Chapter 6

Markets for Homogeneous Products

> Only theory can separate the competitive from the anticompetitive.
> —Robert Bork, *The Antitrust Paradox*

In this chapter we analyze the behavior of firms and consumer welfare under several oligopolistic market structures. The main assumption in this chapter is that the products are *homogeneous,* meaning that consumers cannot differentiate among brands or distinguish among the producers when purchasing a specific product. More precisely, consumers cannot (or just do not bother) to read the label with the producer's name on the product they buy. For example, non-brand-name products sold in most supermarkets—bulk fruit, vegetables, containers of grain—are generally purchased without having consumers learning the producer's name.

In what follows, we assume that consumers are always price takers (henceforth, competitive) and have a well-defined aggregate-demand function. However, firms behave according to the assumed market structures analyzed below.

Our oligopoly analysis starts with section 6.1 (Cournot), which assumes that firms set their output levels simultaneously, believing that the output levels of their rival firms remain unchanged. Historically, as we discuss below, Cournot was the first to provide this modern treatment of oligopoly equilibrium. Section 6.2 (Sequential Moves) modifies the static Cournot setup, by assuming that firms move in sequence, and analyzes whether a firm benefits by setting its output level before any other one does. Following Bertrand's criticism of the use of quantity produced as the actions chosen by firms, section 6.3 (Bertrand) analyzes a market structure where firms set their prices by assuming that the prices of their rival firms remain unchanged. We then discuss how the extreme result of price games leading to competitive prices obtained under the Bertrand competition can be mitigated by introducing capacity constraints. Section 6.4 (Cournot Versus Bertrand) analyzes the relationship between the Cournot and the Bertrand market structures. Section 6.5 (Self-Enforcing Collusion) analyzes the conditions under which firms can maintain higher prices and lower output levels compared with the Cournot levels, assuming that the firms interact infinitely many times. Section 6.6 (International Trade) analyzes international markets in homogeneous products.

6.1 Cournot Market Structure

Noncooperative oligopoly theory started with Antoine Augustin Cournot's book, *Researches into the Mathematical Principles of the Theory of Wealth*, published in France in 1838. In that book, Cournot proposed an oligopoly-analysis method that we today view as identical to finding a Nash equilibrium in a game where firms use their production levels as strategies. Cournot earned his doctorate in science in 1821, with a main thesis in mechanics and astronomy. Cournot's writings extended beyond economics to mathematics and philosophy of science and philosophy of history (see Shubik 1987).

Cournot was central to the founding of modern mathematical economics. For the case of monopoly, the familiar condition where marginal revenue equals marginal cost come directly from Cournot's work (Shubik 1987). In chapter 7 of his book, Cournot employs the inverse-demand function to construct a system of firms' marginal-revenue functions, which could be then solved for what we will call the Cournot output levels. Then, he introduced firms' cost functions and the system of first-order conditions to be solved. Cournot did not consider the possibility that firms with sufficiently high cost may not be producing in this equilibrium.

In what follows, we develop the Cournot oligopoly model where firms sell identical products. In this model, firms are not price takers. Instead, each firm is fully aware that changing its output level will affect the market price.

6.1.1 Two-seller game

Let us consider a two-firm industry summarized by the cost function of each firm i (producing q_i units) given by

$$TC_i(q_i) = c_i q_i, \quad i = 1, 2, \quad \text{where } c_2, c_1 \geq 0, \tag{6.1}$$

and the market-demand function given by

$$p(Q) = a - bQ, \quad a, b > 0, \ a > c_i, \quad \text{where } Q = q_1 + q_2. \tag{6.2}$$

In contrast to chapter 4, where we solved for a competitive equilibrium for this industry, here we solve for a Cournot oligopoly equilibrium. We first have to define a two-firm game that corresponds to a definition of a game given in Definition 2.1. Let each firm's action be defined as choosing its production level, and assume that both firms choose their actions simultaneously. Thus, each firm i chooses $q_i \in A_i \equiv [0, \infty)$, $i = 1, 2$. Also, let the payoff function of each firm i be its profit function defined by $\pi_i(q_1, q_2) = p(q_1 + q_2)q_i - TC_i(q_i)$. Now, the game is properly defined since the players, their action sets, and their payoff functions are explicitly defined. All that is left to do now is to define the equilibrium concept.

DEFINITION 6.1 *The triplet* $\{p^c, q_1^c, q_2^c\}$ *is a* **Cournot-Nash equilibrium** *if*

1. (a) *given* $q_2 = q_2^c$; q_1^c *solves* $\max_{q_1} \pi_1(q_1, q_2^c) = p(q_1 + q_2^c)q_1 - TC_1(q_1) = [a - b(q_1 + q_2^c)]q_1 - c_1 q_1$
   (b) *given* $q_1 = q_1^c$; q_2^c *solves* $\max_{q_2} \pi_2(q_1^c, q_2) = p(q_1^c + q_2)q_2 - TC_2(q_2) = [a - b(q_1^c + q_2)]q_2 - c_2 q_2$
2. $p^c = a - b(q_1^c + q_2^c)$, $\quad p^c, q_1^c, q_2^c \geq 0$.

That is, according to Definition 6.1, a Cournot equilibrium is a list of output levels produced by each firm and the resulting market price so that no firm could increase its profit by changing its output level, given that other firms produced the Cournot output levels. Thus, Cournot equilibrium output levels constitute a Nash equilibrium in a game where firms choose output levels.

Now that the equilibrium concept is well defined, we are left to calculate the Cournot equilibrium for this industry. Firm 1's profit-maximization problem yields the first-order condition given by

$$0 = \frac{\partial \pi_1(q_1, q_2)}{\partial q_1} = a - 2bq_1 - bq_2 - c_1$$

which yields the familiar profit-maximizing condition in which each firm (firm 1 in this equation) sets its marginal revenue ($MR(q_1) = a - 2bq_1 - bq_2$) equal to marginal cost (c_1). The second-order condition guaranteeing a global maximum is satisfied since $\frac{\partial^2 \pi_1}{\partial (q_1)^2} = -2b < 0$ for every q_1 and q_2. Solving for q_1 as a function of q_2 yields the *best-response* *function* (also commonly known as *reaction function*) of firm 1, which we denote by $R_1(q_2)$. Hence,

$$q_1 = R_1(q_2) = \frac{a - c_1}{2b} - \frac{1}{2}q_2. \tag{6.3}$$

Similarly, we can guess that firm 2's best-response function is given by

$$q_2 = R_2(q_1) = \frac{a - c_2}{2b} - \frac{1}{2}q_1. \tag{6.4}$$

The best-response functions of the two firms are drawn in Figure 6.1 in the (q_1, q_2) space.

