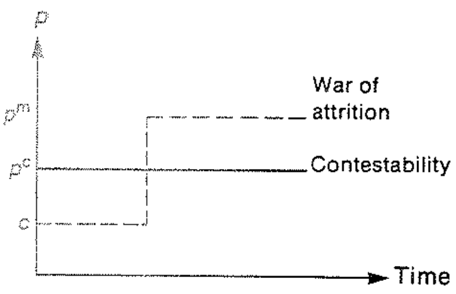


Figure 8.2
Price dynamics in natural monopoly.

13. That is, the cumulative probability that a firm drops out before date t, conditional on the other firm's not dropping out, is $1 - e^{-xt}$ (an exponential distribution).

14. There also exist asymmetric equilibria: For example, at each date firm 1 stays in and firm 2 exits (another equilibrium is obtained by switching the names of the firms).

hold: the rent-dissipation (or zero-profit) postulate, which asserts that the firms' total expenditure on obtaining the monopoly profit is equal to the monopoly profit, and the wastefulness postulate, which asserts that this expenditure has no socially valuable by-products.

Both the contestable allocation and the war-of-attrition equilibrium satisfy the rent-dissipation postulate. Competition for the monopoly position drives industry profits down to zero.[15] The contestable allocation yields an interesting reversal of the wastefulness postulate. Because rent dissipation occurs through low prices, it benefits the consumer and is socially useful. The war-of-attrition equilibrium comes closer to satisfying the wastefulness postulate than the contestable allocation. Some of the profits are dissipated wastefully (for a while, the fixed production cost is duplicated). But the consumers also enjoy marginal-cost pricing for some time before facing the monopoly price (Posner's allocation would correspond to monopoly pricing at each instant). Thus, welfare is higher than that predicted by the rent-seeking literature and lower than that associated with the contestable allocation.

Another interesting analogy relates to the free-entry biases discussed in chapter 7. As Whinston (1986) notes, one can view the exit decision as a reverse entry decision. Therefore, this decision is subject to the same biases—imperfect appropriability of the consumer surplus and business-stealing effect—as the entry decision. Let $w(p)$ denote the social welfare per unit of time, gross of fixed cost. To illustrate the two biases, assume that there are two consumers, with unit demands, and that $c = 0$. First, to focus on the business-stealing effect, suppose that the two consumers have the same valuation v for the good. The flow monopoly profit is then $\tilde{\Pi}^m = v$. A monopolist captures the full consumer surplus and introduces no distortion in consumption. Thus,

$$w(c) - w(p^m) = v - v = 0 < f.$$

The social gain from competition per unit of time is lower than the flow fixed cost of production. It is socially optimal to have a single firm at any point of time even if its pricing behavior cannot be regulated. Thus, there is too little exit. Second, suppose that the two consumers have different valuations $v_1 < v_2$ and that $v_2 > 2v_1$, so that a monopolist charges price v_2. By charging v_1 (which would induce a socially optimal consumption), it would capture only part of the total consumer surplus. Now, if f is lower than

$$w(c) - w(p^m) = (v_1 + v_2) - v_2 = v_1,$$

competition is valuable.[16] That is, when a firm chooses to exit (because its private incentive to stay is zero), a social planner would like it to stay—there is socially too much exit, because the firms do not appropriate the gain in consumer surplus due to competition. Thus, in a second-best world in which pricing cannot be regulated, a social planner would want to prevent any exit.

The preceding analysis relies on strong price competition between the two firms. Suppose they succeed in tacitly colluding in prices while both are still in the market (see chapter 6 for a discussion of tacit collusion). The market price is then equal to p^m, independent of the number of remaining firms. Thus, a social planner would want one of the firms to exit at date 0 in order to avoid wasteful duplication of the fixed cost. However, suppose that the firms wage a war of attrition, and they lose $(f - \tilde{\Pi}^m/2) > 0$ per unit of time while competing. In the symmetric equilibrium, each firm exits with probability $x'dt$ between t and $t + dt$, where x' is given by

$$(f - \tilde{\Pi}^m/2)dt = [x'(\tilde{\Pi}^m - f)/r]dt,$$

which yields $x' < x$. Because fighting for a monopoly position is less costly under tacit collusion, firms exit at a slower rate precisely when a social planner would prefer a single firm. Here we have an example of the business-stealing effect. Staying in has no social value; all profits are derived from diverting half the monopoly profit from one's rival (and the full monopoly profit if this rival exits). Under tacit collusion, there is socially too little exit.[17]

The war-of-attrition paradigm has been used to try to predict whether big firms or small firms are more likely to

15. One way of looking at this is as follows. The monopoly profit in a contestable market turns out to be zero. No expenditure is made to obtain it. In contrast, in the war of attrition, the monopoly profit is the regular one. The expenditure corresponds to the duopoly losses incurred prior to giving up or getting the monopoly situation.

16. In this example, $\tilde{\Pi}^m = v_2 > 2v_1 > 2f$. As long as firms wage Bertrand competition, $f > 0$ is sufficient for the market to be a natural monopoly.

17. See Mankiw and Whinston 1986 and review exercise 24 for analyses of the free-entry biases in a homogeneous-good industry in a static context.

exit first in a declining industry with increasing returns to scale. Ghemawat and Nalebuff (1985) argue that big firms will exit earlier, leaving the industry to the small firms. The intuition is that if demand declines, a big firm loses viability more quickly (it is too big relative to the market). Thus, in a monopoly situation, a big firm would exit earlier than a small firm. In a competitive duopoly, the small firm's anticipation that the big firm will eventually leave is an incentive for the small firm to stay in the market. As Ghemawat and Nalebuff show, this forces the big firm to exit as soon as its instantaneous duopoly profit becomes negative (i.e. no real war of attrition takes place on the equilibrium path).[18] Londregan (1986) extends this model to allow a complete product life cycle, in which the market grows and then declines.[19]

Whinston (1986) shows that the Ghemawat-Nalebuff result depends crucially on the big firms' inability to "go on a diet." He argues that, in practice, a big firm may be able to reduce the number of plants and become a small firm when demand declines. He then solves for equilibrium when firms can scrap plants (exit then occurs when the last plant is closed) and shows that a variety of potential outcomes are feasible. Indeed, Whinston notes that in the declining industry producing the antiknock additive for leaded gasoline, the smallest producer was the first to leave. Ghemawat and Nalebuff (1985) give a few examples, including the synthetic-soda-ash industry and the British steel-castings industry, in which the largest firms exited first.

There is a simple case for which the outcome can be predicted without intimate knowledge of the industry. Ghemawat and Nalebuff (1987) and Whinston (1986) show that if firms can decrease their capacities after a downward shock in demand that calls for exit, the bigger firm reduces its capacity until it is equal in size to its rival, and thereafter the two firms reduce their capacities symmetrically (so they remain of equal size).[20]

In chapter 9 we will consider another aspect of the war of attrition: the possibility that each firm has incomplete information about its rivals' production or opportunity costs. The length of time already spent in a ruinous oligopoly contest is then a signal that a firm is efficient (or has low outside opportunities, or that the market exerts beneficial spillovers on its other product lines). The link between the war of attrition, Bayesian updating, and Darwinian selection in an industry will be discussed.

8.2 Sunk Costs and Barriers to Entry: The Stackelberg-Spence-Dixit Model

A fascinating aspect of sunk costs is their commitment value. A firm that buys equipment today signals that it

18. The Ghemawat-Nalebuff model assumes that each firm faces a flow cost of maintaining capacity, which is proportional to the firm's capacity (there is no fixed cost independent of productive scale). With $P(K, t)$ denoting the inverse demand function at time t, where $K = K_1 + K_2$ is industry capacity, and c denoting the maintenance-cum-production cost, firm i's instantaneous profit (assuming that both firms are still in at t) is

$$[P(K_1 + K_2, t) - c]K_i.$$

Assume that $\partial P/\partial K < 0$, and $\partial P/\partial t < 0$ (i.e., the industry is declining). Assume further that the firm's exit decision is lumpy (so a firm's capacity jumps directly from K_i to 0). Let t_i^* be defined by

$$P(K_i, t_i^*) \equiv c.$$

If $K_1 > K_2$, then $t_1^* < t_2^*$. That is, firm 1 would exit earlier than firm 2 in a monopoly situation. Backward induction shows that firm 1 exits first at time $t < t_1^*$ such that $P(K_1 + K_2, t) = c$, and firm two stays until t_2^*. (Hint: At date t_1^*, it is a dominant strategy for firm 1 to exit. At date $t_1^* - \varepsilon$, for ε small, firm 2 would be foolish to exit: At worst, it loses some profit during ε, and then becomes a profitable monopolist from t_1^* to t_2^*; so, assuming costly reentry, firm 2 stays and firm 1 exits.)

19. See Huang and Li 1986 and Fine and Li 1986 for analyses of the war of attrition when the profits follow a stochastic process.

20. The model of Ghemawat and Nalebuff is a continuous-time, continuous-capacity-adjustment one. That of Whinston assumes discrete periods and indivisible plants of equal sizes; it does not require almost continuous reduction, but uses a Markov-like assumption.

The following is a heuristic description of equilibrium. Consider the continuous-time model in note 18. Let $R(K_j, t)$ denote firm i's static reaction function at t; it maximizes

$$[P(K_i + K_j, t) - c]K_i$$

over K_i. Let $(K^*(t), K^*(t))$ denote the static Nash equilibrium, defined by $K^*(t) \equiv R(K^*(t), t)$. Under mild assumptions, $\partial R/\partial t < 0$, which implies that $dK^*/dt < 0$. Consider now the dynamic model and assume for simplicity that firms can only reduce capacity. The equilibrium strategies are: If

$$K_i(t) < R(K_j(t), t) \text{ for } i = 1, 2,$$

no firm reduces its capacity at date t. If

$$K_i(t) < R(K_j(t), t) \text{ and } K_j(t) \geq R(K_i(t), t),$$

firm i does not reduce its capacity; firm j stays on or moves to its reaction curve (that is, it reduces its capacity continuously if it is on its reaction curve and discontinously if it is above its reaction curve). If $K_i(t) \geq R(K_j(t), t)$ for $i = 1, 2$, both firms move to the static Nash equilibrium $(K^*(t), K^*(t))$. They then reduce their capacities so as to remain on their reaction curve. The equilibrium is nothing but a sequence of myopic (static) Cournot outcomes.

will be around tomorrow if it cannot resell the equipment. Thus, we may conjecture that the buying of equipment—if it is observed by one's rivals—may have strategic effects, and therefore is not a purely internal cost-minimization issue. Rivals may interpret the purchase of equipment as bad news about the profitability of the market and may reduce their scale of entry or not enter at all. The purpose of this section is to verify this conjecture.

For the modeling, we will need an explicitly dynamic model. Sunk costs are, by definition, a multiperiod phenomenon, as is entry deterrence. We will also introduce temporal asymmetries. Some firms will enter the market early, possibly because of a technological lead. We will see that these established firms (also called incumbents) accumulate a quantity of "capital" sufficient to limit the entry of other firms or even to make their entry unprofitable. First-mover advantages thus allow the established firms to restrict or prevent competition. We will think of "capital" as equipment or machines; however, as will be discussed later, the concept of capital can be interpreted more broadly.

8.2.1 Accommodated, Deterred, and Blockaded Entry

We start with a prototypical model whose extremely simplistic structure allows us to highlight the concept of a barrier to entry. This model is due to Heinrich von Stackelberg (1934).

Consider a two-firm industry. Firm 1 (the existing firm) chooses a level of capital K_1, which is then fixed. (We shall return to this assumption later.) Firm 2 (the potential "entrant") observes K_1 and then chooses its level of capital K_2, which is also fixed.

Assume that the profits of the two firms are specified by

$$\Pi^1(K_1, K_2) = K_1(1 - K_1 - K_2)$$

and

$$\Pi^2(K_1, K_2) = K_2(1 - K_1 - K_2).$$

These functions will be interpreted later. (Recall from chapter 5 that they are the reduced-form profit functions that come from short-run product-market competition with given capacities.) For the moment, note that these functions have two properties that are necessary for the generalization of the results to more general profit functions: First, each firm dislikes capital accumulation by the other firm ($\Pi^i_j < 0$). Second, each firm's marginal value of capital decreases with the other firm's capital level ($\Pi^i_{ij} < 0$). That is, the capital levels are strategic substitutes (see the introduction to part II).

For now, assume that there is no fixed cost of entry. The game between the two firms is a two-period one. Firm 1 must anticipate the reaction of firm 2 to capital level K_1. Profit maximization by firm 2 requires that

$$K_2 = R_2(K_1) = \frac{1 - K_1}{2},$$

where R_2 is the reaction function of firm 2 (that is, $R_2(K_1)$ maximizes $K_2(1 - K_1 - K_2)$ with respect to K_2). Therefore, firm 1 maximizes

$$\Pi^1 = K_1\left(1 - K_1 - \frac{1 - K_1}{2}\right),$$

from which we can determine the "perfect" Nash equilibrium:

$$K_1 = \tfrac{1}{2},\ K_2 = \tfrac{1}{4},\ \Pi^1 = \tfrac{1}{8},\ \Pi^2 = \tfrac{1}{16}.$$

Despite identical profit functions, firm 1 is in a position to obtain more profit than firm 2 by limiting the size of firm 2's entry. This illustrates the first mover's advantage. We know that if the two firms were to choose their levels of capital simultaneously, each would react to the other optimally, so that $K_2 = R_2(K_1)$ and $K_1 = R_1(K_2)$. Using the symmetry, the simultaneous-move solution yields

$$K_1 = K_2 = \tfrac{1}{3}$$

and

$$\Pi^1 = \Pi^2 = \tfrac{1}{9}.$$

The simultaneous-move and sequential-move outcomes are illustrated in figure 8.3. The broken lines represent the isoprofit curves. By definition of the reaction curves, firm 1's isoprofit curve is horizontal when it crosses R_1 and firm 2's isoprofit curve is vertical when it crosses R_2. To conform with common usage, S and N are used to denote the equilibrium outcomes in the sequential and the simultaneous game, respectively. They are usually called Stackelberg and Nash equilibria, but that terminology is actually misleading. The equilibrium concept is the same in both cases: (perfect) Nash equilibrium. The games simply differ in their timing. In the Stackelberg game, firm 1 has a chance to choose its level of capital before firm 2 and, therefore, to influence firm 2.

We conclude that temporal asymmetry allows firm 1 to limit firm 2's capital level. To do this, it accumulates more capital than it would have done in a simultaneous equilibrium. Consequently, the profitability of a marginal investment for firm 2 is diminished, providing an incentive for this firm not to accumulate too much capital. The intuition is the same for more general profit functions; by raising K_1, firm 1 reduces the marginal profit from investing (Π_2^2) for firm 2 (as long as $\Pi_{12}^2 < 0$). Thus, firm 2 invests less, which benefits its rival ($\Pi_2^1 < 0$).

The role of the irreversibility of capital levels (i.e., the fact that they may not be reduced in the future) should be stressed. Firm 1 is not on its reaction curve *ex post*; its best response to $K_2 = \frac{1}{4}$ is $K_1 = \frac{3}{8} < \frac{1}{2}$. If, after the choice of K_2, firm 1 could reduce K_1, it would do so. However, firm 2 would then choose $K_2 > \frac{1}{4}$ in anticipation of this response. In this sense, firm 1 loses by being flexible. The fact that the investment cost is sunk is a barrier to exit and allows the incumbent to commit to a high capital level.

Therefore, it is important that the capital investment be somewhat difficult to reverse if it is to have a commitment value. In particular, if the machines operated by the established firm may easily be resold on a second-hand market, then it will not satisfy this condition. The commitment effect is stronger the more slowly capital depreciates and the more specific it is to the firm (that is, when its resale involves large losses).

The value of commitment and the corresponding notion of "burning one's bridges" have widespread applicability beyond economics. An oft-quoted example is that of two armies wishing to occupy an island located between their countries and connected by a bridge to both (figure 8.4). Each army prefers letting its opponent have the island to fighting. Army 1, which is somewhat knowledgeable in game theory, occupies the island and burns the bridge behind it. Army 2 then has no option other than to let army 1 have the island, because it knows that army 1 has no choice other than to fight back if army 2 attacks. This is the paradox of commitment: Army 1 does better by reducing its set of choices.

The above equilibrium demonstrates how the incumbent (firm 1) can reduce firm 2's scale of entry. Following Caves and Porter (1977), we denote this as a *barrier to mobility*. We will also say that firm 1 *accommodates* entry, in that it takes entry for granted and simply tries to affect firm 2's subsequent behavior. Firm 1 cannot deter entry in this model. Firm 2 declines to enter ($K_2 = R_2(K_1) = 0$) only if $K_1 \geqslant 1$, which would yield negative profits to firm 1. In economic terms, this means that it is always worthwhile for firm 2 to enter, even on a small scale. If firm 1 makes positive profits, firm 2 can choose a small level of capital, hardly affect the market price, and make a profit itself.

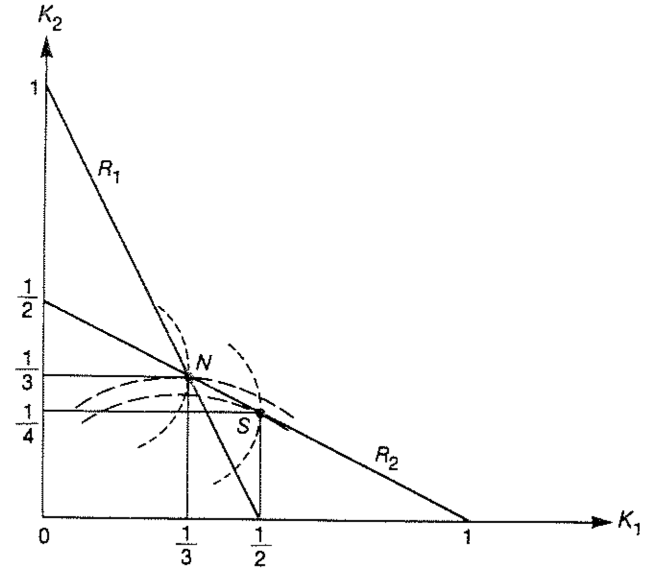


Figure 8.3
Stackelberg outcome.

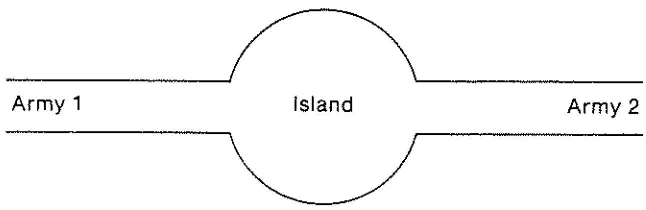


Figure 8.4

Such small-scale entry becomes unprofitable under increasing returns to scale. To illustrate the possibility of *entry deterrence*, let us introduce a fixed cost of entry, f, into our model. Assume that firm 2 has the following profit function:

$$\Pi^2(K_1, K_2) = \begin{cases} K_2(1 - K_1 - K_2) - f & \text{if } K_2 > 0 \\ 0 & \text{if } K_2 = 0. \end{cases}$$

Suppose that $f < \frac{1}{16}$. If firm 1 chooses $K_1 = \frac{1}{2}$ as before, firm 2 chooses $K_2 = \frac{1}{4}$ and makes a profit of $(\frac{1}{16} - f) > 0$.

However, this choice of K_1 may not be optimal for firm 1, which may be able to increase its profit by completely preventing the entry of firm 2. K_1^b, the capital level that discourages entry, is given by[21]

$$\max_{K_2}[K_2(1 - K_2 - K_1^b) - f] = 0,$$

or

$$K_1^b = 1 - 2\sqrt{f} > \tfrac{1}{2}.$$

Firm 2's reaction curve, depicted in figure 8.5, coincides with that in figure 8.3 up to K_1^b, and then coincides with the horizontal axis. When entry is deterred, the profit of firm 1 is

$$\Pi^1 = (1 - 2\sqrt{f})[1 - (1 - 2\sqrt{f})] = 2\sqrt{f}(1 - 2\sqrt{f}).$$

If f is close to $\frac{1}{16}$, this profit is greater than $\frac{1}{8}$. Therefore, firm 1 is interested in completely discouraging, not simply restricting, the entry of firm 2. Firm 1 accomplishes this by choosing $K_1 = K_1^b$, because accumulating beyond K_1^b would reduce profit (K_1^b is greater than the monopoly capital level of $\frac{1}{2}$).[22]

In Bain's terminology, the equilibrium for f a bit below $\frac{1}{16}$ is one of deterred entry, whereas the one for $f = 0$ (or, more generally, f small) is one of accommodated entry. With $f > \frac{1}{16}$, firm 1 blockades entry simply by choosing its monopoly capital level, $K_1^m = \frac{1}{2}$.[23]

*Exercise 8.3** Indivisibilities may, like a fixed cost, lead to a monopolistic structure if combined with a first-mover advantage. Suppose that firms must build an integer number of plants: 0, 1, 2, ... Building n plants costs $(3.5)n$. Each plant produces one unit of output, there is no variable cost, and the market price is $p = 6 - K$, where K is the industry's total capacity (number of plants).

(i) Show that a monopolist installs one plant.

(ii) Consider duopolists simultaneously choosing their numbers of plants, K_1 and K_2. Let $p = 6 - K_1 - K_2$. Show that in the Cournot equilibrium each firm builds one plant.

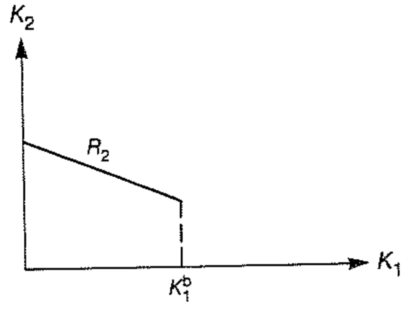


Figure 8.5
A fixed cost of entry implies a minimum capital level.

(iii) Suppose that firm 1 builds before firm 2. Show that firm 1 builds two plants and firm 2 stays out. Comment on the similarities and differences with the continuous-investment-*cum*-fixed-cost case.

8.2.2 Discussion and Extensions

8.2.2.1 Reduced-Form Profit Functions

We now return to the interpretation of the profit functions. Stackelberg actually wrote his two-stage game in terms of quantities. This left (at least) three questions unanswered: What does quantity competition mean? Why does one of the firms enjoy a first-mover advantage (i.e., choose its quantity first)? Why does quantity have a commitment value? Spence (1977, 1979) and Dixit (1979, 1980) made the Stackelberg story consistent, basically by interpreting Stackelberg's quantity variable as a capacity (as we did notationally). Doing so provides answers to the three questions: First, the profit functions represent reduced-form profit functions after one has solved for short-run product-market competition given the capacity levels. Second, the first-mover advantage may come from the fact that one of the firms obtains the technology earlier or is quicker to act than the other firm. Third, capacities have a commitment value to the extent that they are sunk.

Remark 1 In chapter 5, we derive the reduced-form profit functions by solving capacity-constrained price competi-

21. The superscript b stands for *barrier*.

22. Technically, $K_1 = \frac{1}{2}$ satisfies the first-order condition and the second-order condition locally for firm 1. However, since the reaction function of firm 2 is discontinuous at K_1^b, firm 1's objective function is not globally concave. Thus, $K_1 = \frac{1}{2}$ is not necessarily the maximum.

23. That the monopoly level coincides with the incumbent's capital level under accommodated entry is an artifact of the quadratic profit functions we have chosen.

tion. Spence and Dixit depart from this in two respects. First, they take short-run competition to be quantity competition rather than price competition. Second, they allow firms to accumulate more capacity during product-market competition. Consider Dixit's (1980) model. In the first period, firm 1 chooses capacity K_1 at cost $c_0 K_1$. This capacity may subsequently be increased, but cannot be reduced. Firm 2 observes K_1. Then, in the second period, the two firms choose their outputs (q_1 and q_2) and their capacities ($\tilde{K}_1$ and K_2) simultaneously, with $\tilde{K}_1 \geq K_1$. Production involves cost c per unit of output. Output cannot exceed capacity: $q_i \leq K_i$ for all i. The price is equal to the market-clearing price, given the outputs.

Firm 2 faces short-run and long-run marginal costs equal to $c_0 + c$ and clearly chooses identical capacity and outputs ($K_2 = q_2$). For $q_1 \leq K_1$, firm 1 incurs the short-run marginal cost c; any unit of output beyond K_1 costs the long-run marginal cost, $c_0 + c$. The short-run marginal-cost curve is represented in figure 8.6, which suggests why capacity has a commitment value: It lowers the *ex post* marginal cost of producing up to K_1 and hence makes the production of the first K_1 units attractive in the second period. To make this precise, we can consider two reaction functions. Were firms to choose their levels of capital and output simultaneously (that is, were there no first-mover advantage), they would face cost $c_0 + c$ for each unit of output at the date of the production decision. With the demand curve assumed linear ($p = a - bq$), firm i would maximize

$$q_i(a - b(q_i + q_j) - c_0 - c).$$

(There is obviously no point in accumulating capacity that is not used for production here.) Thus, the reaction function is

$$R_i(q_j) = (a - bq_j - c_0 - c)/2b.$$

Next consider the Dixit two-stage game in which firm 1 chooses its capacity in the first period and its production in the second period, whereas firm 2 chooses both capacity and production in period 2. Firm 2's reaction function at date 2 is

$$R_2(q_1) = (a - bq_1 - c_0 - c)/2b.$$

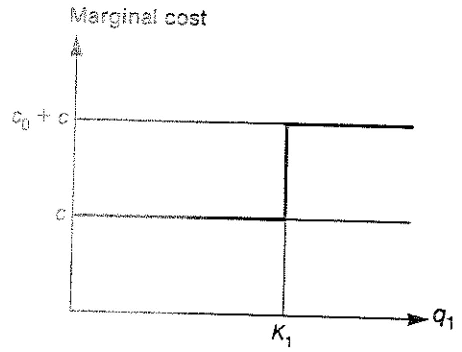


Figure 8.6
Short-run marginal cost.

However, firm 1 at date 2 has a different reaction function, called a *short-run* reaction function. Up to K_1, it incurs marginal cost c only, and therefore it has reaction function

$$\tilde{R}_1(q_2) = (a - bq_2 - c)/2b > R_1(q_2).$$

Beyond K_1, the short-run and long-run reaction functions coincide:

$$\tilde{R}_1(q_2) = R_1(q_2).$$

The second-period equilibrium as a function of K_1 can thus be obtained from the intersection of $\tilde{R}_1$ and R_2 in figure 8.7.

As can been seen from figure 8.7, firm 1 has no incentive to invest in the first period in capacity that it does not use *ex post*.[24] Furthermore, firm 1 gains by investing beyond the Nash capacity K_1^N, because this moves the equilibrium to the right of N along R_2, which increases firm 1's profit.

This does not explain how prices are determined. It is assumed that the market price "clears the market." For instance, the market price for linear demand $p = a - bq$ is

$$p = a - b(q_1 + q_2).$$

Assuming that the firms in the second stage produce to capacity, $q_i = K_i$ (this actually is part of the derivation), we can write the reduced-form profit functions

$$\Pi^i(K_i, K_j) = K_i(a - c_0 - c - b(K_i + K_j)),$$

24. There would be idle capacity if firm 1's short-run reaction curve intersected firm 2's reaction curve at some $q_1 < K_1$. Firm 1 could obtain the same product-market outcome by accumulating only q_1, thus saving $c_0(K_1 - q_1)$.

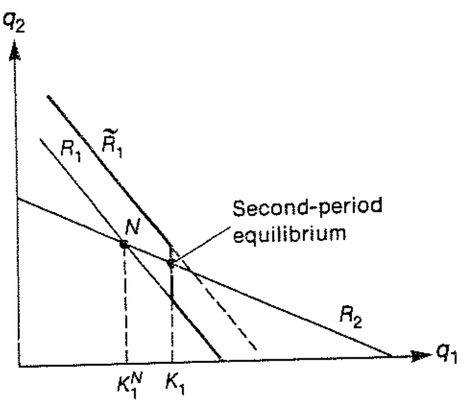


Figure 8.7
Short-run and long-run reaction functions.

which have the previous form for $a - c_0 - c \equiv 1$ and $b \equiv 1$.

As usual, the presence of an auctioneer is not completely satisfactory. A more realistic description of the Spence-Dixit game might involve a "double-capacity-constrained game." The first capacity constraint refers to the *production* capacity, which limits the level of output; the marginal cost is c as long as $q_i \leqslant K_i$. The second capacity constraint refers to the *selling* capacity, which limits the level of sales—firm i cannot sell more than it has produced: $x_i \leqslant q_i$, where x_i is the level of sales. This interpretation simply adds a third stage, in which firms choose prices constrained by their outputs.

A question studied by the literature is whether the incumbent firm uses its capacity after deterring entry. That is, does firm 1 hold idle capacity to deter firm 2's entry? Using quantity competition as the paradigm of short-run product-market competition, Spence answered in the affirmative. But Dixit showed that Spence's result was due to the fact that his equilibrium was not a perfect equilibrium.[25] Indeed, with a concave demand function, any capacity held to deter entry is used by the monopolist. Bulow et al. (1985a) show that Spence's excess capacity may reappear when the demand function is so convex that the reaction curves are upward sloping.

Schmalensee (1981) uses the Spence-Dixit model; however, instead of introducing a fixed cost of entry, he assumes that a firm cannot produce below some minimum level of output K_0 if it produces at all (so $q_i \geqslant K_0$). He interprets K_0 as a minimum efficient scale. Using empirical evidence that the minimum efficient scale is often low relative to industry demand (generally under 10 percent), he argues that such barriers to entry cannot account for established firms' high profits.

*Exercise 8.4*** The first part of this exercise recalls how reduced-form profit functions can be deduced from short-run price competition. The second part (inspired by Matsuyama and Itoh [1985]) shows how the model of barriers to entry can be used to analyze the desirability of protecting an infant industry.

(i) Two firms produce perfect substitutes at zero marginal cost (up to a capacity constraint). The demand function is $p = 4 - (q_1 + q_2)$. Firms are capacity constrained: $q_i \leqslant K_i$. Capacity costs 3 per unit. Use the monopoly solution to show that K_i cannot exceed 1. Use this upper bound to conclude that when firms are capacity constrained and choose their prices simultaneously (with the capacities fixed and common knowledge), both firms quote price $p = 4 - K_1 - K_2$; to show this, posit either the efficient-rationing rule or the proportional-rationing rule.

(ii) Firm 1 is a foreign firm, firm 2 a domestic one. Consider the following "no-protection" 3-stage game:

1. Firm 1 chooses capacity K_1.
2. Firm 2 chooses capacity K_2, knowing K_1.
3. Firms choose prices simultaneously, knowing K_1 and K_2.

(That is, the foreign firm has a first-mover advantage.) The domestic firm faces an entry cost $f = \frac{1}{16}$. Compute the equilibrium and the welfare (where welfare = consumer surplus + profit of domestic firm). Show that a policy of "limited protection," which forces the foreign firm to wait until period 2 to invest domestically (so that both firms choose K_1 and K_2 simultaneously) increases welfare.

Remark 2 Interpreting profit functions as reduced-form functions for price competition under capacity constraints allows us to perform some welfare analysis. In the Stackelberg example (see subsection 8.2.1), let $p = 1 - K$ denote the demand function, where $K = K_1 + K_2$ denotes industry capacity and output. (The intercept of the demand

25. See the Game Theory User's Manual on the notion of perfect equilibrium.

function is net of investment and production costs—see remark 1). The social optimum in this industry is to produce industry output $K = 1$. In duopoly, the welfare loss is measured by the area of the triangle between the demand curve and the marginal-cost curve (which, here, is the horizontal axis, as the marginal cost is normalized to zero); see chapter 1. If p is the market price, the welfare loss from monopoly or duopoly pricing is equal to $p^2/2$. If the entrant enters, the fixed cost f of entry must be added to the welfare loss, because the socially optimal production involves only one firm (the entrant does not bring any cost savings).

First, assume away the entry cost. The market price is higher when the two firms invest simultaneously ($p = \frac{1}{3}$) than when firm 1 invests before firm 2 ($p = \frac{1}{4}$). Thus, a social planner would not mind sequential entry.

The picture may be altered dramatically by the presence of an entry cost. The welfare loss is equal to $f + \frac{1}{18}$ if the two firms invest simultaneously.[26] When firm 1 invests first, the welfare loss is equal to $(2\sqrt{f})^2/2 = 2f$ if the fixed cost is sufficiently large that firm 1 deters entry.[27] Thus, the welfare loss is higher under sequential entry than under simultaneous entry if $f > \frac{1}{18}$ (and the reverse is the case if $f < \frac{1}{18}$ as long as entry is deterred). That the welfare analysis of entry deterrence is ambiguous should not be a surprise, because we know from chapter 7 that entry can result in biases in either direction. While the entrant takes the incumbent's capacity as fixed, the entrant's addition to the capacity of the industry is socially beneficial if it is privately beneficial from the nonappropriability-of-consumer-surplus effect (as long as the capacity is used). The incumbent's increase in capacity to deter entry also yields some increase in the industry's capacity without wasting the entry cost.

8.2.2.2 Multiple Incumbents

Several authors have studied either entry deterrence with several incumbents or models of sequential entry—Bernheim (1984), Gilbert and Vives (1986), McLean and Riordan (1985), Vives (1985), and Waldman (1987), among others.

One of the issues[28] addressed in this literature is whether entry deterrence is a public good. The one incumbent–one entrant model considered above suggests the following conjecture: To deter entry, the incumbent incurs a cost. With several incumbents, entry deterrence becomes a public good; if the first incumbent deters entry by accumulating a large amount of capital, the other incumbents also benefit. Every incumbent would like entry to be deterred but would prefer not to incur the associated cost.

To understand why underinvestment by the incumbents to deter entry might occur, it is useful to come back to the classic noncooperative subscription problem. Consider a community with two individuals. This community can implement a project that costs \$1. Each member of the community has value \$2/3 for the project. Thus, no one is willing to bear the whole cost; however, cooperative action is desirable, because the social value of the project, \$4/3, exceeds its cost, \$1. Suppose that the members choose simultaneously how much to invest in the project. If \$1 or more is collected, the project is implemented; otherwise it is not. (Any remaining money is redistributed according to some rule.) There are two kinds of pure-strategy Nash equilibria in this game. In the first, no one contributes and the project is not implemented. In the second, each member i contributes by subscribing an amount a_i such that $a_1 + a_2 = 1$ and the project is implemented. (There is a continuum of such equilibria, indexed by, say, a_1 in $[\frac{1}{3}, \frac{2}{3}]$, if the money is given back to the contributors because the project is not realized.)[29]

Now consider a situation in which two incumbents (firms 1 and 2) choose their capacities simultaneously. The entrant (firm 3) stays out if and only if $K_1 + K_2 \geq K^b$, where K^b is the entry-deterring industry capacity. The capacities K_1 and K_2 are analogous to the subscriptions in the preceding paragraph. It would seem that we face a public-good problem, with the possibility of too little

26. This assumes that both firms choose their Cournot outputs equal to $\frac{1}{3}$. Note that the entrant in this equilibrium makes profit $\frac{1}{9} - f > 0$. This is the only pure-strategy equilibrium under our assumption that $f < \frac{1}{16}$. (For $f \geq \frac{1}{16}$, there exists another equilibrium, with firm 1 producing its monopoly output, $\frac{1}{2}$, and firm 2 not entering.)

27. That is, $2\sqrt{f} - 4f > \frac{1}{8}$ or $f \geq 0.0054$. We assume as before that the incumbent uses its entry-deterring capacity, $2\sqrt{f}$. See remark 1 above.

28. Some policy interventions against entry deterrence are ambiguous. For instance, as Bernheim notes, making entry deterrence harder for the second firm entering a market reduces this firm's prospects and makes entry deterrence easier for the first firm in the market.

29. See exercise 11.6 in the Game Theory User's Manual for an underinvestment result with a continuous-size project choice.

aggregate investment from the point of view of the incumbents. Gilbert and Vives (1986) show that this intuition may be misleading. The reason is that, contrary to the usual public-good problem, supplying the public good (and thus contributing to entry deterrence) is not necessarily costly. Suppose that entry is deterred and $K_1 + K_2 = K^b$. Let incumbent i's profit be

$$K_i(P(K^b) - c_0 - c),$$

where $P(\cdot)$ is the inverse demand function and where c_0 and c are the investment and *ex post* variable costs per unit. Because the price must exceed the total unit cost, each firm would like to have the highest possible capital level for this given price. Thus, conditional on the actual deterrence of entry, each firm would like to contribute to entry deterrence as much as it can (in contrast with the public-good situation). Gilbert and Vives actually find that only overinvestment can occur. The exercise below suggests in more detail why this is so. For more general models, the conclusions are more ambiguous; see Waldman 1987 and McLean and Riordan 1985.

*Exercise 8.5**** Consider the two incumbents–one entrant game above. Let

$$\Pi^i = K_i(P(K_1 + K_2 + K_3) - c_0 - c)$$

denote firm i's profit for $i = 1, 2$, where

$$K_3 = \begin{cases} R_3(K_1 + K_2) & \text{for } K_1 + K_2 < K^b \\ 0 & \text{for } K_1 + K_2 \geqslant K^b \end{cases}$$

(firm 3 faces a fixed cost of entry). Let

$$\Pi^b \equiv K^b(P(K^b) - c_0 - c)$$

denote the industry profit when firms 1 and 2 just deter firm 3's entry. Show that if the noncooperative equilibrium between the incumbents allows entry, then $\Pi^1 + \Pi^2 \geqslant \Pi^b$ (so there is no underinvestment in entry deterrence by the incumbents).

8.2.2.3 Entry for Buyout

We assumed that the post-entry market organization takes the form of competition between the incumbent and the entrant (if the entrant enters). However, suppose that there are no impediments to mergers—i.e., there is no legal prohibition, there is no asymmetry of information about the value of assets, there are no direct costs of transferring assets, and it is possible for the asset seller to commit not to come back and reinvest in this market. The market structure may then be a monopoly if the incumbent buys the entrant or vice versa. Indeed, if mergers are costless, firms have an incentive to merge after the entrant enters, because a monopoly can do at least as well as duopolists as long as it owns the two firms' assets. Of course, the distribution of the gains from monopolization is determined in the bargaining process for buyout and depends on the "threat point," i.e., on the profits the two firms would make if they were to reach no agreement and compete in the product market. As long as the entrant has some bargaining power, he can extract part of the increase in industry profits associated with the merger. This means that, for a given investment (here, capacity), the possibility of a merger increases the entrant's post-entry profit. The bottom line is that the prospect of buyouts encourages entry. But we should note that the merger *ex post* increases market concentration. One example of a socially perverse effect is that the incumbent may buy the entrant's capacity and scrap part of it (that is, the incumbent may hold excess capacity after the merger). For more on these ideas, see Rasmusen 1987.

8.2.2.4 Uncertainty

Maskin (1986) extends Schmalensee's version of the Spence-Dixit model to allow for uncertainty about demand or short-run marginal cost. He argues that uncertainty forces the incumbent to choose a higher capacity to deter entry than he would under certainty. This increases the cost of entry deterrence, making it less likely.

8.2.2.5 Capital Accumulation

The basic model is very simplistic in that it assumes that firms can accumulate their capacities all at once. Furthermore, these capacities cannot be reduced and do not depreciate. In practice, capacities are accumulated and adjusted over time (possibly in a lumpy way, owing to technological indivisibilities). Capacity expansion imposes adjustment costs.[30] Furthermore, demand grows at the

30. See Prescott and Visscher 1980 for a model of internal organization that explains adjustment costs.

beginning of the product's life cycle, making early complete capacity accumulation a costly strategy. Thus, it is worthwhile to study capital-accumulation games, in which firms vie for a Stackelberg leadership position; see section 8.6.1.

8.2.3 Other Forms of Capital

We saw how physical capital may facilitate the erection of barriers to entry. Other kinds of capital may have the same effect if they have commitment value (that is, they are irreversible, at least in the short run). Consider the following three examples.[31]

• *Learning by doing* In certain industries, the experience acquired by the established firms during previous production periods reduces their current production costs and thus may be considered to be a form of capital. This experience gives the existing firms a competitive advantage, and therefore it can discourage others from entering. Indeed, certain consulting firms (the Boston Consulting Group, for example) have suggested that intense early production promotes learning by doing and thus can be used strategically for this purpose. The argument is, however, a bit less clear-cut than it seems, as will be shown in section 8.4.

*Exercise 8.6*** (i) A monopolist faces demand curve $q = 1 - p$ in each of two periods (A and B). Its unit cost is c in period A and $c - \lambda q^A$ in period B, where q^A is the first-period output (the firm learns by doing). The discount factor between the periods is $\delta = 1$. Show that the first-period output is $d/(2 - \lambda)$, where $d \equiv 1 - c$.

(ii) Suppose now that the monopolist (firm 1) faces an entrant (firm 2, with unit cost c) in the second period. They play Cournot (quantity) competition, which yields profits

$$\Pi_i^B = (1 + c_j^B - 2c_i^B)^2/9$$

and outputs

$$q_i^B = (1 + c_j^B - 2c_i^B)/3.$$

Write the first-order conditions determining q_1^A when (a) q_1^A is not observed by the entrant before second-period competition and (b) q_1^A is observed by the entrant. In which case is the monopolist's first-period output higher? (You need not compute q_1^A; just give the intuition and the interpretation in terms of business strategy.) What could change if the entrant were to face a fixed cost of entry?

• *Developing a clientele* The decision to develop a clientele is a capital decision that increases the demand for the product of the established firm. Clearly, if the clientele attached to the existing firm is considerable, the potential demand for the entrant is weak. This is well understood by firms that launch advertising and promotional campaigns not only to make their product known but also to "preempt" demand. The more imperfect the consumers' information and the more important the costs of switching suppliers, the greater the clientele effect.[32]

• *Setting up a network of exclusive franchises* This is a capital decision that increases the entrant's distribution costs.[33] The established supplier can assure himself of the services of the more capable franchisees by selecting them initially and imposing exclusivity on them.[34] Such an explanation is offered by some economists for the initial difficulty encountered by foreign producers attempting to enter the American automobile market (however, this argument is debatable because exclusive contracts are often of short duration).