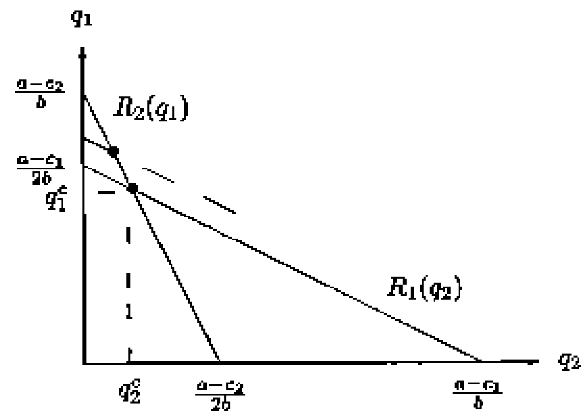

Figure 6.1: Cournot best-response functions (the case for $c_2 > c_1$)

The two best-response functions are downward sloping, implying that for each firm, if the rival's output level increases, the firm would lower its output level. The intuition is that if one firm raises its output level, the price would drop, and hence in order to maintain a high price the other firm would find it profitable to decrease its output level. A perhaps more intuitive explanation for why a firm's best-response function is downward sloping is that an increase in a rival's output shifts the residual demand facing a firm inward. Hence, when a firm faces a lower demand it would produce a smaller amount.

Now, the Cournot equilibrium output levels can be calculated by solving the two best-response functions (6.3) and (6.4), which correspond to the intersection of the curves illustrated in Figure 6.1. Thus,

$$q_1^c = \frac{a - 2c_1 + c_2}{3b} \quad \text{and} \quad q_2^c = \frac{a - 2c_2 + c_1}{3b}. \tag{6.5}$$

Hence, the aggregate industry-output level is $Q^c = q_1^c + q_2^c = \frac{2a - c_1 - c_2}{3b}$, and the Cournot equilibrium price is

$$p^c = a - bQ^c = \frac{a + c_1 + c_2}{3}. \tag{6.6}$$

It is easy to confirm from (6.5) that the output of the high-cost firm is lower than the output level of the low-cost firm. That is, $c_2 \ge c_1$ implies that $q_1 \ge q_2$.

Altogether, the Cournot profit (payoff) level of firm i, as a function of the unit costs for firms i and j, $i \neq j$, is given by

$$\begin{aligned} \pi_i^c &= (p^c - c_i)(q_i^c) = \left(\frac{a + c_i + c_j}{3} - c_i\right)\left(\frac{a - 2c_i + c_j}{3b}\right) \\ &= \frac{(a - 2c_i + c_j)^2}{9b} = b\,(q_i^c)^2. \end{aligned} \tag{6.7}$$

We conclude this section with some comparative static analysis. Suppose that firm 1 invents a new production process that reduces its unit production cost from c_1 to $\bar{c}_1$, where $\bar{c}_1 < c_1$. The type of R&D leading to cost reduction is called "process innovation," to which we will return in Chapter 9. Equation (6.5) implies that q_1^c increases while q_2^c decreases. This is also shown in Figure 6.1, where a decrease in c_1 shifts $R_1(q_2)$ to the right, thereby increasing the equilibrium q_1^c while decreasing q_2^c. Also, (6.6) implies that a decrease in c_1 (or c_2) would decrease the equilibrium price p^c, and (6.7) implies that a decrease in c_1 would increase the profit of firm 1 while lowering the profit of firm 2.

6.1.2 N-seller game

Suppose now the industry consists of N firms, $N \ge 1$. We analyze two types of such industries: (a) N identical firms, all having the same cost function, or (b) heterogeneous firms, where some firms have cost functions different from others. Since solving the general case of firms with different cost functions would require solving N first-order conditions (intersecting N best-response functions), we first solve the model by assuming that all firms have identical technologies. That is, $c_i = c$ for every $i = 1, 2, \dots, N$. In the appendix (section 6.7) we introduce a procedure that makes solving the heterogeneous-firms case easy.

Since all firms have the same cost structure, the first step would be to pick up one firm and calculate its output level as a function of the output levels of all other firms. In other words, we would like to calculate the best-response function of a representative firm. With no loss of generality, we derive the best-response function of firm 1. Thus, firm 1 chooses q_1 to

$$\max_{q_1} \pi_1 = p(Q)q_1 - cq_1 = \left[a - b\left(\sum_{i=1}^{N} q_i\right)\right] q_1 - cq_1.$$

The first-order condition is given by

$$0 = \frac{\partial \pi_1}{\partial q_1} = a - 2bq_1 - b\sum_{i=2}^{N} q_i - c.$$

Hence, the best-response function of firm 1 as a function of the output levels of firms $q_2, q_3, \ldots, q_N$ is given by

$$R_1(q_2, q_3, \ldots, q_N) = \frac{a-c}{2b} - \frac{1}{2}\sum_{i=2}^{N} q_i. \tag{6.8}$$

In the general case, where firms have different cost functions, we would have to derive the best-response function for each of the N firms. However, since all firms are identical, we can guess that in a Cournot equilibrium, the firms would produce the same output levels: we guess (and later verify) that $q_1^c = q_2^c = \ldots = q_N^c$. Thus, we denote the common output level by q, where $q = q_i$ for every i. Note that a common mistake among students is to substitute q for q_i before the best-response functions are derived. This procedure is obviously leading to the wrong solution, since it implies that each firm "controls" the output level of all firms. Therefore, here we substitute the common q only into the already derived best-response functions. The use of symmetry here is purely technical and is done to facilitate solving N equations with N unknowns. From (6.8), we have it that $q = \frac{a-c}{2b} - \frac{1}{2}(N-1)q$. Hence,

$$q^c = \frac{a-c}{(N+1)b} \quad Q^c = Nq^c = \left(\frac{a-c}{b}\right)\left(\frac{N}{N+1}\right). \tag{6.9}$$

The equilibrium price and the profit level of each firm are given by

$$p^c = a - bQ^c = \frac{a+Nc}{N+1} \quad \text{and} \quad \pi_i^c = \frac{(a-c)^2}{(N+1)^2 b} = b\left(q^c\right)^2. \tag{6.10}$$

Varying the number of firms

We now ask how would the Cournot price, quantity produced, and profit levels change when we change the number of firms in the industry? First, note that substituting $N = 1$ into (6.9) and (6.10) yields the monopoly solution described in section 5.1. Second, substituting $N = 2$ yields the duopoly solution described in (6.5), (6.6), and (6.7).

Now, we let the number of firms grow with no bounds, $(N \to \infty)$. Then, we have it that

$$\lim_{N\to\infty} q^c = 0, \quad \text{and} \quad \lim_{N\to\infty} Q^c = \lim_{N\to\infty}\left(\frac{a-c}{b}\right)\left(\frac{N}{N+1}\right) = \left(\frac{a-c}{b}\right). \tag{6.11}$$

That is, in a Cournot equilibrium, as the number of firms grows indefinitely, the output level of each firm approaches zero whereas the industry's aggregate output level approaches the competitive output level given in Proposition 4.1. Also,

$$\lim_{N\to\infty} p^c = \lim_{N\to\infty} \frac{a}{N+1} + \frac{Nc}{N+1} = c = p^e. \tag{6.12}$$

Hence, the Cournot equilibrium price approaches the competitive price that equals the unit production cost of a firm (see Proposition 4.1). These results often cause some confusion among students, leading them to believe that competitive behavior occurs only when there are many (or infinitely many) firms. However, as we pointed out in chapter 4, we can assume a competitive market structure for any given number of firms, and even solve for a competitive equilibrium for the case where $N = 1$. What equations (6.11) and (6.12) say is that the Cournot market structure yields approximately the same price and industry output as the competitive market structure when the number of firms is large.

6.1.3 Cournot equilibrium and welfare

Since our analysis starts with given demand functions (rather than the consumers' utility functions), we cannot measure the social welfare by calculating consumers' equilibrium-utility levels. Instead, we approximate social welfare by adding consumers' surplus and firms' profits (see subsection 3.2.3 on page 52 for a justification of this procedure of welfare approximation). Note that profit should be part of the economy's welfare because the firms are owned by the consumers, who collect the profits via firms' distributions of dividends.

Substituting the Cournot equilibrium price (6.10) into (3.3) on page 52, we obtain the consumers' surplus as a function of the number of firms, N. Hence, $CS^c(N) = \frac{N^2(a-c)^2}{2b(N+1)^2}$. Clearly, $\frac{\partial CS(N)}{\partial N} > 0$, meaning that consumers' surplus rises with the entry of more firms, due to the reduction in price and the increase in the quantity consumed.

We define social welfare as the sum of consumers' surplus plus the industry aggregate profit (see section 4.3 on page 68 for a definition). Thus, if we recall (6.10),

$$W^c(N) \equiv CS^c(N) + N\pi^c(N) \tag{6.13}$$

$$= \left(\frac{(a-c)^2}{2b}\right)\left(\frac{N^2+2N}{N^2+2N+1}\right) = \frac{(a-c)^2}{2b}\bigg|_{N\to\infty}.$$

Also, note that $\frac{\partial W^c(N)}{\partial N} > 0$. Hence, although the industry profit declines with an increase in the number of firms, the increase in consumers' surplus dominates the reduction in the industry profit. Thus, in this economy, free entry is welfare improving!

6.2 Sequential Moves

In the previous section, we analyzed industries where firms strategically choose their output levels. All those games were static in the sense that players simultaneously choose their quantity produced. In this section, we assume that the firms move in sequence. For example, in a two-firm, sequential-moves game, firm 1 will choose its output level before firm 2 does. Then, firm 2, after observing the output level chosen by firm 1, will choose its output level, and only then will output be sold and profits collected by the two firms. This type of market structure is often referred to as *Leader-Follower* on the basis of von Stackelberg's work (1934) (see Konow 1994 for von Stackelberg's biography). This type of behavior defines an extensive form game studied in section 2.2.

In this section we do not raise the important question of what determines the order of moves, that is, why one firm gets to choose its output level before another. We return to this question in chapter 8, where we distinguish among established firms (called incumbent firms) and potential entrants. Here, we assume that the order of moves is given, and we develop the tools for solving an industry equilibrium under a predetermined order of moves.

We analyze a two-stage game, where firm 1 (the leader) chooses the quantity produced in the first stage. The quantity chosen in the first stage is irreversible and cannot be adjusted in the second stage. In the second stage, only firm 2 (the follower) chooses how much to produce after observing the output level chosen by firm 1 in the first stage. Here, the game ends after the second stage, and each firm collects its profit. Our main questions are (a) Is there any advantage for moving in the first stage rather than the second? and (b) How would the equilibrium market price and production levels compare to the static Cournot equilibrium price and output levels?

Following Definition 2.9 on page 26, this game has a continuum of subgames indexed by the output level chosen by firm 1 in the first stage. A finite-horizon dynamic game is generally solved backwards. We look for a subgame perfect equilibrium (Definition 2.10 on page 27) for this game. Hence, we first analyze the players' (firm 2 in our case) action in the last period, assuming that the actions played in previous period are given. Then, we go one period backwards, and analyze firm 1's action given the *strategy* (see Definition 2.3 on page 24) of how firm 2 chooses its output level based on the first-period action. To simplify the exposition, let all firms have identical unit cost, $c_1 = c_2 = c$.

The second-period subgames

In the second period, only firm 2 moves and chooses q_2 to maximize its profit, taking firm 1's quantity produced, q_1, as given. As you probably noticed, we have already solved this problem before, since the second-period problem of firm 2 is identical to the problem firm 2 solves in a Cournot market structure. This maximization results in the best-response function of firm 2 given in (6.4). Hence, $R_2(q_1) = \frac{a-c}{2b} - \frac{1}{2}q_1$. Note that the function $R_2(q_1)$ constitutes firm 2's strategy for this game, since it specifies its action for every possible action chosen by firm 1.

The first-period game

In period 1, firm 1 calculates $R_2(q_1)$ in the same way as firm 2. Thus, firm 1 is able to calculate how firm 2 will best reply to its choice of output level. Knowing that, firm 1 chooses q_1^s to

$$\max_{q_1} \pi_1^s = p(q_1 + R_2(q_1))q_1 - cq_1 = \left[a - b\left(q_1 + \frac{a-c}{2b} - \frac{q_1}{2}\right)\right] q_1 - cq_1. \tag{6.14}$$

We leave it to the reader to derive the first- and second-order conditions. Thus, the quantity produced by the leader is

$$q_1^s = \frac{a-c}{2b} = \frac{3}{2}q_1^c > q_1^c. \tag{6.15}$$

Hence, under the sequential-moves market structure, the leader produces a higher level of output than the Cournot market structure. Substituting (6.15) into $R_2(q_1)$ yields the followers' equilibrium-output level:

$$q_2^s = \frac{a-c}{4b} = \frac{3}{4}q_2^c < q_2^c \tag{6.16}$$

implying that the follower's output level falls compared with the Cournot output level. Thus, the leader's gain in output expansion comes partly from the reduction in the follower's output level. The equilibrium price and aggregate output levels are given by

$$p^s = \frac{a+3c}{4} < \frac{a+2c}{3} = p^c \quad \text{and} \quad Q^s = \frac{3(a-c)}{4b} > \frac{2(a-c)}{3b} = Q^c. \tag{6.17}$$