The last two barriers—developing a clientele and franchising in the distribution network—are preemptive strategies. Two other important examples of such strategies are the following:

• Choosing a "strategic place" in a geographical or product space is often important because of its commitment

---

31. These are enunciations of the conventional wisdom, which may oversimplify reality. Two of the examples will be discussed in more detail below.

32. Considering a clientele as a form of capital suggests that the existing firm should overinvest to block the entry of other firms. Even though such a strategy may be possible, it is not necessarily optimal, for the following reason. If entry does occur, the established firm has two types of customers after entry: its own clients (over whom it still has monopoly power) and the other consumers (for whom it is competing with the entrants). Of course, the firm wants to set a high price for the captive clientele and a lower price for the other consumers. If it cannot price-discriminate, the firm must charge an intermediate price; this intermediate price is higher the more important is the captive clientele. Consequently, the existing firm is less aggressive after entry, when it has a large clientele; it has become a "fat cat," which may make entry profitable. Therefore, overinvesting in clientele may not necessarily be the best way to prevent entry. See Schmalensee 1983. (See also Baldini 1983, Fudenberg and Tirole 1984, and note 43 below.)

33. Salop and Scheffman (1983) include this type of strategic behavior in their category of behaviors that "raise the rival's costs."

34. See subsection 4.6.2.

effect (the fixed cost of establishing oneself cannot be easily recouped; "the firm is here to stay"). See subsection 8.6.2 for further discussion.

- A new product can preempt rival firms, especially when it is patented.

Preemption and the "race" to be first which it engenders are important concepts in the theory of imperfect competition.

- *The problem of "apparently innocent" behavior* From a theoretical viewpoint, it is possible to prescribe policies for government intervention in each situation of noncompetitive behavior by existing firms. Those responsible for fostering competition (antitrust authorities) are well aware that things are not so simple. They have a very difficult time proving that a certain type of behavior is detrimental to competition. In fact, they have less information than the firms about demand functions, cost structures, the quantities of accumulated capital, and so on. Government decision-makers face a dilemma. Certainly they cannot prosecute an existing firm for increasing the demand for its product by providing information to consumers, for decreasing its own costs by investing in R&D and in physical capital, or for accumulating experience. But how can we know if a firm has accumulated its "capital" in a totally innocent fashion? The problem is that most of the decisions that make a firm healthy also elevate it to a power position with respect to potential entrants.[35]

8.3 A Taxonomy of Business Strategies

The point of the Stackelberg model is that commitments matter because of their influence on the rivals' actions. In the capacity-accumulation game, the incumbent overinvests to force the entrant to restrict his own capacity. The goals of this section are to define the notions of "overinvestment" and "underinvestment" and, more generally, to supply a two-period framework within which to think of business strategies, including a taxonomy of possible strategies. The ideas that underlie this section have been known informally for a long time. Recently Fudenberg and Tirole (1984) and Bulow, Geanakoplos, and Klemperer (1985b) have independently offered a framework that systematizes these ideas.[36] The outcomes of many strategic interactions in industrial organization can be predicted using the basic framework of strategic effects in the simple two-period model.

Consider the following two-period, two-firm model. In period 1, firm 1 (the incumbent) chooses some variable K_1 (for example, capacity). We will call K_1 an investment, although, as we will see, that word must be taken in a very large sense. Firm 2 observes K_1 and decides whether to enter. If it does not enter, it makes zero profit. The incumbent then enjoys a monopoly position in the second period and makes profit

$$\Pi^{1m}(K_1, x_1^m(K_1)),$$

where $x_1^m(K_1)$ is the monopoly choice in the second period as a function of K_1 (for instance, x_1 is firm 1's output). If firm 2 enters, the firms make simultaneous second-period choices x_1 and x_2. Their profits are then

$$\Pi^1(K_1, x_1, x_2)$$

and

$$\Pi^2(K_1, x_1, x_2).$$

By convention, firm 2's entry cost is part of Π^2. These functions are assumed to be differentiable.

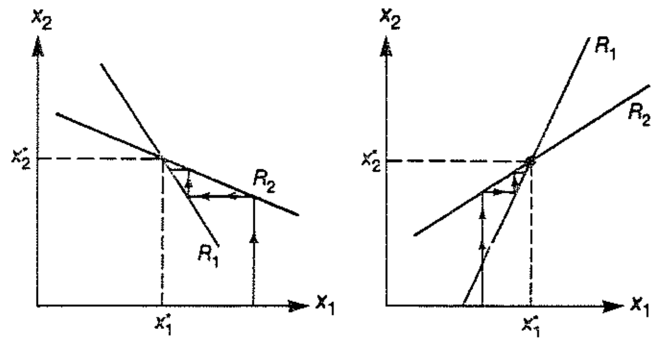
Suppose that firm 1 chooses some level K_1 (take it as given in this paragraph) and that firm 2 enters. The post-entry choices x_1 and x_2 are determined by a Nash equilibrium. The subsequent analysis of the effect of changes in K_1 on the Nash equilibrium assumes that this Nash equilibrium,

$$\{x_1^*(K_1), x_2^*(K_1)\},$$

is unique and stable. "Stability" has to do with the following thought experiment: Suppose that firm 1 picks an arbitrary x_1. Let firm 2 react by choosing an action $R_2(x_1)$ that maximizes $\Pi^2(K_1, x_1, x_2)$ over x_2. Then let firm 1 react to $R_2(x_1)$ by choosing an action $R_1(R_2(x_1))$ that maximizes $\Pi^1(K_1, \tilde{x}_1, R_2(x_1))$ over $\tilde{x}_1$. And so forth. This yields a sequential adjustment process in which both firms

35. In the next section we will see how actions by established firms have direct ("innocent") effects on their profits as well as strategic effects.

36. The terms *strategic complements* and *strategic substitutes* were coined by Bulow et al. The "animal" terminology is taken from Fudenberg and Tirole.



Strategic substitutes **Strategic complements**

Figure 8.8
Stable second-period equilibrium.

are myopic (i.e., they ignore the effect that their adjustment has on their rival; alternatively, they are rational but have discount factor $\delta = 0$). The Nash equilibrium,

$$\{x_1^*(K_1), x_2^*(K_1)\},$$

is stable if such an adjustment process converges to the equilibrium allocation from any initial position.[37] Stability is illustrated in figure 8.8.

Let us now consider the incumbent's first-period choice of K_1. We will say that entry is deterred if K_1 is chosen such that

$$\Pi^2(K_1, x_1^*(K_1), x_2^*(K_1)) \leqslant 0.$$

(This includes the case where entry is blockaded, that is, where the monopoly choice of K_1 deters entry.) Entry is accommodated if

$$\Pi^2(K_1, x_1^*(K_1), x_2^*(K_1)) > 0.$$

Which of the two cases must be examined depends on whether the incumbent finds it advantageous to deter or to accommodate entry. For simplicity, we will also assume that

$$\Pi^1(K_1, x_1^*(K_1), x_2^*(K_1)) \text{ and } \Pi^{1m}(K_1, x_1^m(K_1))$$

are strictly concave in K_1 and that the functions $x_i^*(\cdot)$ are differentiable.

8.3.1 Deterrence of Entry

We ignore the uninteresting case in which entry is blockaded (that case is void of strategic interactions). Thus, the incumbent chooses a level of K_1 so as to just deter entry[38]:

$$\Pi^2(K_1, x_1^*(K_1), x_2^*(K_1)) = 0.$$

Let us consider which strategy firm 1 can use to make firm 2's entry unprofitable. For this, let us take the total derivative of Π^2 with respect to K_1. From the second-period optimization,

$$\frac{\partial \Pi^2}{\partial x_2}(K_1, x_1^*(K_1), x_2^*(K_1)) = 0.$$

Thus, the effect of K_1 on Π^2 through firm 2's second-period choice should be ignored (this is the envelope theorem). Only two terms remain:

$$\frac{d\Pi^2}{dK_1} = \underbrace{\frac{\partial \Pi^2}{\partial K_1}}_{\substack{\text{Direct}\\ \text{effect}}} + \underbrace{\frac{\partial \Pi^2}{\partial x_1}\frac{dx_1^*}{dK_1}}_{\substack{\text{Strategic}\\ \text{effect}}}.$$

By changing K_1, firm 1 may have a *direct effect* on firm 2's profit ($\partial\Pi^2/\partial K_1$). For instance, if K_1 is the clientele accumulated by firm 1 before the entry of firm 2, a greater clientele reduces the size of the market and thus lowers firm 2's profit independent of any strategic effect. Often, however, $\partial\Pi^2/\partial K_1 = 0$. This is the case when K_1 is an investment that affects only firm 1's technology, such as the choice of a capacity or a technique.[39] Any effect on

37. For more on stability in oligopoly models, see Cournot 1838, Fisher 1961, Hahn 1962, Seade 1980, and Dixit 1986. The condition for local stability is $\Pi_{11}^1\Pi_{22}^2 > \Pi_{12}^1\Pi_{12}^2$. (Hint: Compare the slopes at the Nash equilibrium.) For a version in which firms behave rationally (that is, anticipate subsequent reactions and discount the future), see subsection 8.6.1.1.

38. From the continuity of Π^1 and Π^2 and the uniqueness of x_1^* and x_2^*, $x_1^*(K_1)$ and $x_2^*(K_1)$ are continuous in K_1 (from the "theorem of the maximum"). Hence, Π^2 is continuous in K_1. Suppose that

$$\Pi^2(K_1, x_1^*(K_1), x_2^*(K_1)) < 0.$$

Then firm 1 can increase or decrease K_1 a bit while still deterring entry (from the continuity of Π^2). This means that the constraint that firm 1 deters entry is not locally binding at the optimal K_1. From the concavity of Π^1 and that of Π^{1m}, the entry-deterrence constraint is not binding globally, which means that entry is blockaded (the case we ruled out).

39. Unless firm 1's investment bids up the price of investment goods for firm 2, or firm 1's investment has spillover or learning effects on firm 2.

firm 2's profit is then channeled through firm 1's post-entry choice. The *strategic effect* comes from the fact that K_1 changes firm 1's *ex post* behavior (by dx_1^*/dK_1), thus affecting firm 2's profits (in proportion to $\partial\Pi^2/\partial x_1$). The total effect of K_1 on Π^2 is the sum of the direct and strategic effects.

We will say that investment makes firm 1 *tough* if $d\Pi^2/dK_1 < 0$ and *soft* if $d\Pi^2/dK_1 > 0$.

Obviously, to deter entry, firm 1 wants to look tough. Now consider the following taxonomy of business strategies:

top dog: Be big or strong to look tough or aggressive.

puppy dog: Be small or weak to look soft or inoffensive.

lean and hungry look: Be small or weak to look tough or aggressive.

fat cat: Be big or strong to look soft or inoffensive.

If investment makes firm 1 tough, then firm 1 should "overinvest" to deter entry; that is, it should use the "top dog" strategy. If investment makes firm 1 soft, that firm should "underinvest" (i.e., stay lean and hungry) to deter entry.[40]

Example For simplicity, consider a slightly modified version of the Spence-Dixit model of section 8.2 (the same kind of reasoning holds for the original game). In this version, firm 1 chooses an investment K_1. This investment determines firm 1's second-period marginal cost $c_1(K_1)$, with $c_1' < 0$.[41] In the second period, firms 1 and 2 compete in quantities: $x_1 = q_1$, $x_2 = q_2$ (for the sake of exposition, we ignore firm 2's choice of investment). In the second period, firm 1 maximizes

$$q_1(P(q_1 + q_2^*) - c_1),$$

where P is the inverse demand function and c_1 is firm 1's marginal cost. A higher K_1 shifts firm 1's reaction curve to the right.[42] Assuming that quantities are strategic substitutes, the effect of an increase in K_1 can be represented as in figure 8.9. When firm 1's cost decreases, that firm has an incentive to produce more, which lowers the marginal value of output for firm 2. The new equilibrium involves a higher output for firm 1 and a lower output for firm 2. The main point, though, is that investment makes firm 1 tough (it raises q_1^*, which hurts firm 2). Hence, the "top dog" strategy is appropriate to deter firm 2's entry.

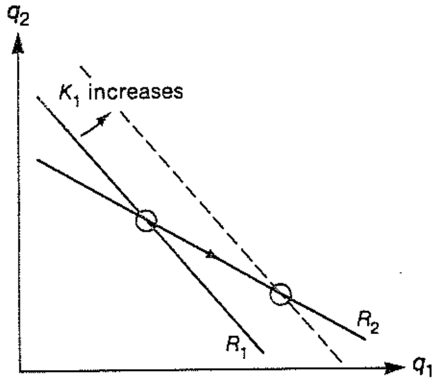


Figure 8.9
A firm's reaction curve moves outward with a decrease in marginal cost.

*Exercise 8.7** Suppose that, in the modified version of the Spence-Dixit game discussed above (where firm 1's investment reduces its marginal cost), the second-period competition is in prices. The two products are differentiated and are substitutes (see, e.g., the location model of chapter 7). Prices are strategic complements. Using a diagram, argue that firm 1 must overinvest to deter entry (assuming that entry is not blockaded).

Example A case was mentioned earlier in which K_1 was firm 1's pre-entry clientele. (One may, for instance, think of K_1 as firm 1's expenditures that makes switching costly

40. The concepts of over- and underinvestment can be characterized in an alternative way. Consider the hypothetical situation in which K_1 is not observed by firm 2 before the entry and second-period decisions. The corresponding equilibrium is usually called an *open-loop* equilibrium, because firm 2's strategy cannot be contingent on the actual choice of K_1, which is not observed at the date of decision. (A closed-loop strategy would depend on the actual level of K_1.) The open-loop case is an interesting benchmark against which to compare the effect of an observable change in K_1. If investment makes firm 1 tough, the equilibrium, entry-deterring level of K_1 exceeds the open-loop level (overinvestment); and conversely if investment makes firm 1 soft.

41. In section 8.2, the investment was a capacity level, and c_1 was not constant with output. But the important feature is that the investment reduces the marginal cost.

42. The proof of this is the same as the proof that a monopoly's optimal price increases with its marginal cost (see chapter 1).

to at least some of its customers.[43]) The direct effect of K_1 is to reduce firm 2's potential market ($\partial\Pi^2/\partial K_1 < 0$). However, the strategic effect has the opposite impact on firm 2's profit if firm 1 is not able to price-discriminate between its consumers; ideally, firm 1 would like to charge a high price to its captive clients and a low price to the noncaptive segment of the market, for which it is competing with firm 2. In the absence of price discrimination, however, an intermediate price is quoted, which intuitively increases with the size of the captive clientele. That is, a sizable clientele may make one a pacifistic fat cat, which is bad for entry deterrence. The overall effect $d\Pi^2/dK_1$ is thus ambiguous, and, depending on the parameters, either the "top dog" strategy or the "lean and hungry look" strategy may be appropriate to deter entry.

8.3.2 Accommodation of Entry

Suppose now that firm 1 finds deterring entry too costly. Whereas firm 1's first-period behavior in the entry-deterrence case was dictated by firm 2's profit, which had to be driven down to zero, it is dictated by firm 1's profit in the entry-accommodation case. The incentive to invest is given by the total derivative of

$$\Pi^1(K_1, x_1^*(K_1), x_2^*(K_1))$$

with respect to K_1.

From the envelope theorem, the effect on Π^1 of the change in firm 1's second-period action is of the second order. Thus, our basic equation in the entry-accommodation case is

$$\frac{d\Pi^1}{dK_1} = \underbrace{\frac{\partial\Pi^1}{\partial K_1}}_{\substack{\text{Direct}\\ \text{effect}}} + \underbrace{\frac{\partial\Pi^1}{\partial x_2}\frac{dx_2^*}{dK_1}}_{\substack{\text{Strategic}\\ \text{effect}}}.$$

Again, we can decompose this derivative into two effects. The *direct* or *"cost minimizing"* effect is $\partial\Pi^1/\partial K_1$. This effect would exist even if firm 1's investment were not observed by firm 2 before the choice of x_2, and therefore could not affect x_2. Thus, we will ignore this effect for the purpose of our classification. The *strategic* effect results from the influence of the investment on firm 2's second-period action. In the case of entry accommodation, we will say that firm 1 should overinvest (underinvest) if the strategic effect is positive (negative).[44]

The sign of the strategic effect can be related to the investment making firm 1 tough or soft and to the slope of the second-period reaction curve. To do this, assume that the second-period actions of both firms have the same nature, in the sense that $\partial\Pi^1/\partial x_2$ and $\partial\Pi^2/\partial x_1$ have the same sign. For instance, if the second-period competition is in quantities (prices), $\partial\Pi^i/\partial x_j < 0$ (>0). Using the fact that

$$\frac{dx_2^*}{dK_1} = \left(\frac{dx_2^*}{dx_1}\right)\left(\frac{dx_1^*}{dK_1}\right) = [R_2'(x_1^*)]\left(\frac{dx_1^*}{dK_1}\right)$$

by the chain rule, and arranging, we obtain

$$\text{sign}\left(\frac{\partial\Pi^1}{\partial x_2}\frac{dx_2^*}{dK_1}\right) = \text{sign}\left(\frac{\partial\Pi^2}{\partial x_1}\frac{dx_1^*}{dK_1}\right) \times \text{sign}(R_2').$$

43. For other examples of clienteles, see Schmalensee 1983, Baldini 1983, and Fudenberg and Tirole 1984. Of particular interest here are the switching-cost models of Klemperer and that of Farrell and Shapiro. Klemperer (1984, 1985a, b) analyzes the effect of switching costs in a two-period duopoly model in which there is competition in the first period and *ex post* monopoly (due to lock-in) in the second period. He shows how frequent-flyer discounts given by airlines to consumers in the first period to be used in the second period lead to weak price competition in the second period and may not benefit the consumers. (Frequent-flyer discounts differ from the most-favored-nation clause, discussed in section 8.4 below, in that the second-period discounts are not attached to the first-period price. They otherwise have similar collusive second-period effects.) More generally, the second-period rents stemming from switching costs induce intense first-period competition. Farrell and Shapiro (1987) introduce successive cohorts of consumers (via a model in which generations overlap) and show how a larger firm with a clientele may "milk" that clientele by charging a high price whereas a smaller firm charges a low price to attract young customers and build a clientele. For background on switching costs, see von Weizsäcker 1984 and the discussion of clienteles in chapter 2 of the present volume.

44. Again the concepts of over- and underinvestment can be characterized by comparing the optimal K_1 to the open-loop solution (i.e., the solution of the same game except that K_1 is not observable by firm 2 prior to its decision). See footnote 40. Given the concavity of Π^1, the optimal K_1 exceeds the open-loop solution if and only if the strategic effect is positive. Hint: In the open-loop solution, $\hat{K}_1$ is given by

$$\frac{\partial\Pi^1}{\partial K_1}(\hat{K}_1, x_1^*(\hat{K}_1), x_2^*(\hat{K}_1)) = 0.$$

This implies that for a positive strategic effect

$$\frac{d\Pi^1}{dK_1}(\hat{K}_1, x_1^*(\hat{K}_1), x_2^*(\hat{K}_1)) > 0.$$

This characterization does not generalize to the case in which both firms make decisions K_1 and K_2 in the first period. Even if both firms' strategic effect is positive, firm 2 (say) may invest less because firm 1 invests more and reduces the marginal value of firm 2's investment (this may occur, e.g., when the strategic effect is much stronger for firm 1).

The sign of the strategic effect, and therefore the over- or underinvestment prescription, is contingent on the sign of the strategic effect in the entry-deterrence case (which is equivalent to whether investment makes firm 1 tough or soft, when there is no direct effect in the entry-deterrence case) and on the slope of firm 2's reaction curve. We are thus led to distinguish four cases, depending on whether investment makes firm 1 tough or soft[45] and on whether second-period actions are strategic substitutes or complements (i.e., whether reaction curves are downward or upward sloping—see the introduction to part II). In all these cases, firm 1 tries to induce a softer behavior by firm 2 through its investment strategy.

- If investment makes firm 1 tough and the reaction curves are downward sloping, investment by firm 1 induces a softer action by firm 2; therefore, firm 1 should overinvest for strategic purposes (i.e., should follow the "top dog" strategy).
- If investment makes firm 1 tough and the reaction curves are upward sloping, firm 1 should underinvest (the "puppy dog" strategy) so as not to trigger an aggressive response from firm 2.
- If investment makes firm 1 soft and the reaction curves are downward sloping, firm 1 should stay lean and hungry.
- If investment makes firm 1 soft and the reaction curves are upward sloping, firm 1 should overinvest to become a fat cat.

These results, and those for the entry-deterrence case, are summarized in figure 8.10.

	Investment makes firm 1	
	Tough	Soft
Strategic complements ($R' > 0$)	A Puppy dog / D Top dog	A Fat cat / D Lean and hungry
Strategic substitutes ($R' < 0$)	A and D Top dog	A and D Lean and hungry

Figure 8.10
Optimal business strategies. (A stands for accommodation of entry, D for deterrence.)

Example

Consider the modified Spence-Dixit game. Firm 1's investment reduces its marginal cost. Second-period competition is either in prices or in quantities. As before, assume that prices are strategic complements and quantities are strategic substitutes. A reduction in marginal cost increases firm 1's output in the quantity game and reduces firm 1's price in the price game (see subsection 8.3.1).

In the quantity game, a higher output for firm 1 yields a lower output for firm 2. Firm 1 thus wants to overinvest—i.e., be a top dog. Thus, firm 1's strategy is the same whether it wants to deter or to accommodate entry, because being tough both hurts and softens firm 2 in the quantity game.

The picture is different in the price game. A lower price for firm 1 forces firm 2 to charge a lower price, which hurts firm 1. Thus, firm 1 should underinvest (i.e., keep a puppy-dog profile) so as not to look aggressive and trigger an aggressive reaction by firm 2. Firm 1's strategy is then very different depending on whether it wants to deter or to accommodate entry (deterrence calls for the "top dog" strategy), because being tough both hurts and toughens firm 2 in the price game.

At this point, one is likely to think: "I have a clear picture of the entry-deterrence case, in which firm 1 ought to overinvest. In the entry-accommodation case, however, the optimal strategy relies too much on the type of *ex post* competition (price or quantity); how can I make up my mind whether firm 1 should be a top dog or a puppy dog?" An element of the answer to this query can be found in chapter 5, where we interpreted quantity competition as capacity competition. To find the optimal strategy, we must wonder whether the investment K_1 reduces the marginal cost of accumulating capacity or that of producing. In the context of this model, we would thus predict strong strategic investment to accommodate entry when this investment reduces the costs of accumulating capacity. In contrast, a firm may be less eager to reduce

45. We assume that $\partial\Pi^2/\partial K_1 = 0$, so that we can identify "toughness" or "softness" with the sign of the strategic effect in the entry-deterrence case. If $\partial\Pi^2/\partial K_1 \neq 0$, the taxonomy under accommodation is relative to the sign of this effect rather than to "toughness" and "softness."

production costs and trigger tough price competition under entry accommodation.

8.3.3 Inducement of Exit

The above model treats only entry deterrence and accommodation. What about firm 1's incentive to invest in period 1, supposing that firm 2 is in the market at that date and must decide whether to stay or to exit in period 2? Inducing exit is very similar to deterring entry. In both cases, firm 1 wants to make firm 2 unprofitable in the second period. That is,

$$\Pi^2(K_1, x_1^*(K_1), x_2^*(K_1)) \leqslant 0$$

is the relevant objective for firm 1, where Π^2 includes entry or exit costs. Thus, firm 1's behavior is driven by firm 2's profit, and the strategic taxonomy is identical in both cases. In particular, D can be replaced by "D or E" in figure 8.10, where E stands for exit inducement.

8.4 Applications of the Taxonomy

We now turn to some applications of section 8.3. Other applications will be given in chapters 9 and 10.[46] We start with two examples of entry accommodation encountered in chapters 5 and 7. We then consider some new ones. We treat these examples mostly in an informal manner; the emphasis is on explaining how often optimal business strategies and market performance can be predicted from educated guesses. (References are given for more formal analyses.)

In section 8.3, K_1 was interpreted as an investment. More generally, it could be any action taken prior to date-2 competition; for instance, in example 4 below, K_1 refers to whether the firm offers a most-favored-customer clause and at what price. What matters is whether this action is observed by firm 2 and whether it makes firm 1 tough or soft in the second-period competition. Actually, K_1 need not even be an action taken by firm 1; it can be any variable that influences date-2 competition. In example 5 below, K_1 refers to firm 1's presence in another market. In example 6, K_1 denotes some variable outside the control of firm 1 (a quota, a tariff, or a subsidy). Again, the taxonomy can be applied as long as we can determine whether K_1 makes firm 1 tough or soft. The only modification is verbal: When K_1 is not controlled by firm 1, the "over- or underinvestment" prescription is replaced by the prescription that a higher K_1 benefits or hurts firm 1. We can also enlarge the set of applications to cases in which all firms play strategically in period 1, or to multiperiod games. The simple model of section 8.3 is again the key to understanding these slightly more complicated models.

In the following applications, we will assume that prices are strategic complements and that quantities (i.e., capacities) are strategic substitutes. This crucial assumption will be discussed in section 8.5.

Example 1: Voluntary Limitation of Capacity

In chapter 5 we analyzed a two-stage (accommodation) game in which firms accumulated capacities and then charged prices. We observed that firms accumulate noncompetitive amounts of capacity (under some circumstances, the Cournot levels). Prior capacity accumulation was seen as one way out of the Bertrand paradox. What prevents a firm from accumulating a large amount (a competitive level) of capacity is that by accumulating a small capacity, each firm signals that it will not play an aggressive price strategy, and there is no point to cutting the price if one cannot satisfy demand. This signal softens the pricing behavior of the firm's rivals. Such a voluntary limitation of capacity is an instance of "puppy dog" behavior. Gelman and Salop (1983) make this point nicely. They consider a model in which an entrant enters on a very small scale so as not to trigger an aggressive response by a large-capacity incumbent.[47] (The entrant is the strategic player in this example.) As Wilson notes,

46. See Shapiro 1986 for a useful and more extensive list of applications.

47. This is similar to the Stackelberg follower behavior in section 8.2. The game considered by Gelman and Salop is, however, different. The entrant, firm 2, chooses both a capacity (K_2) and a price (p_2). The incumbent, firm 1, has no capacity constraint, and chooses price p_1 after observing K_2 and p_2. Clearly, firm 2 does not pick p_2 above the monopoly price p^m, because firm 1 would then undercut to this monopoly price. Hence, when faced with $\{p_2 < p^m, K_2\}$, the incumbent's optimal strategy is either to undercut p_2 by ε (which firm 2 wants to avoid) or to charge $p_1 > p_2$ so as to maximize its profit given the residual demand. For instance, for the efficient-rationing rule (see chapter 5), the residual demand is $D(p_1) - K_2$. Firm 2 thus chooses $p_2 < p^m$ and K_2 so as to maximize $(p_2 - c)K_2$ subject to the no-undercutting constraint:

$$\max_{p_1} \{(p_1 - c)[D(p_1) - K_2]\} \geqslant (p_2 - c)D(p_2).$$

To make undercutting unattractive to firm 1, firm 2 chooses a low enough price and restricts its capacity. Gelman and Salop call this strategy "judo economics." Their paper also includes an interesting theoretical account of the 1979 coupon war among the major-airlines in the United States.

A useful example is an incumbent hotel in a resort location: an entrant that builds a comparable but small hotel with lower rates (a pension) can expect accommodation from a large hotel, since it is in the larger one's interest to serve the overflow from the smaller one rather than to cut its price to compete directly. . . . There are various ways to accomplish a commitment to capacity limitation. Besides a direct restriction on available supplies, the entrant can also tailor its product to a limited market segment. In the hotel example a menu of health foods might suffice. (1984, p. 41)

Example 2: The Principle of Differentiation

In chapter 7 we considered two-stage (accommodation) games of entry and location followed by price competition. We enunciated the principle of differentiation, according to which firms may not want to locate next to one another in the product space in order to avoid vigorous price competition. Here the first-period variable is location.

It is particularly instructive to recall the location game studied in section 7.1, in which two firms choose locations along a segment. When analyzing a firm's optimal decision, we saw that there are two effects: Moving toward the center of the segment increases the firm's market share and profit at given prices. This corresponds to the direct or cost-minimizing (here, profit-maximizing) effect of section 8.3. Moving toward the other firm increases the intensity of price competition. This strategic effect dictates that a firm locate as far as possible from the other firm. We saw that firms differentiate their products. (In our example, the strategic effect is so strong that maximal differentiation occurs.) Product differentiation is another instance of "puppy dog" behavior.[48] Identifying a firm's capital as the closeness of its location to the center of the segment, each firm is willing to accumulate less capital (i.e., locate away from the center) than it would do if its rival's action (price) were fixed.

Example 3: Learning by Doing[49]

It is often argued that experience effects can be used for strategic purposes. Indeed, in the 1970s some consulting firms recommended to their clients that they sacrifice short-run profits early in the product life cycle in order to gain strategic position, on the ground that by producing a lot early (i.e., by cutting its price) a firm can quickly slide down the learning curve and deter the entry (or at least restrict the expansion) of other firms—see, e.g., Boston Consulting Group 1972.

Learning by doing is similar to investing in technology or in capacity in that both reduce the firm's future cost. (Here we assume specific learning. Learning externalities are analyzed later.) There is an important difference, however, between learning by doing and other investments: The cost of learning by doing is not exogenous to the market, but rather follows from the firm's production experience.

For ease of exposition, let us consider a two-firm rivalry. Let each firm's second-period marginal cost decrease with its first-period output. Suppose first that competition takes place in quantities in both periods. By increasing its output in period 1, a firm signals that it will produce a higher output in period 2 because of the learning effect. With strategic substitutes, this reduces the other firm's output at date 2. Thus, the "top dog" strategy of accumulating experience early is optimal under accommodation and quantity competition. It is also optimal to deter entry, because the incumbent firm's lower second-period cost hurts the entrant.

Determining the optimal strategy under price competition is slightly more complex. The "top dog" strategy is still optimal for entry deterrence: By charging a low price today, the incumbent accumulates experience, which induces it to charge a low price tomorrow. Entry accommodation when there is only one firm in the market at date 1 (as in section 8.3) yields the opposite result: Experience induces a low price, which triggers a low price from the rival. The "puppy dog" strategy of underinvestment in experience (i.e., high first-period price) is then called

48. This is a bit loose, because the case of entry deterrence or exit inducement involves a direct effect: If $K_1 = a$ (in the notation of chapter 7), then $\partial \Pi^2/\partial K_1 < 0$. But $(\partial \Pi^2/\partial p_1)(\partial p_1^*/\partial K_1)$ is also negative, so we can identify "toughness" and the strategic effect.

49. The strategic aspects of learning by doing with and without spillovers have been analyzed by Spence (1981, 1984), Fudenberg and Tirole (1983a), Stokey (1986), and Mookherjee and Ray (1986), among others. The present discussion follows Fudenberg and Tirole 1986. This field owes much to the early analysis of Arrow (1962).

for. In contrast, accommodation when both firms are in the market at date 1 yields ambiguous results. On the one hand, a low price today increases the firm's output and hence its experience, making the firm aggressive tomorrow and triggering a low price from its rival (this is the previous strategic effect); however, a low price also reduces the rival's market share and, therefore, reduces its experience. The rival faces a higher second-period cost and, therefore, is less aggressive in that period. This second effect, which does not exist for quantity competition because a firm cannot affect its rival's current output, calls for the "top dog" strategy. It is not clear *a priori* which effect dominates.

To summarize: With specific learning by doing, the "top dog" strategy of accumulating lots of experience is optimal to deter entry or induce exit. It is also optimal for entry accommodation under quantity competition, but it may or may not be optimal under price competition. Thus, for entry accommodation it matters whether the learning by doing refers to a reduction in investment costs (quantity competition) or to a reduction in production costs (price competition).

Let us now consider the possibilities of diffusion of learning across firms (spillovers).[50] Such externalities can take place through interfirm mobility of employees, through spying, or through reverse engineering (i.e., taking a product apart to learn how it was built). Learning by doing then somewhat resembles a public good and is therefore likely to be undersupplied. The new strategic effect arising from the diffusion of learning runs counter to the "top dog" tendency associated with specific learning: No firm is willing to accumulate experience that helps its rival to reduce its cost and thus to be more aggressive.[51]

Example 4: Most-Favored-Customer Clause

A firm competing in price in an accommodation framework ought to look inoffensive so as not to force its rivals to cut price. It would thus like to take actions that commit it to charge a high price. As we saw earlier, this can be achieved by restraining investments that reduce production costs. There are, of course, other ways to commit to a high price. One way is to grant current customers a most-favored-customer status or price protection. (See Hay 1982 and Salop 1986. The analysis here relies more particularly on Cooper's [1986] formal treatment of such policies.)

The most-favored-customer policy guarantees a firm's current customers that they will be reimbursed the difference between the current price and the lowest price offered in the future (up to some specified date). For instance, in the 1960s and the early 1970s the two manufacturers of turbine generators, General Electric and Westinghouse, offered a price-protection policy effective during the six months following a sale.[52]

Before we consider why such a policy may help firms collude, it may be useful to recall the Stackelberg *price-leadership* story. Consider a duopoly producing differentiated products. Figure 8.11 depicts the reaction curves and the Nash (simultaneous-move) equilibrium (p_1^*, p_2^*). Suppose now that firm 1 chooses its price before firm 2. If it raises its price slightly above p_1^* to $\mathring{p}_1$, its profits are affected only to the second order by the fact that p_1^* is

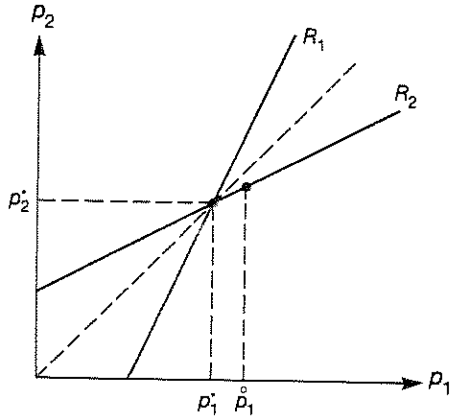


Figure 8.11
Stackelberg price leadership.

50. See Lieberman 1984 for evidence of the diffusion of experience in the chemical industry.

51. For instance, with linear demand and quantity competition, the firms' first-period output can be shown to decrease with the degree of diffusion by learning (in spite of the fact that diffusion increases total experience and therefore increases second-period output, making first-period learning more desirable). See, e.g., Fudenberg and Tirole 1983a.

52. These firms ended the practice as part of a settlement to avoid antitrust action.

an optimal response to p_2^* (according to the envelope theorem). To this direct effect can be added the indirect effect that firm 2 reacts by raising its price. This indirect effect raises firm 1's profit to the first order. Firm 1, the Stackelberg leader, thus chooses a price exceeding p_1^*.[53]

The Stackelberg story gives the intuition of why even unilateral commitments to a price-protection policy may be desirable. Consider a two-period duopoly price game. The firms choose their prices simultaneously in each period. The demand functions $D_i(p_i, p_j)$ and the cost functions (which we take as linear for notational simplicity) are independent of time. To simplify, there is no discounting between the two periods. In the absence of price-protection policies, the price equilibrium is the Nash equilibrium (p_1^*, p_2^*) in each period.

With price protection, the previous Nash equilibrium (without price protection) is no longer an equilibrium. That is, it would pay a firm to *unilaterally* impose price protection. To see this, suppose that firm 1 charges a price $\mathring{p}_1$ a bit above p_1^* in the first period *and* offers to reimburse the difference between $\mathring{p}_1$ and the price it will charge in the second period if the former exceeds the latter. Assume that the buyers behave myopically; that is, it is not because they expect firm 1 to lower its price and pay some cash back that they buy from that firm. (We will see later that these myopic consumers are actually rational, as firm 1 will not lower its price.) Thus, firm 1's first-period demand is

$$\mathring{q}_1 \equiv D_1(\mathring{p}_1, p_2^*).$$

Firm 1's second-period profit is thus

$$\mathring{\Pi}^1(p_1, p_2) = \begin{cases} \Pi^1(p_1, p_2) & \text{if } p_1 \geqslant \mathring{p}_1 \\ \Pi^1(p_1, p_2) - (\mathring{p}_1 - p_1)\mathring{q}_1 & \text{if } p_1 < \mathring{p}_1 \end{cases}$$

where $\Pi^1(p_1, p_2) \equiv (p_1 - c)D_1(p_1, p_2)$.

Thus, firm 1's marginal profit in the second period exhibits a discontinuity at $p_1 = \mathring{p}_1$. To see that

$$\{p_1 = \mathring{p}_1, p_2 = R_2(\mathring{p}_1)\}$$

is the second-period price equilibrium, draw firm 1's second-period reaction curve $\mathring{R}_1$. (Firm 2, which by assumption has not imposed price protection, has its usual reaction curve, R_2.) Whenever the optimal reaction to p_2 calls for $p_1 \geqslant \mathring{p}_1$ in the usual (no price protection) case, $\mathring{R}_1$ and R_1 clearly coincide. Let $\mathring{p}_2$ be such that $R_1(\mathring{p}_2) = \mathring{p}_1$. By definition of $R_1(\cdot)$,

$$\Pi_1^1(\mathring{p}_1, \mathring{p}_2) = 0.$$

This implies that

$$\Pi_1^1(\mathring{p}_1, \mathring{p}_2) + \mathring{q}_1 > 0.$$

It is thus easy to see that for $\mathring{p}_2 - \varepsilon$, firm 1 wants to react by $\mathring{p}_1$ rather than $R_1(\mathring{p}_2 - \varepsilon)$. This is also true for a range of prices for firm 2. Only when firm 2's price becomes very low will firm 1 cut below its first-period price and (with regret) bring the price-protection policy into play.[54] Firm 1's second-period discontinuous reaction curve is depicted in figure 8.12. A useful way to under-

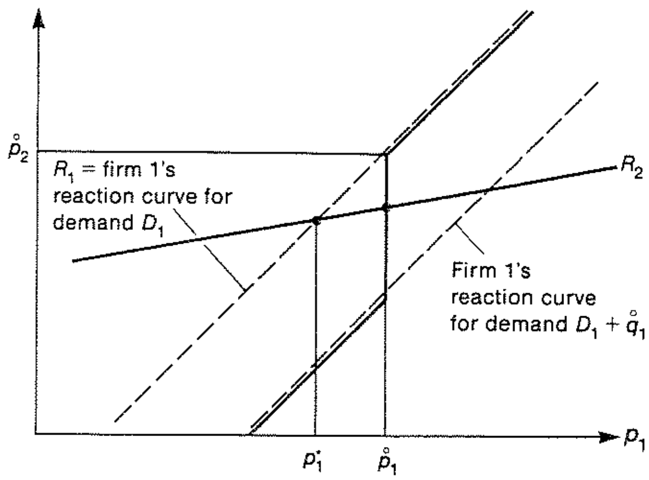


Figure 8.12
Second-period reaction curves when firm 1 offers price protection at p_1.

53. However, in contrast with quantity competition, being the leader may not be desirable. For instance, with symmetric profit functions, one has $R(p) < p$ for $p > p^*$. That the follower benefits more than the leader from the sequential timing follows from

$$\Pi(R(\hat{p}), \hat{p}) > \Pi(\hat{p}, \hat{p}) > \Pi(\hat{p}, R(\hat{p})),$$

where the first inequality comes from the optimality of the firm's reaction to $\hat{p}$ and the second inequality from the fact that each firm's profit increases with its rival's price.

54. This very low price, p_2', is given by

$$\Pi^1(\mathring{p}_1, p_2') = \max_{p_1 < \mathring{p}_1} [\Pi^1(p_1, p_2') - (\mathring{p}_1 - p_1)\mathring{q}_1].$$

stand this reaction curve is to notice that for $p_1 < \mathring{p}_1$

$$\mathring{\Pi}^1 = (p_1 - c)[D_1(p_1, p_2) + \mathring{q}_1] - (\mathring{p}_1 - c)\mathring{q}_1.$$

The second term in this expression is irrelevant as far as marginal choices are concerned, so everything is as if firm 1 faced demand $D_1(p_1, p_2) + \mathring{q}_1$ for $p_1 < \mathring{p}_1$. Thus, the reaction is that for demand $D_1(p_1, p_2) + \mathring{q}_1$ whenever the reaction falls strictly below $\mathring{p}_1$, and that for demand $D_1(p_1, p_2)$ whenever the reaction exceeds $\mathring{p}_1$.

Now recall that we choose $\mathring{p}_1$ just greater than p_1^*. From figure 8.12, the second-period price equilibrium is given by $p_1 = \mathring{p}_1$ and $p_2 = R_2(\mathring{p}_1)$.[55] In words: Firm 1 has succeeded in becoming the Stackelberg leader, thus driving firm 2's price up. Firm 1 increases its profit to the first order, which more than offsets the second-order loss in the first period. Thus, it pays for a firm to impose a price-protection policy, even if the other firm does not impose such a policy. By making future price cuts costly, a firm uses a profitable "puppy dog" strategy. It shifts down its second-period profit function (i.e., it becomes weak) in order to commit to a high second-period price (so as to look inoffensive). But, as in the Stackelberg price game, the firm that offers price protection gains less from this policy than its rival. See Cooper 1986 for the complete solution to the game (in equilibrium, either one or both firms will offer price protection).