Therefore,

Proposition 6.1 *A sequential-moves quantity game yields a higher aggregate industry-output level and a lower market price than the static Cournot market structure.*

Thus, the equilibrium market outcome under a sequential-moves game is more competitive than the Cournot equilibrium outcome in the sense that this outcome is somewhere in between the competitive equilibrium outcome derived in chapter 4 and the Cournot outcome derived in section 6.1. The intuition behind Proposition 6.1 is as follow: Under the Cournot market structure, firm 1 perceives the output produced by firm 2 as given. However, under sequential-moves market structure, firm 1 knows firm 2's best-response function and therefore calculates that firm 2 will reduce its output level in response to its increase in output level. Hence, when firm 1 expands output, it expects the price to fall faster under Cournot than under sequential-moves market structure. Therefore, in order maintain a high price, firm 1 will produce more under the sequential game than it will under Cournot. Now, (6.15) and (6.16) demonstrate that the increase in aggregate output stems from the fact that the follower does not find it profitable to cut its output level by the same amount as the increase in the leader's output level. This happens because the reaction functions are sloped relatively flat (slope is negative but exceeds -1), implying that a firm reduces its output level by less than the increase in the output level of the rival firm.

We now compare firms' profit levels under sequential moves to the Cournot profit levels. We leave it to the reader to verify that the leader's profit increases while the follower's declines. That is,

$$\pi_1^s = \frac{(a-c)^2}{8b} > \pi_1^c \quad \text{and} \quad \pi_2^s = \frac{(a-c)^2}{16b} < \pi_2^c, \qquad (6.18)$$

where π_1^c and π_2^c are given in (6.7). Note that we could have concluded even without going into the precise calculations that the leader's profit under the sequential-game equilibrium will be higher than under the Cournot. How? It is very simple! Since firm 2 reacts in a "Nash fashion," firm 1 could just choose to produce the Cournot output level q_1^c. In this case, firm 1 would earn exactly the Cournot profit. However, since in the sequential game firm 1 chooses to produce a different output level, it must be increasing its profit compared with the Cournot profit level. The kind of reasoning we just described is called a *revealed profitability* argument, and the reader is urged to learn to use this kind of reasoning whenever possible because performing calculations to investigate economic effects does not generate an intuitive explanation for these effects. In contrast, logical deduction often provides the necessary intuition for understanding economic phenomena.

Finally, we can logically deduce how industry profit under sequential moves compare with industry profit under Cournot. Equations (6.17) show that the market price under sequential moves is lower than it is under Cournot. Since the Cournot market price is lower than the monopoly's price, and since monopoly makes the highest possible profit, it is clear that industry profit must drop when we further reduce the price below the monopoly's price. Hence, whenever $c_1 = c_2$, industry profit must be lower under sequential moves. In a more general environment, this argument may not holds when the industry profit is not a concave function of p.

6.3 Bertrand Market Structure

In a Cournot market structure firms were assumed to choose their output levels, where the market price adjusted to clear the market and was found by substituting the quantity produced into consumers' demand function. In contrast, in a Bertrand market structure firms set prices rather than output levels. The attractive feature of the Bertrand setup, compared with the Cournot market structure, stems from the fact that firms are able to change prices faster and at less cost than to set quantities, because changing quantities will require an adjustment of inventories, which may necessitate a change in firms' capacity to produce. Thus, in the short run, quantity changes may not be feasible, or may be too costly to the seller. However, changing prices is a relatively low-cost action that may require only a change in the labels displayed on the shelves in the store. Let us turn to the Bertrand market structure.

In 1883 Joseph Bertrand published a review of Cournot's book (1838) harshly critical of Cournot's modeling. It seems, however, that Bertrand was dissatisfied with the general modeling of oligopoly rather than with the specific model derived by Cournot. Today, most economists believe that quantity and price oligopoly games are both needed to understand a variety of markets. That is, for some markets, an assumption that firms set quantities may yield the observed market price and quantity produced, whereas for others, a price-setting game may yield the observed market outcomes. Our job as economists would then be to decide which market structure yields a better approximation of the observed price and quantity sold in each specific market.

We now analyze the two-firm industry defined in (6.1) and (6.2) and look for a Nash equilibrium (see Definition 2.4) in a game where the two firms use their prices as their actions. First, note that so far, our analysis has concentrated on a single market price determined by our assumption that consumers are always on their demand curve. However, in a Bertrand game we have to consider outcomes where each firm sets a different price for its product. Therefore, we now make two explicit assumptions about consumers' behavior under all possible prices announced by both firms:

1. Consumers always purchase from the cheapest seller.
2. If two sellers charge the same price, half of the consumers purchase from firm 1 and the other half purchase from firm 2.

Formally, we modify the demand given in (6.2) to capture the quantity demand faced by each firm i, $i = 1, 2$. Therefore, we assume that

$$q_i = \begin{cases} 0 & \text{if } p_i > a \\ 0 & \text{if } p_i > p_j \\ \frac{a-p}{2b} & \text{if } p_i = p_j \equiv p < a \\ \frac{a-p_i}{b} & \text{if } p_i < \min\{a, p_j\} \end{cases} \qquad i = 1, 2,\ i \neq j. \tag{6.19}$$

Equation (6.19) is the quantity demand facing firm i at any given p_1 and p_2 and incorporates what is commonly called a *rationing rule*, which tells us how the market demand is divided between two firms selling a homogeneous product. Thus, if firm i charges a higher price than firm j, then no consumer would purchase the product from firm i. In contrast, if $p_i < p_j$, then all the consumers will purchase only from firm i, and none will purchase from firm j. In this case, the quantity demanded from firm i is calculated directly from (6.2). Finally, if both firms charge the same prices, then the quantity demand determined in (6.2) is equally split between the two firms.

Definition 6.2 *The quadruple* $\{p_1^b, p_2^b, q_1^b, q_2^b\}$ *is a* **Bertrand-Nash equilibrium** *if*

1. *given* $p_2 = p_2^b$, p_1^b *maximizes* $\max_{p_1} \pi_1(p_1, p_2^b) = (p_1 - c_1)q_1$
2. *given* $p_1 = p_1^b$, p_2^b *maximizes* $\max_{p_2} \pi_2(p_1^b, p_2) = (p_2 - c_2)q_2$
3. q_1 *and* q_2 *are determined in (6.19).*

Definition 6.2 states that in a Bertrand-Nash equilibrium, no firm can increase its profit by unilaterally changing its price.

In the next two subsections we apply Definition 6.2 to two types of markets: the first, where firms do not have capacity constraints and can produce any amount they wish under the assumed cost structure; and the second, where we assume that firms' capacities are limited and therefore, in the short run, they are unable to expand production.