Remark 1 Despite its strategic attractiveness, the most-favored-customer clause is not widespread. Several reasons can be found for this: (1) Rebates to other customers must be made observable to each buyer, because unrecorded rebates would benefit (*ex post* but not *ex ante*) a manufacturer who had offered price protection in the past. That is, discount secrecy removes the credibility of the price-protection policy. Other transaction costs include the cost of indexing the price to inflation and input costs. (2) The design of the good may be altered over time, so that again the price-protection policy has little applicability.[56] (3) The practice may face antitrust prosecution. (4) Price-protection policies are not very profitable when other firms threaten to enter the market. Indeed, a "puppy dog" strategy (here, the established firms' commitment to a high price) encourages entry. (5) Even in the accommodation case, the application of price-protection policies may be delayed by the fact that each firm wants to be the follower rather than the leader (as is the case when only one firm elects to offer the policy in the simultaneous-move game). This may give rise to situations similar to the war of attrition.[57]

Remark 2 Price protection is one method of softening future price competition. Another method (as in example 2) is to increase product differentiation. Klemperer (1984) has argued that discounts for repeated purchases increase the cost of customers' interbrand switching and thus differentiate the products in the future. This raises prices in the future. However, price competition is more intense at the beginning, because the value of a customer to a firm is raised.[58]

Example 5: Multimarket Oligopoly

The presence of a firm in one market may affect its strategic position in another market if the two markets are somehow related. This is the case when producing for two markets involves economies (or diseconomies) of scale or scope. Alternatively, the demands on the two markets may be interdependent.[59]

Bulow et al. (1985b) consider a duopoly model in which firms 1 and 2 are rivals in market 1 and firm 1 is a monopoly in market 2. (For concreteness, the two markets can be thought of as two different regions.) Here K_1, rather than a choice variable for firm 1, is a parameter related to profitability in market 2 (it can be thought of as a demand parameter). Bulow et al. show that an increase in firm 1's profitability in market 2 may actually reduce its total profit. This is because of the strategic effect in

55. The reaction curves for demands D_1 and $D_1 + \mathring{q}_1$ do not converge to each other (i.e., stay far apart) when $\mathring{p}_1$ converges to p_1^*.

56. As Cooper (1986) observes, General Electric and Westinghouse published books that contained relative prices for each component in order to face the issue that turbine generators are custom-made. They changed the prices by adjusting the multiplier.

57. Still another possibility is that firms compete in capacities. We know that results under accommodation are usually reversed with quantity competition. In this case, a firm wants to look tough (which a price-protection policy does not help it achieve). Checking this intuition would require solving the game in which firms choose capacities, prices, and protection policy.

58. See note 43 above.

59. For an example of multimarket rivalry with interdependent demands, see subsection 8.6.2 below.

market 1. Suppose that the firms compete in quantities, that firm 1's production cost depends on the sum of its outputs in the two markets, and that this technology exhibits decreasing returns to scale. All quantities (the two outputs for firm 1 and the output for firm 2) are chosen simultaneously and noncooperatively. Suppose that demand increases in market 2. This induces firm 1 to sell more in that market, which raises firm 1's marginal cost of production and lowers its output in market 1. Firm 2, observing the increase in demand in market 2, will infer that firm 1 will decrease its output in market 1, and so firm 2 will raise its output in market 1. In other words, the increased profitability of market 2 raises firm 1's marginal cost in market 1, which puts it at a strategic disadvantage (a "puppy dog" look is detrimental under strategic substitutes).[60] Similarly, if firms compete in prices and firm 1's technology exhibits increasing returns to scale, an increase in firm 1's profitability in market 2 lowers its marginal cost in market 1 and therefore makes it aggressive in this market, which triggers a low price by firm 2. Again, this strategic effect may offset the profitability increase for firm 1. Either with quantity competition and economies of scale or with price competition and decreasing returns to scale, an increase in the profitability of market 2 unambiguously increases firm 1's profit.

Example 6: Quotas and Tariffs

Strategic interaction in an international context is affected by countries' trade policies. Exercise 8.4 showed how a protection policy can help a domestic firm to gain an edge in a domestic capacity-accumulation contest with a foreign firm (when goods are costly to trade between countries). More generally, subsidies, tariffs, and quotas (which can be interpreted as the variable K_1 of the general model) may have a non-negligible impact on the strategic positions of foreign and domestic firms (Brander and Spencer 1984; Dixit 1984; Dixit and Grossman 1986; Eaton and Grossman 1983; Eichberger and Harper 1986; Krishna 1983; Krugman 1984).[61]

For instance, if a domestic and a foreign firm compete in quantities in the foreign market, an export subsidy induces the domestic firm to expand its output, which induces the foreign firm to contract its own. That is, the export subsidy makes the domestic firm a top dog (to its advantage). The following exercises develop other examples.

*Exercise 8.8** "A foreign firm that competes in prices with a domestic firm in the domestic market suffers from facing a quota." True or false?

*Exercise 8.9** Suppose that two firms, producing substitute but differentiated products, compete in prices. (The equilibrium is unique and "stable," and the profit functions are concave.) Show that a government-imposed floor on firm 1's price may increase that firm's profit. Explain.

Example 7: Vertical Control

The contracts signed between owners and managers or between manufacturers and their retailers influence competition between downstream units (managers or retailers) if these contracts are observable. For instance, Rey and Stiglitz (1986) show how exclusive territories may soften not only intrabrand competition but also interbrand competition. Exclusive territories may allow firms to behave like puppy dogs in a price game.[62] Bonanno and Vickers (1986) show that in duopoly a manufacturer may prefer to sell his product through an independent retailer rather than directly to consumers, in order to induce more friendly behavior from the rival manufacturer (see also McGuire and Staelin 1983 and Moorthy 1987). For some general results on the link between observable agency contracts and interbrand competition, see Fershtman and Judd 1986 and Katz 1987.

Example 8: Tying

Whinston (1987) reconsiders the old leverage theory, according to which tying may allow a firm with monopoly power in one market to monopolize a second market. His simplest model is as follows: Suppose there are two firms and two completely unrelated markets (the reasoning can be extended to the case in which the goods are comple-

---

60. See exercise 5.5.

61. See Itoh and Kiyono 1987 for other reasons why export subsidies may be desirable.

62. See review exercise 19. In a financial context, Brander and Lewis (1986) show that the contract between a bank and a firm affects market competition. In their model, a high level of debt makes a firm a top dog in a quantity competition. See also Mathewson and Winter 1985 for a strategic analysis of exclusive dealing.

ments). Market A is monopolized by firm 1. Consumers all have willingness to pay v for good A. Normalize the demand in this market to be 1 (as long as the price does not exceed v). Market B is a differentiated market and is served by firms 1 and 2. Let $q_i = D_i(p_i, p_j)$ denote firm i's demand in this market. For simplicity, assume that the consumers are the same and have unit demands in the two markets, so that $D_i(\cdot, \cdot) \leqslant 1$. Let c and c_1 denote firm 1's unit production costs in markets A and B. Does firm 1 have an incentive to tie its products? For simplicity, assume that firm 1 offers the two products either separately or together (i.e., that there is no mixed bundling, in the language of chapter 4).

Suppose first that firm 1 takes price p_2 as given. This situation arises when firms 1 and 2 choose their prices, and firm 1 chooses whether to tie its two products simultaneously. It is easily seen that firm 1 does not gain from tying the two goods—under tying, firm 1 offers the bundle at price P_1 so as to maximize

$$(P_1 - c_1 - c)D_1(P_1 - v, p_2),$$

as the fictitious price for its good in market B is $P_1 - v$. Where P_1^* denotes the optimal price, firm 1 can realize at least the tying profit by selling the two goods separately at prices v and $P_1^* - v$, respectively:

$$(v - c) + [(P_1^* - v) - c_1]D_1(P_1^* - v, p_2)$$
$$\geqslant (P_1^* - c - c_1)D_1(P_1^* - v, p_2),$$

where $D_1 \leqslant 1$. Absent strategic considerations (as is the case here, where firm 1 takes p_2 as given and acts as a monopolist on its residual-demand curve), tying generally hurts firm 1 by reducing the number of degrees of freedom in its pricing strategy. (We know from chapter 4 that this conclusion is not general, as a monopolist may gain from tying. On this, see also review exercise 27. But this modeling choice will make the conclusions more striking and help us to identify the strategic effect.)

An important property is that under pure bundling, the fictitious price $\tilde{p}_1 \equiv P_1^* - v$ is lower than the price p_1 in market B, under no bundling, for any p_2. That is, *bundling shifts firm 1's reaction curve westward in market B*. To see this, note that under bundling P_1 maximizes

$$(P_1 - c_1 - c)D_1(P_1 - v, p_2),$$

which means that $\tilde{p}_1$ maximizes

$$\{\tilde{p}_1 - [c_1 - (v - c)]\}D_1(\tilde{p}_1, p_2).$$

In the absence of bundling, however, p_1 maximizes

$$(p_1 - c_1)D_1(p_1, p_2).$$

(The constant term $(v - c)$ in firm 1's profit function can be ignored.) Thus, everything is as if bundling reduced firm 1's cost of producing in market B by $v - c$ as far as pricing in market B is concerned. This is very natural, as a unit loss of sales in market B costs $v - c$ to firm 1 in market A under bundling, so that the "real" marginal cost of selling in market B is reduced by $v - c$. Now, we know from chapter 1 that a monopoly price increases with marginal cost. The consequence of this is that a firm's reaction curve in oligopoly shifts outward when the marginal cost increases (because this firm is a monopoly on its residual-demand curve). For any p_2, therefore, $\tilde{p}_1 < p_1$. Furthermore, in this model bundling is formally identical to an investment in cost reduction. Firm 1 pays a fixed investment cost $v - c$ (which corresponds to the loss in revenue from selling in market A separately) for a fictitious technology that reduces its marginal cost in market B from c_1 to

$$c_1 - (v - c).$$

As the number of units sold in market B is generally lower than the number sold in market A ($D_1 < 1$), such an investment cannot be profitable in the absence of strategic considerations.

Suppose now that firm 1 decides whether to bundle before the two firms compete in prices. Thus, firm 1 first chooses to market the two goods separately (no bundling) or together (pure bundling); then the two firms choose their prices simultaneously. Think of a technological decision concerning the packaging of the product or (more likely in the case of complements) the decision whether to make the product intended for market A incompatible with firm 2's product.[63] Here bundling hurts not only firm 1 but also firm 2, as it commits firm 1 to charge a low fictitious price in market B. Thus, bundling is not a good strategy if firm 2's entry or exit decision is not at stake. It hurts firm 1 both directly and indirectly,

63. See example 9 for related arguments.

as it forces firm 2 to reduce its price. However, if firm 1 wants to deter the entry or induce the exit of firm 2, bundling may be profitable. In terms of our cost-reduction analogy, firm 1 may overinvest in cost reduction (i.e., bundle) to foreclose market B. Again, the "top dog" strategy is optimal under entry deterrence and the "puppy dog" strategy is optimal under accommodation. (The "puppy dog" strategy is actually the no-bundling strategy; firm 1 cannot underinvest in bundling here.)

We conclude, with Whinston, that technological precommitment to bundling has important strategic effects and may allow a firm to use the leverage provided by its power in one market to foreclose another market.

Example 9: Systems and Product Compatibility

This example, which is related to example 8, involves systems of complementary products (for example, computer hardware and software; cameras, lenses, and film processing; tape decks, amplifiers, and speakers). The products in these lines cannot be consumed separately, but they can be purchased individually—consumers can "mix and match" products as long as they are compatible. In contrast, a manufacturer that makes its system incompatible with other systems imposes a *de facto* tie-in (see chapter 4) of its various components.

Matutes and Regibeau (1986) analyze the compatibility decisions of two producers of competing lines. They consider a duopoly in which each firm produces two complementary products, X and Y, which constitute a system. Their model extends the differentiation on the line model of chapter 7 to a two-dimensional case. The consumers are uniformly located on a square of dimension 1, and the firms' products are at diametrically opposed locations. Firm 1 is located at the origin and firm 2 at the point (1, 1). (See figure 8.13.) A consumer located at the coordinates (x_1, y_1) is $tx_1 + ty_1$ away from his preferred system when buying firm 1's system, where t is a taste parameter (the analog of a "transportation cost"). Thus, if p_1 is the price of the system (i.e., the two bundled components) sold by firm 1, the generalized price paid by the consumer for this system is

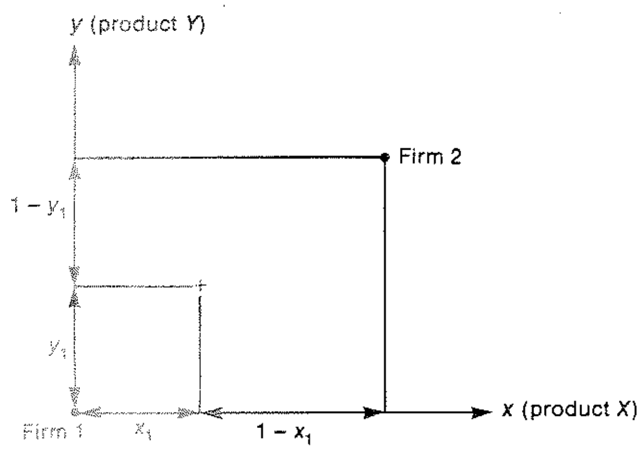


Figure 8.13

$$\tilde{p}_1 = p_1 + t(x_1 + y_1).$$

Similarly, the generalized price for the system sold by firm 2 is

$$\tilde{p}_2 = p_2 + t[(1 - x_1) + (1 - y_1)].$$

Under incompatibility, the consumer chooses the system with the lower generalized price, $\tilde{p} = \min\{\tilde{p}_1, \tilde{p}_2\}$. His demand is then downward sloping: $q = a - b\tilde{p}$.[64] If the systems are compatible and the products are sold separately, the consumer can mix and match. For example, the generalized cost of buying product X from firm 1 and product Y from firm 2 is

$$p_1^X + p_2^Y + t[x_1 + (1 - y_1)],$$

where p_1^X and p_2^Y are respectively the price charged by firm 1 for product X and the price charged by firm 2 for product Y. By mixing, the consumer can choose among four systems—*product variety is increased.* Again, the consumer chooses the system with the lowest generalized price.

Thus, as in chapter 7, the demand function for each firm's system (under incompatibility) or each firm's product (under compatibility) can be determined as a function of the prices. One can then solve for the Nash equilibrium in prices. Suppose, for simplicity, that the unit production cost is the same for both firms and both products. The

64. In chapter 7 we assumed that demand was equal to 1 (although we could just as well have considered the case of a downward-sloping demand at each location). For audio systems, the demand function is mainly a unit demand per consumer. The way to justify a downward-sloping demand *at a given location* is to envision a large number of consumers with different tastes for the system at this location.

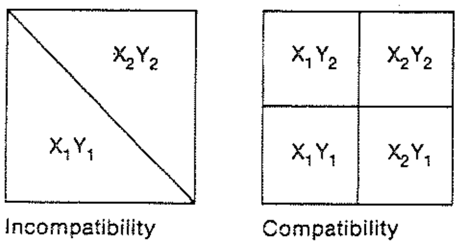


Figure 8.14

equilibrium demands for each product are symmetric and are as represented in figure 8.14 (assuming that the whole market is covered in equilibrium). In the figure, $\{X_i, Y_j\}$ means that the consumer consumes good X from firm i and good Y from firm j. Under compatibility, the consumers located in the northwest and in the southeast corner of the square buy a system that is better suited to their preferences than the system they would buy under incompatibility.

What are the incentives to achieve compatibility? First, compatibility raises demand, because it makes products better adapted to the consumers' tastes. Second, compatibility softens price competition, as Matutes and Regibeau show. To see the latter, note that when firm 1 decreases its price for good X_1 it increases the demand for systems that include X_1 (because the generalized prices for those systems decrease). Under incompatibility, the only system that includes X_1 is X_1Y_1. (Properly speaking, a decrease in the price of X_1 is then equivalent to the same decrease in the system's price.) Thus, firm 1 enjoys the full benefit associated with the increase in demand. Under compatibility, there are two systems that include X_1 (X_1Y_1 and X_1Y_2), so some of the benefit from increased demand accrues to firm 2. This noninternalization of part of the increase in demand reduces firm 1's incentive to cut its price.[65] Thus, the firms price their components less aggressively than they would do if the components were bundled in an incompatible system. These two effects imply that *firms have a common interest in achieving compatibility.*[66]

Remark The desire for compatibility stems from the assumption that the firms accommodate each other (i.e., do not try to force each other out of the market). In contrast, we know from Whinston's version of the leverage theory (see example 8) that tie-ins may serve as a barrier to entry. Similarly, a dominant firm that wants to induce its rival's exit might well want to make its products incompatible with those of the rival.[67] Incompatibility hurts the rival in two ways: It reduces demand and it leads to more aggressive price competition. It is therefore apt to induce exit. Thus, a firm's optimal strategy (here, concerning the compatibility decision) again hinges on whether it wants to accommodate its rivals or to deter entry or induce exit.[68]

8.5 Epilogue: Prices versus Quantities

A crucial assumption in the interpretation of examples 1 through 9 is that prices are strategic complements and quantities are strategic substitutes. This characterization is particularly crucial in accommodation games, where firms wanting not to look aggressive when they compete in prices may take actions that will later turn them into puppy dogs and where firms competing in quantities may try to become top dogs in the future. It is therefore not surprising that two-period price games (respectively, quantity games) are often more collusive (respectively, more competitive) than their static (one-period) counterparts.[69] The strategies in an entry-deterrence or an exit-inducement situation usually differ less between price and quantity games than in an entry-accommodation situation. As we saw in section 8.3, the important thing is then

65. This effect is reminiscent of the observation in chapter 4 that producers of complementary products tend to charge prices that are too high from an industry viewpoint. The incompatible case makes systems fairly good substitutes, whereas compatibility introduces some complementarity.

66. The welfare analysis is less clear cut. In particular, the social welfare relative to consumers purchasing X_1Y_1 and X_2Y_2 under compatibility has gone down, as they face a higher price for their systems and their product selection is the same as under incompatibility. The consumers purchasing X_1Y_2 and X_2Y_1 under compatibility buy a more suitable system than under incompatibility but also pay a higher price, so the welfare analysis is ambiguous without further assumptions.

67. It has been alledged that IBM makes its products incompatible with those of its rivals in order to maintain dominance.

68. See Ordover and Willig 1981 for a discussion of predatory incompatibility decisions.

69. The infinite-horizon Markov analysis of Maskin and Tirole (1987, 1988b) also emphasizes the role of the cross-partial Π^i_{ij}, and suggests that these results are somewhat robust. There, repetition yields a collusive outcome in the price game and a more competitive outcome than the Cournot one in the quantity game.

to look tough. For instance, by reducing costs, one hurts one's rival, whether competition is in price or in quantity. The bottom line is that, before applying the above taxonomy, one should look at the microstructure of the industry and determine the type of competition that is being waged.

The characterization of prices and quantities as strategic complements and substitutes is a presumption, not a general law, as the following will show.

Quantities Assume that profits have the exact Cournot form (see chapter 5)

$$\Pi^i(q_i, q_j) = q_i P(q_i + q_j) - C_i(q_i),$$

where C_i is firm i's cost function. A simple computation yields the cross-partial derivative:

$$\Pi^i_{ij} = P' + q_i P''.$$

We already know that $P' < 0$. To obtain the strategic-substitute property, it suffices that the price function be linear ($P'' = 0$) or concave ($P'' < 0$). The property may fail for sufficiently convex price functions.[70]

Prices Let $q_i = D_i(p_i, p_j)$ denote the demand curves. The profit functions are

$$\Pi^i(p_i, p_j) = p_i D_i(p_i, p_j) - C_i(D_i(p_i, p_j)).$$

This yields the cross-partial derivative:

$$\Pi^i_{ij} = \frac{\partial D_i}{\partial p_j} + (p_i - C_i')\frac{\partial^2 D_i}{\partial p_i \partial p_j} - C_i'' \frac{\partial D_i}{\partial p_i}\frac{\partial D_i}{\partial p_j}.$$

As in the case of quantities, this cross-partial derivative depends on the details of the demand function. Assume that the demand is linear (over the relevant range),

$$D_i(p_i, p_j) = a - bp_i + dp_j,$$

and that the marginal cost is constant. If the goods are demand substitutes ($d > 0$), then $\Pi^i_{ij} > 0$, so the goods are strategic complements. If they are demand complements ($d < 0$), then they are strategic substitutes. More generally, if we assume that the goods are demand substitutes, and we note that $p_i - C_i' > 0$ in equilibrium (from firm i's first-order condition), it suffices that $\partial^2 D_i/\partial p_i \partial p_j$ be non-negative in order for the goods to be strategic complements in the neighborhood of a price equilibrium.[71]

70. For instance, Bulow et al. (1985b) note that for $P(q_1 + q_2) = (q_1 + q_2)^{-\alpha}$ where $0 < \alpha < 1$, Π^i_{ij} is proportional to $\alpha - q_j/q_i$. Thus, if because of cost differences the equilibrium involves a big firm and a small firm (q_1/q_2 very large, say), quantities are strategic complements for one firm and strategic substitutes for the other near the equilibrium point. In particular, an increase in the small firm's output raises the big firm's optimal reaction to this output.

71. It is easy to construct examples in which this property is not satisfied. What is more, the goods generally are not strategic complements in prices over the whole range of potential prices, as Maskin and Tirole (1988b) note. To see why, suppose that the goods are fairly good demand substitutes. Fix p_j and let p_i vary. When $p_i \gg p_j$, firm j obtains the whole demand, and firm i's demand and profit are not affected much by a unit change in firm j's price. (In the limit, with perfect substitutes, demand remains 0 and thus is not affected at all.) So Π^i_j is very small. Similarly, when $p_i \ll p_j$, a unit change in p_j has little effect on firm i's demand and profit; again, Π^i_j is very small. When p_i is close to p_j, a unit change in p_j has a big effect on firm i's demand and profit (think of perfect substitutes); hence, Π^i_j is big. Thus, Π^i_j cannot be monotonic in p_i. Now, this did not matter in our applications, because the second-period simultaneous-move price equilibrium occurred in the region where $\partial D_i/\partial p_j$ is large and Π^i_{ij} is positive. In more dynamic games, this may have some relevance. For instance, in Maskin and Tirole 1988a,b the reaction curves are monotonic (downward sloping) in the quantity game and nonmonotonic in the price game.

8.6 Supplementary Section: Strategic Behavior and Barriers to Entry or Mobility

This section, which covers some of the recent research on barriers to entry, serves two purposes. First, on a technical level, it goes beyond the somewhat contrived two-period model of sections 8.2–8.4 to analyze the full-fledged dynamic interaction between firms. Second, and perhaps more important, it studies in detail two distinct barriers to entry. Section 8.6.1 compares short-term and long-term capital accumulation. (The analysis follows Fudenberg and Tirole 1986 and 1987.) Section 8.6.2 deals with differentiated markets; it shows how a firm may want to preempt its rivals to occupy the profitable market niches and how a firm can use product proliferation to restrict entry.[72]

8.6.1 Capital Accumulation

The commitment value of capital is higher the longer its lifetime and the more costly its disposal or resale. Thus, the extent to which capital is sunk determines the monopoly power and profit enjoyed by established firms. We will examine two polar cases here: one in which investments are sunk only in the very short run and one in which capital cannot be resold and does not depreciate (i.e., is completely sunk).

8.6.1.1 Short-Term Capital Accumulation and Contestability

This subsection explores two related investment-based dynamic models of a natural monopoly. In these models there is room for only one firm in the market, and there is actually a single firm in equilibrium. This firm makes a profit and deters entry through capital accumulation. Capital is sunk only in the short run and must be "renewed" periodically. The length of time over which capital is sunk determines the period of commitment. When commitment is short, the established firm enjoys only small incumbency advantages over potential entrants (because an entrant can kick the incumbent out of the market quickly). Thus, it must accumulate capital to deter entry. In the limit for very short commitments, the incumbent firm makes almost no profit, so Posner's rent-dissipation postulate (which says that monopoly profit is dissipated through competition—here, potential competition) is satisfied for very-short-run commitments. Posner's wastefulness postulate (according to which profits are dissipated in a socially wasteful way) may or may not hold, depending on whether the incumbent's capital is excess capital or contributes to production.

Wasteful Rent Dissipation

The first theory of short-run commitments, developed by Eaton and Lipsey (1980), considers an industry with two firms. Time is continuous, and the horizon is infinite. One unit of capital (e.g., a plant) is necessary for production and gives access to constant marginal cost, c. A second unit of capital is useless in the sense that it does not reduce the marginal cost of production. One unit of capital costs f per unit of time and has deterministic durability H (after the unit of capital is installed, it undergoes no physical depreciation for H units of time and full depreciation thereafter[73]). The fixed cost of production (equal to $\int_0^H fe^{-rt}dt$, where r is the rate of interest) is paid when the unit is installed, so the firm cannot avoid paying the fixed cost by leaving the market before H units of time have elapsed. Therefore, with equipment of age $\tau < H$, the firm never has an incentive to leave the market, even if another firm enters. Thus, H is a measure of commitment.

If at date t only one firm is active (i.e., has at least one unit of capital), that firm's flow profit, gross of capital cost, is

$$\tilde{\Pi}^m = \max_q [P(q)q - cq].$$

Suppose that $f < \tilde{\Pi}^m < 2f$. A monopoly is feasible, because $\tilde{\Pi}^m > f$. If two firms operate (i.e., have at least one unit of capital each), they wage Bertrand competition with marginal cost c and make zero gross profit; thus, each loses f per unit of time. The Bertrand assumption is meant to simplify computations. More generally, the firms could make a positive gross duopoly profit; the assumption $\tilde{\Pi}^m < 2f$ would still guarantee a negative net profit for at least one of them, since the gross monopoly

72. Readers not familiar with dynamic games may want to skip subsection 8.6.1 in a first reading; it is technically more difficult than the rest of the section.

73. This is the "one-horse shay" manner of depreciation.

profit is an upper bound on gross industry profits under duopoly.

The firms' sole decision is when to build units of capital. One firm invests at time 0. (Think, for instance, of a technological edge that allows this firm to enter first.) The strategies constructed by Eaton and Lipsey are otherwise symmetric. They also are Markovian, in that they depend only on the current payoff-relevant state (here, the two firms' capital structures, i.e., the number and the age of their productive plants). The incumbent firm (the one with capital) always purchases a second unit of capital Δ ($<H/2$) years before its current unit depreciates. The other firm invests in a unit of capital if the incumbent has only one unit and this unit is more than $H-\Delta$ years old. In equilibrium, the length Δ is chosen such that when the incumbent's unit of capital is $H-\Delta$ old, the potential entrant is indifferent between entering and not entering. If he does not enter, the incumbent remains a monopoly forever, and the entrant makes no profit. If he enters, he makes a profit of $-f$ for Δ years (because the incumbent is still committed: The fixed cost on his current unit is sunk) and enjoys monopoly profit forever after. The incumbent's investment path is represented in figure 8.15. Along the equilibrium path, the incumbent always renews his capital before it depreciates. The potential entrant never enters; he is kept out of the market by the incumbent's commitment to stay in for at least Δ years after entry (which inflicts short-term losses on the entrant).

Let us now compute Δ. In equilibrium, the incumbent's present discounted profit from date 0 on (or from any date at which he buys one new unit of capital) is

$$V = \int_0^{\infty} \tilde{\Pi}^m e^{-rt} dt - \left(\int_0^H f e^{-rt} dt\right)(1 + e^{-r(H-\Delta)} + e^{-r2(H-\Delta)} + \cdots).$$

The first term represents the flow monopoly profit forever. The second is the cost of one unit of capital, repeated at dates 0, $H-\Delta$, $2(H-\Delta)$, ..., $n(H-\Delta)$, ... Some simple mathematics yields

$$V = \frac{\tilde{\Pi}^m}{r} - \frac{f}{r}\left(\frac{1-e^{-rH}}{1-e^{-r(H-\Delta)}}\right). \tag{8.1}$$

Now suppose the potential entrant wants to enter. Obviously, there is no point in entering strictly before the

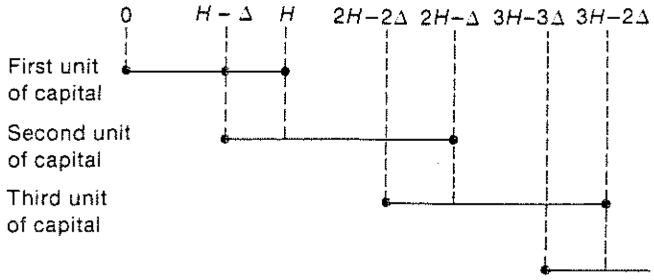


Figure 8.15
Incumbent's equilibrium investment strategy.

incumbent buys a second unit of capital (because the net duopoly profit flow is negative), so the entrant will wait and preempt the incumbent just before the latter buys its second unit (i.e., when the incumbent's current unit is $H-\Delta$ old). If the entrant does so, the incumbent does not buy a second unit, but he sticks around with his old unit for Δ units of time before exiting the market. The entrant's profit from the entry date on is thus equal to V minus the monopoly profit forgone during the entrant's first Δ years of existence (that is, the only difference between the incumbent at date 0 and an entrant who takes over is that the entrant is in a duopoly situation for Δ units of time):

$$V - \int_0^{\Delta} \tilde{\Pi}^m e^{-rt} dt = V - \tilde{\Pi}^m \frac{1-e^{-r\Delta}}{r}.$$

Because the second unit of capital is costly and is useless for productive purposes, the incumbent chooses Δ as small as is consistent with deterring entry:

$$V - \tilde{\Pi}^m \frac{1-e^{-r\Delta}}{r} = 0, \tag{8.2}$$

or, substituting V,

$$\frac{\tilde{\Pi}^m}{f} = \frac{1-e^{-rH}}{e^{-r\Delta}-e^{-rH}}. \tag{8.3}$$

Under our assumptions, equation 8.3 implies that $\Delta < H/2$.

We are particularly interested in what happens in cases of very short commitments. Let H (and thus Δ) tend to zero. Performing first-order Taylor approximations on equation 8.3, we get

$$\frac{\tilde{\Pi}^m}{f} \simeq \frac{rH}{rH - r\Delta} = \frac{H}{H-\Delta}.$$

Thus,

$$\frac{\Delta}{H} \simeq 1 - \frac{f}{\tilde{\Pi}^{\mathrm{m}}}. \tag{8.4}$$

This means that the incumbent has two units of capital $\Delta/(H - \Delta) \simeq (\tilde{\Pi}^{\mathrm{m}} - f)/f$ percent of the time. More interesting, perhaps, equation 8.2 yields

$$V \simeq 0. \tag{8.5}$$

Thus, even though there is only one firm in the industry in equilibrium, this firm does not make profits. The monopoly rent is entirely dissipated by the accumulation of the second unit of capital. This is natural. If the value V of being a monopolist were large, the entrant would come in, lose money for a very short period of time (because H is small and therefore Δ is small), and capture V. Thus, for short-run commitments, *potential* competition drives the monopolist's profit down to zero.

Even for short-run commitments, we do not obtain the contestability outcome. Actually, we obtain Posner's wasteful-rent-dissipation postulate exactly: The monopolist charges the monopoly price and yet makes no profit. The total welfare loss per unit of time is equal to the loss in consumer surplus (see the triangle in figure 1.2) plus the net monopoly profit ($\tilde{\Pi}^{\mathrm{m}} - f$) (see chapter 1). This should not be surprising; the only possible avenue for rent dissipation in this model is excess capital, which by definition has no social value. This brings us to our second model, in which rent dissipation is socially useful.

Contestability

Our second model is due to Maskin and Tirole (1988a). Though it is similar in spirit to the first model, the formulation differs in some respects. Time is discrete, and the horizon is infinite. There are two firms, which compete in capacities. A capacity is locked in for two periods once chosen. Let $\tilde{\Pi}(K_i, K_j)$ denote the per-period profit of a firm with capacity K_i when its rival has capacity K_j (gross of the per-period fixed cost f). As usual, Π decreases with K_j, and the cross-partial derivative $\partial^2 \tilde{\Pi}/\partial K_1 \partial K_2$ is negative (capacities are strategic substitutes). The period length is T, and the discount factor between periods is $\delta = e^{-rT}$.

The firms choose their capacities sequentially. (The model actually is equivalent to a continuous-time model in which firms choose capacities K, which, as in Eaton and Lipsey 1980, depreciate in a "one-horse shay" manner, but according to a Poisson process—i.e., H is stochastic.[74]) Firm 1 picks capacities in odd periods and firm 2 in even periods.[75] A firm picks a capacity for two periods of production and sinks in the first of these two periods a fixed cost: $f(1 + \delta)$ for capacity $K > 0$, zero for $K = 0$. Let

$$\tilde{\Pi}^{\mathrm{m}} \equiv \max_K [P(K)K - (c + c_0)K],$$

where c is the marginal cost of production and c_0 is the marginal cost of installing capacity. As before, assume that $f < \tilde{\Pi}^{\mathrm{m}} < 2f$. Thus, one firm is viable, but not two. Strategies are required to be "Markov" (i.e., payoff relevant)—that is, firm i reacts to the capacity K_j chosen by

---

74. Consider a continuous-time model with rate of interest r. Let $\Pi^i(K_1, K_2)$ denote firm i's gross profit flow per unit of time. When a firm chooses a capacity, its period of commitment to that capital is stochastic. The probability that the commitments will lapse between date t and date $t + \Delta t$ is independent of time and is equal to $\lambda \Delta t$. One can think of this technology as an uncertain working lifetime (the time independence of depreciation is clearly an extreme assumption). Letting $V^i(K_j)$ (respectively, $W^i(K_i)$) denote the present discounted value of firm i's profit when firm i renews its capital and reacts to firm j's current capital K_j (respectively, when firm j renews its capital and reacts to firm i's current level K_i). From dynamic programming, we have

$$V^1(K_2) = \max_{K_1} \{[\Pi^1(K_1, K_2) - f]\Delta t + \lambda \Delta t\, W^1(K_1) e^{-r\Delta t} + (1 - \lambda \Delta t) V^1(K_2) e^{-r\Delta t}\},$$

which yields

$$V^1(K_2) = \max_{K_1} \left(\frac{\Pi^1(K_1, K_2) - f}{\lambda + r} + \frac{\lambda}{\lambda + r} W^1(K_1) \right).$$

Thus, the continuous-time model is equivalent to the discrete-time, sequential-move model with gross profit function

$$\tilde{\Pi}^1(K_1, K_2) \equiv \frac{\Pi^1(K_1, K_2)}{\lambda + r}$$

and discount factor

$$\delta = \frac{\lambda}{\lambda + r}.$$

For the dynamic programming equations in the discrete time framework, see the next subsection.

75. When we "endogenize" timing by letting firms choose their capacities whenever they want, subject to the constraint that a capacity is locked in for two periods once chosen, the symmetric equilibrium is the same as described below. (Another way of endogenizing the timing is given in note 74.)

its rival in the last period and still in place for the current period by choosing a capacity $K_i = R_i(K_j)$.[76]

As in Eaton and Lipsey 1980, there exists a unique symmetric equilibrium. For δ sufficiently large, it takes the following form (illustrated in figure 8.16): In equilibrium only one firm operates, at capacity level K^*. A firm chooses to enter if and only if its rival's capacity is less than the entry-deterring capacity K^*; if it enters, it accumulates capacity K^* itself. In equilibrium, K^* is such that the entrant is indifferent between entering and not entering:

$$[\tilde{\Pi}(K^*, K^*) - f] + \frac{\delta}{1-\delta}[\tilde{\Pi}(K^*, 0) - f] = 0. \quad (8.6)$$

This equation reflects the fact that the entrant gets $\tilde{\Pi}(K^*, K^*) - f < 0$ when entering (recall that $2\tilde{\Pi}(K, K) \leqslant \tilde{\Pi}^m < 2f$ for all K). The incumbent exits in the following period, and the entrant becomes a monopolist who keeps deterring entry by choosing K^* forever. Then the entrant's future profits are given by

$$\delta[\tilde{\Pi}(K^*, 0) - f] + \delta^2[\tilde{\Pi}(K^*, 0) - f] + \cdots$$
$$= \frac{\delta[\tilde{\Pi}(K^*, 0) - f]}{1-\delta}.$$

The incumbent chooses its own capacity so as to just deter entry; accumulating beyond K^* is costly because, as we will see shortly, K^* already exceeds the monopoly capacity K^m.

In summary: The equilibrium involves a single firm, operating at a capacity level K^*.[77] This firm engages in some form of limit pricing. It accumulates more capacity than a monopolist facing no threat of entry would. It therefore charges a price that is less than the monopoly price (see below).

We will now investigate the case of *short-run commitments* (where T tends to 0—i.e., δ tends to 1). From equation 8.6 we see that when δ converges to 1, $\tilde{\Pi}(K^*, 0) - f$ converges to 0. That is, the monopolist's profit converges to 0. (Note in particular that K^* exceeds K^m.) The intuition for this rent-dissipation result is the same as

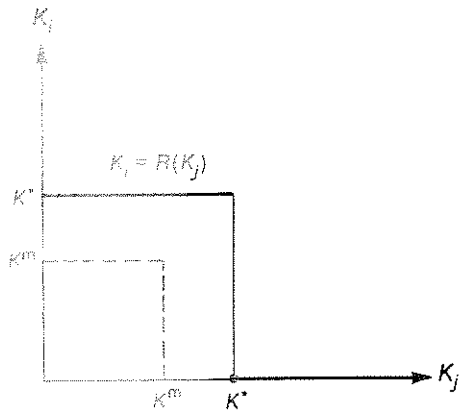


Figure 8.16

before. The established firm enjoys an important incumbency advantage only if it can inflict a duopoly loss on the entrant for a sufficient amount of time. The entrant is lured by the prospect of becoming a monopolist after only a brief fight, so the incumbent must raise its capacity to deter entry.

An important difference from the previous model is that the rent dissipation need not be wasteful. Indeed, if the established firm's capacity K^* is used (so output q is equal to K^*), the dissipation is socially useful. Rent dissipation occurs through price reduction rather than through excess capacity. The outcome in the limit is the one predicted by the contestability school (see section 8.1). Whether the monopolist uses all of his capacity K^* is an empirical question. As in chapter 5, K^* installed is used if the marginal cost of investment c_0 is sufficiently large relative to the marginal cost of production c.

For smaller discount factors, entry is blockaded. That is, by accumulating the monopoly capacity, the established firm deters entry.

No Fixed Cost: The Dynamics of Cournot Competition

In the two preceding subsections we assumed the existence of large fixed costs, which made the industry a natural monopoly. The incumbent firm overinvested to deter entry. In the absence of fixed costs (or in the presence of low fixed costs), there is room for two firms.

---

76. In other words, the strategies do not depend on the payoff-irrelevant history of the game.

77. In both the Eaton-Lipsey model and the Maskin-Tirole model, there are two asymmetric Markov-perfect equilibria for sufficiently short commitments. In these equilibria, one of the firms enjoys unconstrained monopoly power in a steady state (that is, it does not renew its capital early in the first model, and it accumulates K^m in the second). This firm never quits the market, and it reacts to entry by assuming that the entrant will exit once its commitment has elapsed. This aggressive behavior is self-fulfilling and ends up deterring entry.

Rather than deter entry, the firms accommodate each other. This subsection analyzes accommodation in an industry with short-run commitments and presents the argument that some of the principles governing accommodation under quantity competition in two-period models (see sections 8.3 and 8.4) carry over to full-fledged dynamic games.

Consider the sequential-move capacity competition model of the preceding subsection, but assume that the firms incur no fixed cost ($f = 0$). (The analysis here follows one presented by Maskin and Tirole [1987] and is based on the earlier model of Cyert and DeGroot [1970].) Firm 1 chooses capacities in odd periods (which are locked in for two periods and can be freely changed after two periods), and firm 2 chooses capacities in even periods. Firm i's intertemporal profit at time t is

$$\sum_{s=0}^{\infty} \delta^s \Pi^i(K_{1,t+s}, K_{2,t+s}).$$

As before, we make the usual assumptions on the profit function: $\Pi^i_{ii} < 0$, $\Pi^i_j < 0$, $\Pi^i_{ij} < 0$. We look for a pair of dynamic reaction functions, $R_1(\cdot)$ and $R_2(\cdot)$, that form a Markov perfect equilibrium. Thus, if firm 2's current (locked-in) capacity is K_2, firm 1 reacts by choosing capacity $K_1 = R_1(K_2)$ to maximize its present discounted profit given that both firms will then move according to R_1 and R_2. As in section 6.7, let $V^i(K_j)$ denote the present discounted profit of firm i when it reacts to its rival's capacity K_j, and let $W^i(K_i)$ be firm i's present discounted profit when it is locked into K_i and its rival reacts. The equilibrium conditions are the following:

$$V^1(K_2) = \max_K \{\Pi^1(K, K_2) + \delta W^1(K)\}, \tag{8.7}$$

$$R_1(K_2) \text{ maximizes } [\Pi^1(K, K_2) + \delta W^1(K)], \tag{8.8}$$

$$W^1(K_1) = \Pi^1(K_1, R_2(K_1)) + \delta V^1(R_2(K_1)), \tag{8.9}$$

and similarly for firm 2.

The first common feature with traditional analysis is that, because capacities are strategic substitutes ($\Pi^i_{ij} < 0$), the reaction curves are downward sloping. To show this, it suffices to write the optimality of reaction functions (an identical technique is used to prove the monotonicity of incentive-compatible allocations in incentive problems). Consider two capacity levels, K_2 and $\tilde{K}_2$. Let $R_1(K_2)$ and $R_1(\tilde{K}_2)$ denote the optimal reactions to K_2 and $\tilde{K}_2$. By definition, $R_1(K_2)$ is a better response to K_2 than $R_1(\tilde{K}_2)$:

$$\Pi^1(R_1(K_2), K_2) + \delta W^1(R_1(K_2)) \geqslant \Pi^1(R_1(\tilde{K}_2), K_2) + \delta W^1(R_1(\tilde{K}_2)). \tag{8.10}$$

Similarly, $R_1(\tilde{K}_2)$ is a best response to $\tilde{K}_2$:

$$\Pi^1(R_1(\tilde{K}_2), \tilde{K}_2) + \delta W^1(R_1(\tilde{K}_2)) \geqslant \Pi^1(R_1(K_2), \tilde{K}_2) + \delta W^1(R_1(K_2)). \tag{8.11}$$

Adding equations 8.10 and 8.11, we obtain

$$\Pi^1(R_1(K_2), K_2) - \Pi^1(R_1(\tilde{K}_2), K_2) + \Pi^1(R_1(\tilde{K}_2), \tilde{K}_2) - \Pi^1(R_1(K_2), \tilde{K}_2) \geqslant 0, \tag{8.12}$$

which is equivalent to

$$\int_{\tilde{K}_2}^{K_2} \int_{R_1(\tilde{K}_2)}^{R_1(K_2)} \Pi^1_{12}(x, y)\,dx\,dy \geqslant 0. \tag{8.13}$$

But, by assumption, $\Pi^1_{12} < 0$. Thus, equation 8.13 implies that $R_1(K_2) \leqslant R_1(\tilde{K}_2)$ if $K_2 > \tilde{K}_2$. The reaction curves are necessarily downward sloping.