6.3.1 Solving for Bertrand equilibrium

Before we characterize the Bertrand equilibria, it is important to understand the discontinuity feature of this game. In the Cournot game, the payoff (profit) functions are continuous with respect to the strategic variables (quantities); in the Bertrand price game, by contrast, equation (6.19) exhibits a discontinuity of the payoff functions at all the outcomes where $p_1 = p_2$. That is, if one firm sells at a price that is one cent higher than the other firm, it would have a zero market share. However, a two-cent price reduction by this firm would give this firm a one 100 percent market share. The action of a firm to slightly reduce the price below that of its competitor is called *undercutting*. Since undercutting involves setting a price slightly lower than the competitor's, we need to examine the types of currencies used in order to determine the smallest possible undercutting actions available to firms. Therefore, we make the following definition:

Definition 6.3 *Let* ϵ *be the smallest possible monetary denomination (smallest legal tender). The medium of exchange (money) is said to be* **continuous** *if* $\epsilon = 0$, *and* **discrete** *if* $\epsilon > 0$.

Examples of discrete smallest legal tenders are: in China, $\epsilon = 1$ Fen; in Finland, $\epsilon = 10$ Penniä; in Israel, $\epsilon = 5$ Agorot; and in the US, $\epsilon = 1$ cent.

The following proposition characterizes Bertrand equilibria.

Proposition 6.2

1. *If the medium of exchange is continuous and if the firms have the same cost structure,* $(c_2 = c_1 \equiv c)$, *then a Bertrand equilibrium is* $p_1^b = p_2^b = c$, *and* $q_1^b = q_2^b = (a - c)/(2b)$.
2. *Let the medium of exchange be discrete, and assume that* c_2 *is denominated in the medium of exchange. That is,* $c_2 = \lambda\epsilon$, *where* $\lambda \geq 2$ *is an integer. Also let* ϵ *be sufficiently small, that is, satisfying* $(c_2 - \epsilon - c_1)\left(\frac{a-c_2+\epsilon}{b}\right) > (c_2 - c_1)\left(\frac{a-c_2}{2b}\right)$. *If* $c_2 - c_1 > \epsilon$, *then* $p_2 = c_2$, $p_1 = c_2 - \epsilon$, $q_2^b = 0$, *and* $q_1^b = (a - c_2 + \epsilon)/b$ *constitute a Bertrand equilibrium.*

Thus, if firms have equal unit costs, the Bertrand equilibrium price and aggregate output are the same as for the competitive equilibrium. In other words, undercutting reduces the prices to marginal cost. In cases where firm 1 has a lower unit cost than firm 2, firm 1 undercuts firm 2 by charging the highest possible price that is lower than c_2, which is given by $p_1 = c_2 - \epsilon$.

Proof. Part 1: In equilibrium, each firm must make nonnegative profit. Hence, $p_i^b \ge c_i$, $i = 1, 2$.

We first establish that in a Bertrand equilibrium both firms charge the same prices. By way of contradiction suppose that $p_1^b > p_2^b > c$. Then, by (6.19), firm 1 makes zero profit. However, since the medium of exchange is continuous, firm 1 can increase its profit by reducing its price to $p_2^b > \tilde{p}_1 > c$ and grab the entire market, thereby making strictly positive profit, a contradiction.

By way of contradiction suppose that $p_1^b > p_2^b = c$. Then, since the medium of exchange in continuous, firm 2 can raise its price slightly while still maintaining a lower price than firm 1. Hence, firm 2 will deviate, a contradiction.

Now that we have established that $p_1^b = p_2^b$, by way of contradiction assume that $p_1^b = p_2^b > c$. Clearly, this cannot constitute a Nash equilibrium in prices since firm 1, say, would have an incentive unilaterally to reduce its price to $\tilde{p}_1 = p_1^b - \epsilon$, where ϵ can be as small as one wants, thereby grabbing the entire market. For ϵ sufficiently small, this deviation is profitable for firm 1.

Part 2: To briefly sketch the proof of part 2, observe that firm 2 makes a zero profit and cannot increase its profit by unilaterally raising its price above $p_2^b = c_2$. Hence, firm 2 does not deviate. Now, for firm 1 to be able to sell a positive amount, it must set $p_1^b \le c_2$. If $p_1^b = c_2 = p_2^b$, then (6.19) implies that the firms split the market by selling each $q_i = \frac{a-c_2}{2b}$. In this case, the profit of firm 1 is

$$\pi_1 = (c_2 - c_1)q_1 = (c_2 - c_1)\frac{a - c_2}{2b}. \tag{6.20}$$

However, if firm 1 undercuts by the smallest legal tender, then it becomes the sole seller and sells $q_1 = \frac{a-(c_2-\epsilon)}{b}$. In this case,

$$\pi_1 = (c_2 - \epsilon - c_1)q_1 = (c_2 - \epsilon - c_1)\frac{a - (c_2 - \epsilon)}{b}. \tag{6.21}$$

Comparing (6.20) with (6.21) yields the condition stated in part 2. ■

6.3.2 Bertrand under capacity constraints

The previous section demonstrated that when the firms have the same cost structure, price competition reduces prices to unit costs, thereby making firms earn zero profits. Economists often feel uncomfortable with this result, especially since it makes the number of firms in the industry irrelevant, in the sense that under symmetric Bertrand competition, price drops to unit cost even when there are only two firms. Now, if most industries are indeed engaged in a Bertrand competition

as described in this section, then we should observe unit-cost prices for those industries with two or more firms. If this case is realistic, then the antitrust authority should not have to worry about industries' concentration levels and should devote all its effort to fighting monopolies. Clearly, we rarely observe intense price competition among industries with a small number of firms, and therefore the antitrust authority challenges mergers of firms that lead to highly concentrated industries (see Section 8.6).

One way to overcome this problem is to follow Edgeworth (1925) and to assume that in the short run, firms are constrained by given capacity that limits their production levels. The Irish economist Francis Ysidro Edgeworth, who made enormous contributions to economic theory and other disciplines, identified some discontinuity properties of the firms' profit functions when firms produce under increasing marginal cost (decreasing returns to scale) technologies. In Edgeworth's words (Edgeworth 1925, 118):

> In the last case there will be an intermediate tract through which the index of value will oscillate, or rather vibrate irregularly for an indefinite length of time. There will never be reached that determinate position of equilibrium which is characteristic of perfect competition.

We demonstrate Edgeworth's argument by assuming an extreme version of increasing marginal cost, which is letting the cost of expanding production beyond a certain output level (which we call capacity) be infinite. Figure 6.2 illustrates a market-demand curve composed of four consumers, each buying, at most, one unit.