To find equilibrium reaction functions, we must solve the system of equations 8.7–8.9.[78] For quadratic profit functions, such as

$$\Pi^i = K_i(d - K_i - K_j),$$

there exists a particularly simple solution. Each firm's reaction function is linear in its rival's capacity: $R_1 = R_2 = R$, where $R(K) = a - bK$. This solution also has the remarkable property that it is the limit of each firm's reaction function at any date when the horizon is finite but tends to infinity.[79]

78. To find a differentiable solution (if such a solution exists), we can differentiate equation 8.9 and take the first-order condition in 8.7. After some substitutions, we obtain a system of difference-differential equations in the two reaction functions. This system is generally hard to solve, but is easy in the case of quadratic profit functions.

79. The finite-horizon solution is too complex to be derived in closed form. Indeed, Cyert and DeGroot (1970) computed it numerically. To show convergence toward the infinite-horizon, linear, Markov-perfect equilibrium, one shows that the finite-horizon solution belongs to the class of linear reaction functions with slopes between $-\frac{1}{2}$ and 0 and intercepts between 0 and d, that it is obtained by backward induction through a contraction mapping in the space of such functions, and that the fixed point of the contraction mapping (which is the limit of the reaction function for large horizons) satisfies the difference-differential equations for (R_1, R_2) derived from equations 8.7 through 8.9.

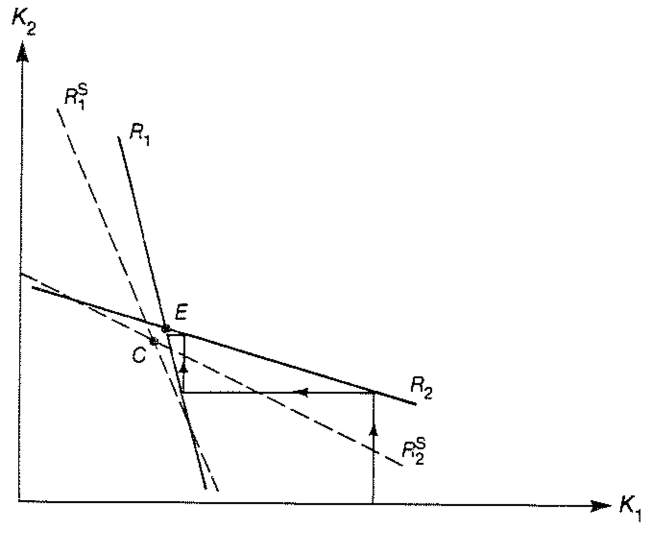


Figure 8.17

The dynamics of the game are illustrated in figure 8.17. The solid lines depict the dynamic reaction functions for δ in (0, 1), the broken lines represent the static Cournot reaction functions R_1^s and R_2^s, E denotes the steady-state allocation, and C denotes the Cournot outcome.

For $\delta = 0$, the firms are myopic. They react according to the static reaction function

$$R(K) = \frac{d}{2} - \frac{K}{2},$$

which maximizes $\tilde{K}(d - K - \tilde{K})$. Thus, $a = d/2$ and $b = \frac{1}{2}$. The industry dynamics are then called the *tâtonnement process*. The steady state is the Cournot allocation C. For $\delta > 0$, each firm takes into account not only its current profit but also its rival's future reaction. Because reaction curves are downward sloping, the intuition is that a firm should invest beyond its short-run interest so as to induce its rival to curtail capacity (as in the Stackelberg game of section 8.2). Indeed, it can be shown that when δ increases, the steady-state symmetric level of capacity, given by $K = a - bK$ or $K = a/(1 + b)$, increases and thus moves away from the Cournot level. The process is dynamically stable—for any initial level of capacity, the capacities of the two firms converge to the steady-state capacities. This generalizes the Cournot tâtonnement process in that each firm rationally anticipates the influence of its capacity choices on its rival's behavior.

The moral of such a simple infinite-horizon model is that the intuitions derived in two-period models carry over: Strategic substitutes yield downward-sloping reaction curves, so each firm overinvests for strategic reasons. The outcome can be thought of as symmetric Stackelberg leadership.

8.6.1.2 Long-Term Capital-Accumulation Games

In the other polar case, investment creates a long-term commitment to be in the market. Specifically, we assume that investment, once in place, does not depreciate and cannot be resold. That is, investment is irreversible. The following model is due to Spence (1979); the version presented here is from Fudenberg and Tirole 1983b.

Consider a duopoly, with firms indexed by $i = 1, 2$. Time is continuous, and the horizon is infinite. Firm i's flow profit at any time t, gross of investment expenditures, is given by

$$\Pi^i(K_1(t), K_2(t)),$$

where $K_i(t)$ is firm i's capital stock at date t (as usual, $\Pi^i_{ii} < 0$, $\Pi^i_j < 0$, and $\Pi^i_{ij} < 0$).

Capital at date t is equal to cumulative investment to date:

$$\dot{K}_i(t) \equiv \frac{dK_i(t)}{dt} = I_i(t),$$

where $I_i(t)$ is the rate of investment. It is assumed that the cost of investment is linear. One unit of investment costs \$1. To avoid instantaneous investment at date 0, we bound each firm's investment above by $\overline{I}_i$. This technology is an example of convex investment cost. Investment must be non-negative, and there is no depreciation. Thus, the capital stocks are nondecreasing. Firm i's net profit at time t is

$$\Pi^i(K_1(t), K_2(t)) - I_i(t).$$

Firm i's strategy is a path of investment $\{I_i(t)\}$ satisfying $0 \leqslant I_i(t) \leqslant \overline{I}_i$. Each firm's investment at date t depends on the current capital stocks $(K_1(t), K_2(t))$ (again, we assume that the strategies are of the Markov type, in that they depend only on the payoff-relevant state of the game and not on the whole history). Both firms enter the market at time $t = 0$ without any capital.

Firm i's objective function is equal to its present discounted profit:

$$\int_0^{\infty} [\Pi^i(K_1(t), K_2(t)) - I_i(t)]e^{-rt}\,dt.$$

In this subsection we will only consider the limit game in which both firms become infinitely patient (that is, r tends to 0). In this case, the firms maximize their time-average payoffs, so that only the eventual steady-state capital levels matter (no firm will choose to invest forever). Thus, firm i's objective function is $\Pi^i(K_1^{ss}, K_2^{ss})$, where ss stands for steady state. This simplification allows us to ignore the private cost of investment and focus on its strategic aspect, and to use a simple diagrammatic approach.[80]

Let us first examine the "precommitment" or "open-loop" equilibria.[81] In a precommitment equilibrium, the firms simultaneously commit themselves to entire time paths of investment. Thus, the precommitment equilibria are really static, in that there is only one decision point for each firm. The precommitment equilibria are just like Cournot-Nash equilibria, but with a larger strategy space. In the capacity game, the precommitment equilibrium is exactly the same as if both firms built their entire capital stocks at the start (because there is no discounting). In the resulting "Cournot" equilibrium, each firm invests to the point at which the marginal productivity of capital equals zero, given the steady-state capital level of its opponent. All of the many different paths that lead to this steady state are precommitment equilibria. For example, each firm's strategy could be to invest as quickly as possible to its Cournot level. We can highlight the similarity of this solution to a Cournot equilibrium by defining the "steady-state reaction curves" that give each firm's desired steady-state capital level as a function of its rival's steady-state capital level. Under our assumptions, these reaction curves look the same as usual "nice" Cournot reaction curves. The reaction curves R_1 and R_2 are displayed in figure 8.18, where IGP is the investment-growth path (the path along which both firms are investing as rapidly as they can). The precommitment (open-loop) equilibrium is at $C = (C_1, C_2)$, the intersection of the two curves. We have seen that the use of the concept of precommitment transforms an apparently dynamic game into a static one. As a modeling strategy, this transformation is ill advised: "... one should not allow precommitment to enter by the back door.... If it is possible, it should be explicitly modeled ... as a formal choice in the game." (Kreps and Spence 1984)

Now allow firm i's investment at time t to depend on the capital stocks at that time (the firms employ closed-loop strategies). The capital stocks are the "state variables" (i.e., the capital stocks at any date and the investment programs from that date on are all the information one needs to compute the payoffs). A Markov perfect equilibrium is a pair of Markov strategies

$$\{I_i(K_1(t), K_2(t))\}_{i=1,2}$$

that form a "closed-loop" Nash equilibrium from any possible initial state (K_1^0, K_2^0) and not only from the initial state (0, 0).

Consider figure 8.19, which depicts a Markov perfect equilibrium. The arrows indicate the direction of motion. The motion is vertical if only firm 2 is investing, horizon-

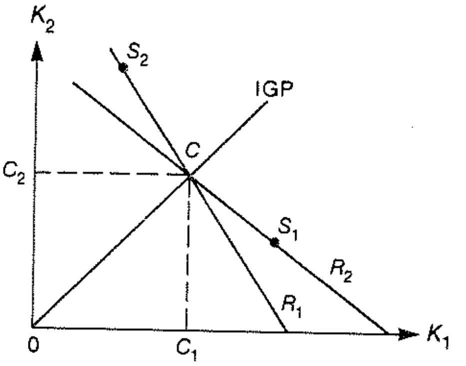


Figure 8.18

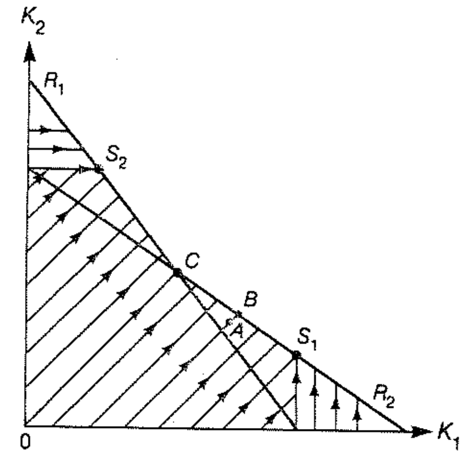


Figure 8.19

80. For analyses of the discounting case, see Fudenberg and Tirole 1983b and Nguyen 1986.

81. The analysis here follows Fudenberg and Tirole 1986, pp. 8–13.

tal if only firm 1 is investing, diagonal if each is investing as quickly as possible, and nonexistent if neither firm invests (because of the linearities, the optimal strategies are, in the jargon of optimal-control theory, "bang-bang"). Note that we have defined choices at every state, and not just those along the equilibrium path—we must do this in order to test for perfectness: Looking at figure 8.19, we see that unless firm 1 has a head start it cannot enforce its Stackelberg outcome S_1, because it cannot accumulate enough capital before firm 2 reaches its reaction curve. If firm 1 can invest to its Stackelberg level before firm 2 reaches its reaction curve, it does so and then stops; firm 2 then continues investing up to R_2. If for some reason firm 1's capital stock already exceeds its Stackelberg level, it stops investing immediately. The situation is symmetric on the other half of the diagram, which corresponds to states in which firm 2 has a head start. Thus, this equilibrium demonstrates how an advantage in investment speed or initial conditions can be exploited. The conditions of the growth phase (which firm got there first, the costs of adjustment, and so on) have a permanent impact on the structure of the industry.

It turns out that the equilibrium depicted in figure 8.19 is not unique. There are many others. To understand why, consider point A in the figure. This point is close to firm 2's reaction curve and past firm 1's reaction curve. The strategies specify that from point A on both firms invest until R_2 is reached. However, both firms would prefer the status quo at A. Firm 1 in particular would not want to invest even if firm 2 stopped investing; it just invests in self-defense to reduce firm 2's eventual capital level. Both firms' stopping at A is an equilibrium in the subgame starting at A, enforced by the credible threat of going to B (or close to B) if anyone continues investing past A. Thus, the Markov restriction does not greatly restrict the set of equilibria in the investment game.[82]

In this study we presume that capital does not depreciate. An open issue, analyzed by Hanig (1985) and Reynolds (1987), is the behavior of investment in the industry when capital depreciates. Intuition suggests th capital ought to lose some of its commitment value a that the steady-state levels of capital should be less sen tive to the initial head start of one of the firms. Hanig a Reynolds consider quadratic payoff functions,

$$\Pi^i = K_i(1 - K_i - K_j),$$

and quadratic investment costs,

$$C^i(I_i) = cI_i^2/2.$$

They allow depreciation ($\dot{K}_i = I_i - \mu K_i$) and discountin and they look for Markov-perfect-equilibrium investme strategies that are linear in the capital levels ($I_i($ $= -\alpha K_i(t) - \beta K_j(t) + \gamma$, where $\alpha, \beta, \gamma > 0$). They us differential-games techniques[83] to obtain such a solutio The main result is that the steady-state level of capital fo both firms strictly exceeds the Cournot level; thus, bot firms are beyond their reaction curves in the long run The intuition is the same as for the model of short-ru commitments (and no fixed cost) described above. Each firm at each instant keeps more capacity than it would if it could not influence its rival's accumulation. It thus forces its rival to reduce its capacity. Because both firms behave in this Stackelberg fashion, their capital levels exceed the Cournot ones. The commitment value of capital is inversely related to its rate of depreciation. In particular, capital that depreciates rapidly involves only a short-term commitment.

If we ignore fixed costs and barriers to entry, these models point at the following conclusion for dynamic competition under Markov strategies: Relative to static competition (see chapter 5), repeated interaction promotes collusion under price competition (see chapter 6) and fosters competition under capacity competition in the Hanig (1985), Maskin-Tirole (1987), and Reynolds (1987) models.[84] This conclusion makes economic sense. By raising its price, a firm creates incentives for its rival to do the same; by increasing its capital level, it induces its rival to reduce its own. Thus, the distinction between strategic

82. Fudenberg and Tirole (1983b) single out a reasonable "early-stopping" equilibrium (i.e., an equilibrium with steady state under the upper envelope of the reaction curves) through arguably intuitive arguments, including the elimination of Pareto-dominated equilibria. In the symmetric case, this equilibrium coincides with the joint profit-maximizing outcome. MacLeod (1985) offers a more formal argument that lends some support to this selection.

83. See Starr and Ho 1969 and Fudenberg and Tirole 1986. The differentiability of the investment strategies in the capital levels required by the theory of differential games is not an innocuous assumption. Its rules out the above early-stopping equilibria, in which a firm invests to some level, stops, and threatens to resume investment if its rival does so.

84. We must be careful here because of the potential multiplicity of equilibria. The previous no-depreciation capital-accumulation game admitted nondifferentiable equilibria, which are quite collusive. See note 82.

substitutes and complements has some relevance for the study of long-run competition.

The Stackelberg-Spence-Dixit model (see section 8.2) illustrates the fact that with low fixed costs and in the absence of substantial indivisibilities in production, established firms do not deter entry but only try to limit the expansion of entrants. The dynamic rivalry models discussed above make this point even more forcefully. The Stackelberg-Spence-Dixit model also shows that under large fixed costs and/or indivisibilities, entry deterrence becomes optimal for the incumbents. This point too was confirmed by the dynamic rivalry models discussed above.

8.6.2 Product Proliferation, Preemption, and the Persistence of Monopoly

In many industries, firms do not choose a continuous scale variable (like capacity in the previous investment game). Rather, because of indivisibilities or fixed costs, they face a discrete choice: They invest in plants that are most efficient in terms of scale (as in the case of a U-shaped cost curve); they choose among a limited set of products; they locate at a restricted set of geographical places; and so on. The advantage of being the first mover then takes an extreme form: that of preemption. Certainly, preemption occurs in the above long-term capital-accumulation game. Each firm would like to enter first so as to reach its Stackelberg capacity before its rival has accumulated enough capital to dissuade it from doing so. The effect of indivisibilities is that firms wish to preempt with a vengeance. In the investment game, a firm that delays its investment a bit loses a bit of its first-mover advantage (in the absence of depreciation, the steady state will involve a bit less capital for this firm and a bit more for its rival). In contrast, a firm that fails to install a plant or to occupy the right market niche on time may not be able to prevent an entrant from installing a plant or occupying the niche. A little delay may allow entry and thus have large consequences for the firms' profits.[85]

In this subsection we will study discrete choice and preemption. The discussion will be limited to situations in which the preempting firm does not physically deter entry but, rather, makes it unprofitable. (Exclusionary investments will be studied in chapter 10 in the context of patentable innovations.)

A natural focus of preemption games is the timing of the introduction of plants or products. As in the model of Eaton and Lipsey (1980), the established firms will tend to invest early. Another focus is the persistence of monopoly. Will the established firm always be able to deter its rivals from entering by investing early? Should we expect a monopolistic or an oligopolistic structure in the long-run?

8.6.2.1 Product Proliferation

As we saw in chapter 7, firms want to differentiate their products so as to avoid intense price competition (with some exceptions). Therefore, potential entrants look for unfilled market niches. To deter entry, the established firms may try to pack the product space and leave no profitable market niche unfilled. For instance, Scherer (1980, pp. 258–259) describes General Motors' 1921 decision to offer a complete spectrum of automobiles, and GM chairman Alfred P. Sloan's strategic approach to the decision. He also notes how the Swedish Tobacco Company, upon losing its legal monopoly position in 1961, reacted by offering twice as many brands (and by increasing its advertising twelvefold in the following years). Schmalensee (1978) observes that the six leading manufacturers of ready-to-eat breakfast cereals introduced eighty brands between 1950 and 1972 (the year in which the Federal Trade Commission issued a complaint against the four largest manufacturers, who had cornered 85 percent of the market and who enjoyed large profits).

Schmalensee (1978) shows formally how a cartel (a group of firms that act as a single monopolist) crowds a product space. In the context of a circular location model, he asks how many products a cartel should introduce to make further entry unprofitable; and he shows that it is indeed in a cartel's interest to deter entry in this way. Schmalensee's model is static and therefore is silent on the optimal timing of preemption. Subsequent research has developed models in which, over time, demand grows or the cost of introducing new products decreases, and the date of introduction of a new product is a choice variable.

85. A similar phenomenon would occur in an investment game with a fixed cost of entry.

Further persistence-of-monopoly results were discovered by Eaton and Lipsey (1979), who described preemption in a location model[86], Gilbert and Newberry (1982), who demonstrated a similar result and clearly identified the reason for persistence in the context of a patent race; and Gilbert and Harris (1984), who determined the "threat dates" at which an established firm can build indivisible plants so as to deter entry.[87] We now look at the results for a simple product-differentiation model.[88]

Let us return to the simple model developed in chapter 7 and consider a linear city of length 1. We will assume that there are only two possible store locations: one at each of the extremities of the city. This assumption, although not crucial, simplifies the presentation. The consumers, who are distributed evenly along the segment, have a transportation cost t per unit of distance. Time is continuous and belongs to the interval $[0, +\infty)$. At date 0, the consumer density is unitary. It remains unitary until date T, when it doubles instantly; then it remains 2 forever. (The discontinuous growth of the population, which is somewhat contrived, gives us a simple way of modeling locational choice in an expanding city.)

There are two firms. At date 0, firm 1 (the existing firm) serves the entire market from its single store at the left end of the city. At any future moment, each of the two firms can build a shop at the right end of the city with an accompanying fixed investment cost f. For the time being, we assume that a firm does not exit after sinking the investment cost.

We could assume that each firm can build a store at a location where its rival already has a store. However, because Bertrand competition with undifferentiated products yields zero profit, it is easy to see that such a policy is not profitable in our model; therefore, we will not consider it (but see below). The problem is to determine which firm will invest in building at the second location and at what moment this will occur.

Before date T, firm 1 earns a profit per unit of time of Π_0^m if neither of the two firms has built at the right end of the city, Π_1^m (without subtracting the fixed building cost) if it has built, and Π^d if firm 2 (the entrant) was the first to build. In the last case, firm 2 also earns Π^d per unit of time. If the unit production cost (net of the fixed building cost of the store) is constant, these flow profits are doubled after date T because of the population growth. We assume that $\Pi_1^m > \Pi_0^m$ and $\Pi_1^m > 2\Pi^d$. The first of these inequalities simply says that, if we ignore the cost of building the store, the existing firm prefers to have two stores rather than one; the second inequality says that for a given number of stores (here, two), the total profit of the industry is smaller for a duopoly because of competition. These conditions are quite general. That they hold in the case where the consumers have unitary demands, where $\bar{s} > 2t$ ($\bar{s}$ is the consumers' valuation for the good sold by both stores), and where the production cost c is zero is due to $\Pi_0^m = \bar{s} - t$, $\Pi_1^m = \bar{s} - t/2$, and $\Pi^d = t/2$.