Figure 6.2 assumes that consumer 1 is willing to pay a maximum of \$3 for one unit, consumer 2 a maximum of \$2, consumer 3 a maximum of \$1, and consumer 4 will not pay at all. Such prices are commonly termed as consumers' *reservation prices*.

Suppose now that there are two firms and that each is capable of producing at zero cost, $c_1 = c_2 = 0$. Then, Proposition 6.2, proved in the previous subsection, shows that if firms are not subject to capacity constraints, then Bertrand competition would lead to prices of zero, $p_1^b = p_2^b = 0$.

To demonstrate Edgeworth's argument, suppose now that in the short run each firm is limited to producing, at most, two units. Then, it is easy to show that the prices $p_1 = p_2 = 0$ no longer constitute a Nash equilibrium. To see this, observe that firm 1 can increase its profit from $\pi_1 = 0$ to $\pi_1 = 3$ by increasing its price to $p_1 = 3$, and selling its unit to the consumer with the highest reservation price. In this outcome, firm 1 sells one unit to the consumer with a reservation price of 3, whereas firm 2 sells a unit to one of the other consumers for the price of $p_2 = 0$. Since one firm would always want to deviate from the unit cost pricing, we conclude that the Bertrand equilibrium prices under no capacity constraints need not be Nash equilibrium prices under capacity constraints.

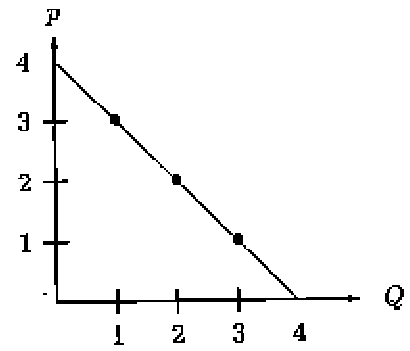

Figure 6.2: Edgeworth Cycles: Bertrand competition under capacity constraints

We are left to show that in the present example there does not exist a Nash equilibrium in prices. This result is sometimes referred to as *Edgeworth Cycles* since under any pair of firms' prices, one firm would always find it profitable to deviate. To see this, let us look at the outcome $p_1 = 3$ and $p_2 = 0$. Clearly, firm 2 would deviate and undercut firm 1 by setting $p_2 = 3 - \epsilon$, where ϵ is a small number. In this case firm 1 sells nothing, whereas firm 2 sells its unit to the consumer with the highest reservation price, and earns a profit of $\pi_2 = 3 - \epsilon \approx 3$. Clearly, firms continue undercutting each other's prices and a Nash equilibrium in prices is never reached. Hence, we showed that marginal-cost pricing is not an equilibrium under capacity constraint, and that firms will keep changing prices without reaching any Nash equilibrium in prices.

Finally, it should be pointed out that introducing capacity constraints on the firms is not the only way to generate above-marginal-cost equilibrium prices. Above-marginal-cost pricing can be an equilibrium outcome (a) when products are differentiated (see next chapter), (b) when demand randomly fluctuates, and (c) when firms are engaged in an infinitely pricing repeated game.

6.4 Cournot versus Bertrand

In sections 6.1 and 6.3 we analyzed the same industry where in the Cournot-market-structure firms use quantity produced as actions, whereas

in the Bertrand-market-structure firms use prices as actions. The analyses of these sections show that in general, the two types of market structures yield different market outcomes (prices and quantity produced). Thus, when we change the firms' actions from choosing quantities to choosing prices, the Nash equilibrium yields a completely different outcome because under Cournot, firms make positive profit, since the resulting market price exceeds unit cost, whereas under Bertrand, prices drop to unit cost. Moreover, in a Bertrand game, only the low-cost firm produces, which is generally not the case for the Cournot game. Therefore, we can state that in a one-shot (static) game there is no correspondence between the Cournot solution and the Bertrand solution.

However, Kreps and Scheinkman (1983) constructed a particular environment (a particular two-period dynamic game) where; in the first period, firms choose quantity produced (accumulate inventories) and in the second period, the quantities are fixed (cannot be changed) and firms choose prices. They showed that the quantities chosen by firms in the first period and the price chosen in the second period are exactly the Cournot outcome given in (6.5) and (6.6). That is, they show that for some market games where two firms choose how much to produce in period 1, and then set prices in period 2, a subgame perfect equilibrium (see Definition 2.10 on page 27) yields the exact quantity produced and price as those in a one-shot Cournot-market-structure game, where firms choose only how much to produce.

We will not bring a complete proof of their proposition; however, we illustrate the idea in our simple two-firm industry for the case where $p = 10 - Q$, and both firms have a unit cost of $c = 1$.

As we discussed earlier, the easiest way of solving for a subgame perfect equilibrium for a dynamic finite game is to solve it backwards. Therefore, we begin with the second period and ask what prices will be chosen by firms in a Nash-equilibrium one-shot price game, where the quantity produced is taken as given by first-period choices. Then, we analyze the first period looking for a subgame perfect equilibrium in first-period production levels, where firms can calculate and take into account the second-period equilibrium market prices, which depend on first-period production levels.

The second-period subgame

Assume that for some reason, the firms choose to produce the Cournot capacity levels $q_1^c = q_2^c = 3$. Hence, total industry output is $Q^c = 6$. We now show that in a Nash equilibrium for the second-period subgame both firms will choose to set prices that clear the market under the Cournot outcome. That is, each firm will set $p_i = 4 = p^c$. Figure 6.3

illustrates the Cournot outcome.

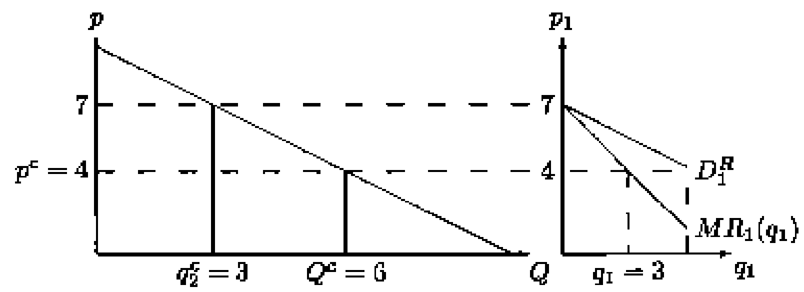

Figure 6.3: Residual demand when firms have fixed inventories

Note that in the second period, firms are free to choose any price they wish so that the Nash equilibrium prices may differ from $p^c = 4$. To demonstrate that this is not the case, we now show that given $p_2 = 4$, firm 1 will not deviate and will also choose $p_1 = 4$. First, note that firm 1 will not lower its price below $p_1 = 4$ because a price reduction will not be followed by an increase in sales (the capacity is limited to $q_1 = 3$.) Thus, lowering the price will only lower its revenue.

Second, we must show that firm 1 cannot increase its profit by raising its price and selling less than $q_1^c = 3$. The right side of Figure 6.3 exhibits the *residual* demand facing firm 1 when it raises its price above $p_1^c = 4$. Residual demand is the demand facing firm 1 after the quantity supplied by firm 2 is subtracted from the aggregate industry demand. In the present case, we subtract $q_2^c = 3$ from the aggregate demand curve to obtain the residual demand curve facing firm 1, given by $q_1 = 10 - p - 3 = 7 - p$ or its inverse $p = 7 - q_1$. The most important observation to be made about Figure 6.3 is that the marginal-revenue curve derived from this residual-demand function ($MR_1(q_1) = 7 - 2q_1$) is strictly positive for all output levels satisfying ($q_1 \le 3.5$), implying that the residual demand is elastic at this interval. Therefore, increasing p_1 will only reduce the revenue of firm 1. This establishes the following claim.

Lemma 6.1 *If the output (capacity) levels chosen in period 1 satisfy $q_1 + q_2 \le 6$, then the Nash equilibrium exhibits both firms choosing the market-clearing price in the second period.*

Lemma 6.1 shows that, given firms' choices of output levels, in the second-period price game firms will strategically choose to play the market price that clears the market at the given aggregate output level.

The first-period game

In the first period, firms observe that the second-period price would be the market-clearing price (Lemma 6.1). Therefore, for each firm, the first-period-capacity-choice problem is precisely the Cournot-quantity-choice problem as formulated in Definition 6.1. Hence, in the first period, firms would choose the Cournot quantity levels $q_1^c = q_2^c = 3$. Intuitively, in the first period both firms know that the second-period price choices by both firms would be the price that clears the market for the first-period production levels. This knowledge makes the firms' first-period-output-choice problem identical to firms' output choices in a Cournot market structure as defined in Definition 6.1.

Finally, note that this illustration does not provide a complete proof for this statement, since in Lemma 6.1 we assumed that the firms did not choose "very high" capacity levels in the first period. In that respect, Lemma 6.1 is not proven for output levels exceeding $q_1 + q_2 > 6$. We refrain from proving that in order to avoid using mixed strategies in this book. Also, from time to time this result causes some confusion among students and researchers, leading them to state that there is no reason for using Bertrand price competition anymore since the two-period, capacity-price game would yield the same outcome as the Cournot market structure. Note that this statement is too strong, since it holds only for the particular two-period game analyzed in the present section.

6.5 Self-Enforcing Collusion

In this section we extend the basic static Cournot game to an infinitely repeated game in which firms produce output and collect profits in each period. Although the analysis in this section is self-contained, the reader is urged to obtain some background on repeated games by reading section 2.3.

One very important result will emerge from analyzing an infinitely repeated Cournot game, namely, that the outcome in which all firms produce the collusive output levels (see the cartel analysis in subsection 5.4.1) constitutes a subgame perfect equilibrium for the noncooperative repeated Cournot game. More precisely, in subsection 6.1.2 we proved that under the Cournot market structure with two or more firms, aggregate industry output exceeds the monopoly output level (which equals the cartel's total output level). Moreover, we showed that as the number of firms increases, the output level increases and converges to the competitive output level. Altogether, firms have a lot to gain by colluding rather than competing under any market structure. In this section we show that if the Cournot game is repeated infinitely, then the collusive output level can emerge as a noncooperative equilibrium. The importance of this result is that it implies that observing an industry where production levels are limited and firms make strictly positive profits does not imply that the firms are engaged in any cooperative activities. In fact, what we show in this section is that the cooperative collusive output levels can be sustained as a noncooperative equilibrium.

In the subsection 6.5.1 we develop a simple Cournot duopoly model and analyze the incentives to collude among firms and the incentive for each firm to unilaterally deviate from collusion when the game is played only once. Subsection 6.5.2 analyzes equilibrium outcome when the one-shot game is repeated infinitely.

6.5.1 The one-shot game

Consider the following basic one-shot Cournot game: There are two firms denoted by $i = 1, 2$. We denote by q_i the output level of firm i. The demand facing the industry is $p = 1 - q_1 - q_2$. Let $Q \equiv q_1 + q_2$ denote the aggregate industry-output level, and assume that production is costless.

In the following subsubsections we quickly derive the already familiar Cournot duopoly equilibrium, the collusion (cooperative) monopoly equilibrium, and then the incentives to deviate from the cooperative outcome.

Duopoly: Non-cooperative behavior

In view of Definition 6.1, in a Cournot market structure firm 1 maximizes $\pi_1 = (1-q_1-q_2)q_1$, yielding a best-response function: $q_1(q_2) = (1-q_2)/2$ and the equilibrium output levels $q_1 = q_2 = 1/3 \equiv M$, where M stands for medium production level. Hence, $Q = 2/3$, and $p = 1/3$, implying that $\pi_i = 1/9$. The profits of the firms under duopoly are displayed in the second column and second row of Table 6.1.

		Firm 2: $q_2 = L = \frac{1}{4}$		$q_2 = M = \frac{1}{3}$		$q_2 = H = \frac{3}{8}$	
	$q_1 = L = \frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{5}{48}$	$\frac{5}{36}$	$\frac{3}{32}$	$\frac{9}{64}$
Firm 1	$q_1 = M = \frac{1}{3}$	$\frac{5}{36}$	$\frac{5}{48}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{7}{72}$	$\frac{7}{64}$
	$q_1 = H = \frac{3}{8}$	$\frac{9}{64}$	$\frac{3}{32}$	$\frac{7}{64}$	$\frac{7}{72}$	$\frac{3}{32}$	$\frac{3}{32}$

Table 6.1: Cooperation L; Noncooperative Cournot duopoly M; Defection from cooperation H.

Collusion: Cooperative behavior

We assume that when the two firms collude, they act as a cartel, analyzed in subsection 5.4.1. Since the firms have identical technologies that exhibit constant returns to scale, the present case is easy to analyze because under CRS there is no difference whether under collusion they operate one or two plants. In any case, the cartel's profit-maximizing output is found by equating $MR(Q) = 1 - 2Q = 0 = MC_i$, implying that $Q = 1/2$, $p = 1/2$. Hence, equal division of output between the two colluding firms imply that $q_i = L = 1/4$, where L stands for "low" output levels. Thus, as expected, collusion implies that both firms restrict their output levels below the Cournot output levels. The two firms equally divide the profit, so $\pi_i = pQ/2 = 1/8$, which is displayed in the first column and row in Table 6.1.