Let $t_1 > 0$ denote the preemption date, that is, the date at which one of the firms invests (first) and builds at the right end of the city. Let $L_i(t_1)$ (respectively, $F_i(t_1)$) be the present discounted value of profit, at date 0, for firm i when it is first to invest and does so at date t_1 (respectively, is preempted). L and F stand for *leader* and *follower*. (The leadership is endogenous.) For $t_1 < T$, these functions are given by

$$L_1(t_1) = \int_0^{t_1} \Pi_0^m e^{-rt} dt + \int_{t_1}^{T} \Pi_1^m e^{-rt} + \int_T^{\infty} 2\Pi_1^m e^{-rt} dt - fe^{-rt_1},$$

$$F_2(t_1) = 0,$$

$$L_2(t_1) = \int_{t_1}^{T} \Pi^d e^{-rt} dt + \int_T^{\infty} 2\Pi^d e^{-rt} - fe^{-rt_1},$$

$$F_1(t_1) = \int_0^{t_1} \Pi_0^m e^{-rt} dt + \int_{t_1}^{T} \Pi^d e^{-rt} dt + \int_T^{\infty} 2\Pi^d e^{-rt} dt,$$

86. Precursors include Hay 1976, Prescott and Visscher 1977, and Rothschild 1976. For an account of these contributions, see Gabszewicz and Thisse 1986. Bonanno (1987) analyzes a Prescott-Visscher-type model of sequential entry in a spatial market. He allows firms to open several stores, whereas Prescott and Visscher restrict the analysis to one store or none per firm. At date i ($i = 1, \ldots, n$) firm i decides whether to enter and (if it chooses to enter) how many stores to open and where to locate them. At date $n + 1$, after the n firms have made their investment decisions, price competition takes place. The main result is that monopoly persists: Firm 1 deters entry. In fact, for some values of the parameters, entry deterrence is not achieved through product proliferation; rather, firm 1 opens the same number of stores as a protected monopolist, but rearranges the locations of these stores to deter entry. If this strategic location choice is not sufficient or not the most profitable way to deter entry, product proliferation occurs.

87. See Rao and Rutenberg 1979 for the optimal timing of plant installation when entry cannot be deterred.

88. The following analysis follows Fudenberg and Tirole 1986, pp. 41–45, which builds on Eaton and Lipsey 1979.

where r is the interest rate and f is the investment cost. For $t_1 > T$, we can define L_i and F_i in a similar way.

Now suppose that

$$\frac{2\Pi^d}{r} > f > \frac{\Pi^d}{r}.$$

The first inequality tells us that after the doubling of the population, the present discounted value of duopoly profit is greater than the investment cost. This condition guarantees that firm 2's entry into the industry is profitable. The second inequality says that at any moment before T, the duopoly profit (Π^d) does not cover the interest (rf) on the investment cost. These two inequalities imply that in the absence of a preemptive threat by firm 1, firm 2 wishes to invest exactly at date T (that is, L_2 reaches its maximum at date T). The functions L_i and F_i are depicted in figure 8.20.

We define $T_2 < T$ such that, at date T_2, firm 2 is indifferent between preempting and being preempted; i.e.,

$$L_2(T_2) = F_2(T_2) = 0.$$

We can verify that $L_2(t_1) > F_2(t_1)$ if and only if $t_1 > T_2$, and we can verify that $L_1(t_1) > F_1(t_1)$ for any $t_1 \geqslant T_2$ (using $\Pi_1^m - \Pi^d > \Pi^d$).

Suppose that $(\Pi_1^m - \Pi_0^m) < rf$. In other words, in the absence of a threat of entry, the established firm does not choose to invest before T.[89] That is, L_1 is increasing before T. Figure 8.20 fully summarizes the preemption game between the two firms.

We can now solve the preemption game. To do this, we consider the problem by looking back in time from date T. At that moment, the established firm (firm 1) wishes to invest (if no one has done so before) regardless of firm 2's subsequent strategy. Knowing this, firm 2 will not allow investment by firm 1; it will preempt at some earlier moment $T - \varepsilon$, because $L_2(T - \varepsilon) > F_2(T)$. Firm 1, knowing the entrant's preemptive choice at $T - \varepsilon$, will wish to preempt by investing just before that moment, and so forth. This preemptive spiral stops at moment T_2, when firm 2 finds further preemption too costly. Therefore, in order to preempt firm 2, it is sufficient for firm 1

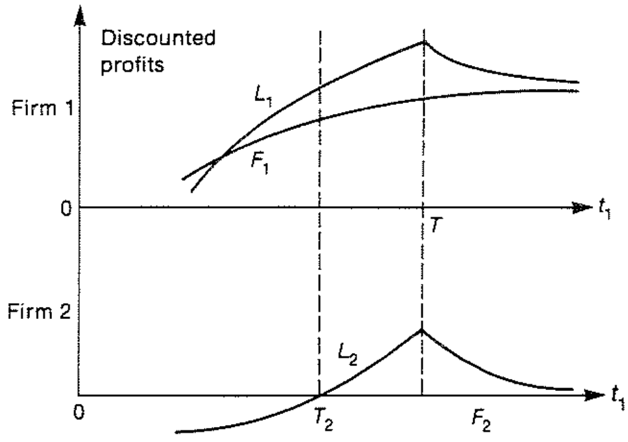


Figure 8.20

to invest just before T_2. Since L_1 is increasing before T_2, firm 1 waits until that date (or just slightly before it) to preempt. Therefore, equilibrium is characterized by the following two properties:

- The established firm preempts the entrant and retains its monopoly.
- Preemption occurs before the increase in the population, at the first date when the entrant would have been willing to enter in the absence of preemption.

A correct formalization of the equilibrium strategies can be found for a similar game in Fudenberg and Tirole 1985.[90]

The basic result of the above example is the persistence of monopoly. The intuitive reasoning behind the more general property goes as follows: Competition is destructive of profits; a monopolist with the same production technology as a duopoly industry earns more profit than the two rival firms together (at worst, the monopolist can always make its stores choose the strategies followed by the competing firms). This property, called the *efficiency effect* and reflected here by the inequality $\Pi_1^m \geqslant 2\Pi^d$, is very general and forms the basis for the phenomenon of monopoly persistence. At the time of entry, the potential entrant bases his decisions on duopoly profit per unit of time Π^d. Now consider the choices available

89. This inequality was satisfied above: $\Pi_1^m - \Pi_0^m = t/2 = \Pi^d < rf$.

90. The above reasoning is very loose. Familiar strategies for continuous-time games—called "distributional strategies," and specifying a (right continuous) cumulative probability distribution that a firm has moved by any date t—are not "rich" enough to describe such preemption games. Richer and more satisfactory strategies are obtained by taking the limit of the discrete-time model while allowing reasonable behavior. See Simon and Stichcombe 1986 and Simon 1987 for useful elaborations on this theme.

to the existing firm, i.e., either to allow or to preempt entry. Allowing entry implies a loss of $\Pi_1^m - \Pi^d$ per unit of time. Since $\Pi_1^m - \Pi^d > \Pi^d$, *the existing firm has more of an incentive to preempt than the entrant has to enter.*

The monopolist's rent is dissipated, although not fully, by the necessity to invest earlier than it desires (in order to preempt the entrant). In the above example with identical unit demands and linear transportation cost, this rent dissipation turns out to be socially wasteful, as in Eaton and Lipsey 1980. Thus, a social planner would wish to eliminate the threat of entry.[91]

It is worthwhile to compare preemption games such as the one just solved with war-of-attrition games such as the one considered in section 8.1. Both are "games of timing." In such games, each firm makes a single decision (when to enter in the preemption game; when to exit in the war of attrition). In the preemption game, each firm prefers to be first (at least over a period of time preceding the optimal date for moving), but would like to move "late" if it could be sure that its rival would not preempt. In the war of attrition, each firm prefers to "move" second (e.g., not to exit) but would like to move "early" if it could be sure that it would be outlasted by its rival. These two standard games are only polar examples of games of timing, and more general industrial-organization situations may involve different patterns; however, the techniques and intuitions derived for these games help us to apprehend the more complex situations (see Katz and Shapiro 1984).

8.6.2.2 Is Spatial Preemption Credible?

The general line of reasoning based on the efficiency effect suggests that monopoly situations remain monopoly situations, which of course is not always the case. We will consider what may be wrong with this reasoning. For the moment, let us observe that the incumbent's investment has a preemptive value only if the incumbent is somehow committed to this investment (see Judd 1985). A multiproduct incumbent who can withdraw some of his products at low cost may not be able to use crowding as a barrier to entry. This seems logical; we have been insisting all along that investment deters entry more easily when committed. Judd's interesting insight is that if a multiproduct firm competes with a single-product rival on some market, *the multiproduct firm has more incentive to quit the market than its rival as long as a low price in this market depresses the demand for its other goods.* Thus, existing products may have little commitment value.

To see how the multiproduct firm may be forced to exit a market, consider the previous model of a linear city. Suppose that the incumbent has preempted the entrant and has two stores, located at the two extremities. Suppose further that the entrant follows suit and enters the right-end location itself. If no firm exits, Bertrand competition drives the price of the two right-end stores down to marginal cost c. Hence, each firm makes a zero profit at its right-end store. Firm 1 makes a positive profit at its left-end store. Because of the transportation cost, the goods sold by the left- and right-end stores are differentiated, and firm 1 can keep its price a bit above c without losing all its customers (see chapter 7). However, its profit is meager, because the good sold at the right-end location is sold at the low price c. Now compare the two firms' incentives to quit the right-end location, assuming that the firms do not recoup their building cost f when exiting and do not incur any extra exit cost. Firm 2 has little incentive to exit, because it makes a zero profit whether it exits or stays when firm 1 stays. Firm 1, however, makes more money by exiting than by staying, if firm 2 stays. By exiting the right-end location, it raises the price at this location and therefore increases the residual demand for the good sold at the left-end location. For instance, for linear transportation costs, uniform density of consumers, and a city of length 1, the duopoly price is $c + t > c$ (see chapter 7). Because firm 1 was not making any money from the consumers purchasing at its right-end branch, it cares only about the residual demand faced by its left-end branch; thus, it increases its profit by exiting. To summarize loosely: Exiting the right-end location, a weakly dominated strategy for firm 2, raises firm 1's profit. Thus,

91. Because of the inelastic demand structure, the flow increase in welfare (before T) associated with the introduction of the right-end store is equal to the savings in average transportation costs: $t/2 - t/4 = t/4$ (assuming that the monopoly always covers the market). The flow cost of the store is equal to rf. But, by assumption, $rf > \Pi^d = t/2 > t/4$.

in equilibrium, *firm 1 exits immediately and firm 2 stays; the result is a duopoly.*[92]

Solving for the overall game, notice that firm 2 kicks firm 1 out of its right-end location immediately if both firms are located there, and recall that no firm wanted to enter the right-end location before date T except for preemptive purposes. Thus, the equilibrium is such that *firm 1 never enters the right-end location; firm two enters this location at date T.* No preemption occurs. We conclude that small exit costs together with product substitution may place the incumbent firm at a disadvantage and prevent it from credibly preempting the entrant through product proliferation.

*Exercise 8.10** Two differentiated goods, apples and oranges, are located at the two extremes of a linear product space (a segment of length 1). The utility of a consumer located at x is

$$\bar{s} - tx^2 - p_1$$

if he consumes one apple,

$$\bar{s} - t(1 - x)^2 - p_2$$

if he consumes one orange, and 0 otherwise (consuming both yields indigestion). The price of an apple is p_1; the price of an orange is p_2. Consumers are located uniformly along the segment. (This is exactly like the transportation model, where spatial preferences are reinterpreted as tastes, except that transportation costs are quadratic instead of linear.) The marginal cost of each good is c. Firm 1 is an apple monopoly, firm 2 an orange monopoly.

(i) Show that the demand functions are

$$D_1 = (p_2 - p_1 + t)/2t$$

and

$$D_2 = (p_1 - p_2 + t)/2t$$

in the relevant range ($|p_2 - p_1| \leqslant t$ and prices not too high).

(ii) Solve for the Bertrand equilibrium. Compute the profits.

(iii) Suppose that firm 1 is an apple monopoly, but that both firms produce oranges. Compute the Bertrand equilibrium. Show that Π^1 is smaller (by a factor of 4) than in question ii. Explain.

(iv) Suppose that there are no exit costs, that entry costs are sunk, that firm 1 is in both markets, and that firm 2 is in the orange market (as in question iii). Which firm has an incentive to exit the orange market? What do you conclude about the role of sunk costs or exit costs with regard to the possibility of entry deterrence through product proliferation (e.g., firm 1 entering first in orange markets)?

8.6.2.3 Do Monopolies Persist?

Subsection 8.6.2.1 unveiled an important factor favoring the persistence of monopolies: the efficiency effect. *Because competition destroys industry profits, an incumbent has more incentive to deter entry than an entrant has to enter.*[93] In general, however, this efficiency effect is not sufficient for the persistence-of-monopoly result. (This is fortunate for the theory: In the United States there are

92. This is a very informal description of the game. Exiting is a weakly dominated strategy for firm 2, because by always charging c it can guarantee itself a zero intertemporal profit. Furthermore, staying is profitable if firm 1 exits. (This weak-domination argument actually assumes that reentry is impossible, but a more sophisticated argument can be used to derive the same outcome when reentry is allowed.) Suppose now that we rule out weakly dominated strategies as eligible for equilibrium behavior (as is done, for instance, in Selten's notion of trembling-hand perfect equilibrium in discrete games—see the Game Theory User's Manual). Firm 2 stays, and firm 1 has no other choice than to exit.

If firms could recoup part of their fixed cost f when exiting, exiting would no longer be a weakly dominated strategy for firm 2, but firm 1 would still gain more by exiting than firm 2. The exit game would then resemble a war of attrition. (In the mixed-strategy equilibrium of this war of attrition, firm 2's probability of exit exceeds firm 1's.)

93. This efficiency effect rests on the comparison between a monopoly and a duopoly. One might conjecture that, more generally, a big firm has more incentive to preempt than a small firm. This, however, is not correct. Suppose, for instance, that the initial market structure is a duopoly. Firm 1 (the big firm) has unit cost 1 and firm 2 (the small firm) has unit cost 3. There is no fixed cost, and the firms wage Cournot competition. Suppose that an innovation comes along that makes a technology with unit cost 2 accessible at low adoption cost. Even in the case where a firm can buy an exclusive right to this technology (thus excluding its rival), it is not clear that firm 1 will preempt its rival and purchase the new technology (which it would do only for competitive purposes, and not for productive purposes). It may be the case that the cost reduction for firm 2 offsets the loss in industry revenues stemming from more intense competition. Thus, firm 2 may have more incentive to buy the technology than firm 1 (see Leung 1984 and Kamien and Tauman 1983 for related ideas). This comes from the fact that we are comparing an initial duopoly situation with a subsequent duopoly situation. If the initial industry configuration were an unconstrained monopoly (firm 2 starting with a large unit cost), then the efficiency effect would prevail.

very few pure monopolies. In the absence of regulatory restrictions, multifirm markets are the norm.)

First, preemption must be *effective*. Either it allows the preemptor to establish a property right on the technology (e.g. through a patent or exclusive licensing) or it commits the firm to intense price competition if the rival follows suit. An example in which preemption is not fully effective (in the sense that it allows entry) is the Stackelberg game without fixed costs (see section 8.2). In this example, the Stackelberg leader does not own property rights on capital; furthermore, capacity constraints prevent fierce price competition, so that the only way the leader can deter entry is to accumulate enough capacity to serve the whole market at a price equal to the marginal cost of investment plus production (and thus make no profit itself). By simply accumulating the Stackelberg follower's capacity on top of the leader's, the leader does not deter entry. Another example in which preemption is not effective is when the incumbent's investment has no commitment value, as in the example of product dropping considered in the preceding subsection.

Second, the technology of preemption must be *deterministic*. That is, firms must have a means of preempting their rivals. With a nondeterministic technology (as in the patent contests considered in chapter 10), there may be no way for the incumbent to guarantee that it will obtain the technology first.[94]

Third, even in a situation in which preemption is both effective and deterministic, it is hard to believe that a monopolist always keeps his privileged situation. Indeed, several variations of the preemption model give rise to a positive probability that an oligopolistic structure emerges:

(1) The incumbent does not possess the entrant's technology. In this obvious case, the incumbent cannot duplicate the entrant's strategy a bit earlier, which may leave scope for entry. In the product-differentiation model, for instance, the incumbent might not be able to build at the right-end location.

(2) The incumbent may not have time to preempt the entrant. This is the case when an innovation appears that both the incumbent and entrant would like to adopt immediately. Preemption would then require the incumbent to adopt the innovation before its appearance, which is impossible. This lack of time is also present (in a disguised way) in the simultaneous models of entry, in which there is only one period to invest. The models of competition in location and of monopolistic competition analyzed in chapter 7 belong to this category. In the case of the linear city with two locations, it is easy to see that there exist two equilibria in pure strategies if entry decisions are made simultaneously. In one equilibrium, the existing firm is the only one to invest and build at the right end of the city (monopoly persistence); in the other, the entrant is the only one to invest and build at this location (entry).[95]

The lack of time is also implicit in the investment models in which firms are not allowed to build more than one plant or introduce more than one product. The implicit assumption there is that rapid investment (in a second or third plant or product) is very costly, and that single-plant or single-product firms can enter before the established firms can pursue their expansion.

In a sense, the simultaneous-entry models of chapter 7 correspond to very long information lags: There is no way a firm can observe its rivals' choices before making its own. This is clearly an extreme assumption, even in situations where the firms' investment decisions take a while to be observed.[96] More generally, one may consider dynamic rivalry under non-negligible information lags (imperfect information). Fudenberg et al. (1983) consider a game in which, under perfect information (no information lags), only the incumbent conducts research and development, whereas with information lags competition may arise. If the entrant did not try to enter, the incumbent would delay its investment decision (as in the preemption game, the incumbent would like to move "slowly") and it would pay the entrant to enter. But the entrant will try to enter only if it has a chance of being first. Thus,

94. Furthermore, we will see in chapter 10 that, because the date of preemption is random, the incumbent may not want to hasten its own replacement and may therefore have less incentive than the entrant to invest in R&D.

95. As is usual in the simultaneous-entry models of a natural monopoly (where "natural monopoly" refers to the right-end location and not to the whole market), there exists a third, mixed-strategy equilibrium, in which both firms are indifferent between entering and not entering.

96. And where firms cannot communicate their investment decisions in a credible way. Indeed, it may be in the interest of a firm to release investment decisions so as to deter entry. (Firms actually do announce construction of plants or, like IBM, preannounce their products.) Conversely, the absence of announcement signals the absence of investment (unless announcements reveal precious technological information).

there is a positive probability that the monopoly will not persist.

(3) *The existing firm may not have complete information about the entrant's characteristics.* Without complete information, it cannot calculate exactly the optimal preemption date T_2. (In the case of an information lag, the existing firm does not observe its rival's action; in the case of incomplete information, it may not observe its rival's cost structure.) Since the existing firm wants to invest as late as possible (L_1 is increasing) and to preempt entry, it may bide its time even when this entails the risk of entry. The firm must evaluate the gain related to waiting and estimate the probability of not preempting entry. Therefore, incomplete information introduces a nonzero probability that the potential entrant will actually enter. To see this in the context of the model of subsection 8.6.2.1, suppose that the entrant's entry cost is either "high" or "low," and that it is known only to the entrant. The optimal date of preemption, T_2, is much later for a high-cost entrant than for a low-cost entrant. Intuitively, if the probability that the entrant has a high entry cost is sufficiently high, it does not pay for the incumbent to make sure he preempts the low-cost entrant. The benefits of obtaining a higher L_1 offset the loss associated with the probability of being preempted.

Answers and Hints

Exercise 8.1

The situation is depicted in figure 8.21. Let $C = \{p^c, q^c\}$ denote the point at which the average-cost curve intersects the demand curve, let q^* denote the most efficient scale, and let p^* denote the minimum average cost. A market-clearing, profit-making allocation must lie on the demand curve to the northwest of C. In particular, the market price must (weakly) exceed p^c. Now suppose that an entrant comes in at a lower price p^e between p^* and p^c, and produces $q^e = q^*$. That is, the entrant rations the consumers at price p^e. Because the price charged by the entrant strictly exceeds its average cost (which is p^*), the entrant makes a strictly positive profit and the initial allocation is not sustainable. Thus, there exists no sustainable allocation.

Exercise 8.2

(i) There are several meanings to "natural monopoly" (the notion depends on the application to be made). One meaning refers to the socially efficient production pattern. Because of increasing returns to scale, one firm is the optimal arrangement (if its price can be controlled). Another meaning looks at the maximum (upper bound on the) number of firms in the industry. Here, even if firms can somehow collude, they make at most $\Pi^m = \max[p(1-p)] = \frac{1}{4}$. Since $f = \frac{3}{16}$, if there are two firms at least one of them loses money.

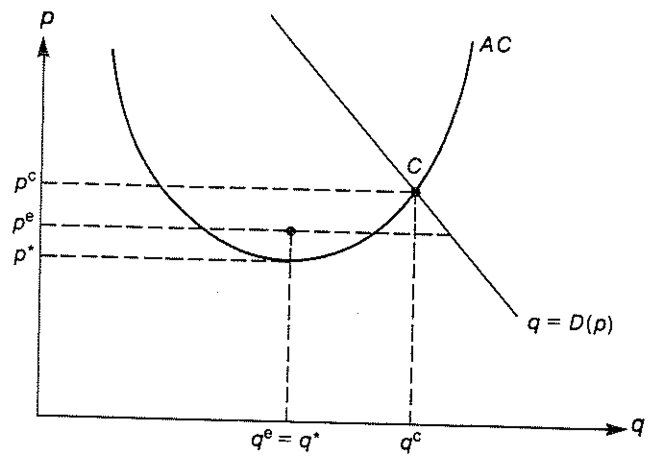


Figure 8.21