Deviation from collusion

Suppose that firm 2 plays the naive collusive output level $q_2 = L$. We now show that in this one-shot game, firm 1 can increase its profit by unilaterally increasing its output level. To see that, for given $q_2 = 1/4$, firm 1 chooses q_1 to $\max \pi_1 = (1 - q_1 - 1/4)q_1$, yielding $0 = 3/4 - 2q_1$. Hence, $q_1 = 3/8 \equiv H$. Thus, if firm 2 does not deviate from $q_2 = L$, firm 1 has the incentive to increase its output to a high level. In this case, $Q = 3/8 + 1/4 = 5/8$, $p = 3/8$, $\pi_1 = 9/64$ and $\pi_2 = 3/32$; both are displayed in the first column, third row in Table 6.1.

Equilibrium in the one-shot game

The first part of the next proposition follows directly from equation (6.5) and also from Table 6.1. The second part follows from Definition 2.6 and Table 6.1.

Proposition 6.3 *In the one-shot game:*

1. *there exists a unique Cournot-Nash equilibrium, given by* $q_1 = q_2 = M = 1/3$;
2. *the equilibrium outcome is Pareto dominated by the "cooperative outcome"* $q_1 = q_2 = L = 1/4$.

Note that we use the Pareto criterion to refer only to the profit of firms, thereby disregarding consumers' welfare.

6.5.2 The infinitely repeated game

Suppose now that the two firms live forever. The game proceeds as follows: In each period t both firms observe what both firms played in *all* earlier periods (observe period t history as defined in Definition 2.11) and then play the one-shot game described in Table 6.1. That is, in each period t, each firm i chooses $q_i(t)$, where $q_i(t) \in \{L, M, H\}$, $i = 1, 2$ and $t = 1, 2, \ldots$. A strategy of firm i is a list of output levels chosen each period by firm i after the firm observed all the output levels chosen by each firm in all earlier periods (see Definition 2.11 for a precise definition of a strategy in repeated games).

Let $0 < \rho < 1$ be the discount factor. Note that in perfect capital markets, the discount factor is inversely related to the interest rate. Let r denote the interest rate. Then, $\rho = \frac{1}{1+r}$. As r rises, ρ falls, meaning that future profits are less valuable today. Following Assumption 2.1, we assume that the objective of each firm is to maximize the sum of present and discounted future profits given by

$$\Pi_i \equiv \sum_{t=1}^{\infty} \rho^{t-1} \pi_i(t) \tag{6.22}$$

where the values of $\pi_i(t)$ are given in Table 6.1.

The trigger strategy

We restrict the discussion here to one type of strategies called *trigger strategies*, meaning that in every period τ each player cooperates (playing $q_i(\tau) = L$) as long as all players (including himself) cooperated in all periods $t = 1, \ldots, \tau - 1$ (see Definition 2.11 for a precise definition). However, if any player deviated in some period $t \in \{1, \ldots, \tau - 1\}$, then player i plays the noncooperative (duopoly) strategy forever! That is, $q_i(t) = M$ for every $t = \tau, \tau + 1, \tau + 2, \ldots$. Formally, let us restate Definition 2.12 for the present game.

DEFINITION 6.4 *Player i is said to be playing a* **trigger strategy** *if for every period* τ, $\tau = 1, 2, \ldots,$

$$q_i(\tau) = \begin{cases} L & \text{as long as } q_1(t) = q_2(t) = L \text{ for all } t = 1, \ldots, \tau - 1 \\ M & \text{Otherwise.} \end{cases}$$

In other words, firm i cooperates by restricting its output as long as all firms restrict their output levels in earlier periods. However, if any firm deviates even once, then firm i produces the static Cournot-Nash duopoly output level forever.

Equilibrium in trigger strategies

We now seek to investigate under what conditions playing trigger strategies constitutes a subgame perfect equilibrium (see Definition 2.10). It turns out that for a small discount factor, a firm may benefit by deviating from the cooperative output level, thereby collecting a temporary high profit by sacrificing the extra future profits generated by cooperation. However, for a sufficiently large discount factor we can state the following proposition:

Proposition 6.4 *If the discount factor is sufficiently large, then the outcome where both firms play their trigger strategies is a SPE. Formally, trigger strategies defined in Definition 6.4 constitute a SPE if* $\rho > 9/17$.

Proof. We look at a representative period, call it period τ, and suppose that neither firm has deviated in periods $t = 1, \ldots, \tau - 1$. Then, if firm 1 deviates and plays $q_1(\tau) = H$ (the best response to $q_2(\tau) = L$), Table 6.1 shows that $\pi_1(\tau) = 9/64 > 1/8$. However, given that firm 1 deviates, firm 2's equilibrium strategy calls for playing $q_2(t) = M$ for every $t \geq \tau + 1$. Hence the period $\tau + 1$ sum of discounted profits of firm 1 for all periods $t \geq \tau + 1$ is $\frac{1}{1-\rho}\frac{1}{9}$. Note that we used the familiar formula for calculating the present value of an infinite stream of profits given by $1 + \rho + \rho^2 + \rho^3 + \ldots = \sum_{t=0}^{\infty} \rho^t = \frac{1}{1-\rho}$. Hence, if firm 1 deviates in period τ, its sum of discounted profits is

$$\Pi_1 = \frac{9}{64} + \frac{\rho}{1-\rho}\frac{1}{9}. \tag{6.23}$$

However, if firm 1 does not deviate in period τ, then both firms continue producing the collusive output yielding

$$\Pi_1 = \frac{1}{1-\rho}\frac{1}{8}. \tag{6.24}$$

Comparing (6.23) with (6.24) yields the conclusion that deviation is *not* profitable for firm 1 if $\rho > 9/17$.

As we noted in the proof of Proposition 2.5, to prove subgame perfection we need to show that each firm would find it profitable to respond with deviation when it realizes that deviation occurred in an earlier period, as stated in the definition of the trigger strategy described in Definition 6.4. That is, we still need to show that a firm would produce a level of M forever once either firm deviated in an earlier period. In the language of game theorists, we need to show that the trigger strategy is the best response even if the game "drifts" off the equilibrium path. However, Definition 6.4 implies that if firm j deviates, then firm j would produce M in all future periods. Then, Table 6.1 shows that firm i's best response to firm j's playing M is to play M. Hence, the trigger strategies defined in Definition 6.4 constitute a SPE. ■

Discussion of trigger strategies and extensions

The purpose of section 6.5 was to demonstrate that in an infinitely repeated game, the set of oligopoly equilibria is larger than that of a one-shot game and includes cooperative outcomes in addition to the familiar noncooperative outcome. Readers who wish to learn more about cooperation in oligopolistic market structures are referred to Abreu 1986, Friedman 1971, 1977, Green and Porter 1984, Segerstrom 1988, Tirole 1988, chap. 5, and more recent books on game theory noted in the references to chapter 2.

We conclude our analysis of dynamic collusion with two remarks: (a) We have not discussed what would happen to our cooperative equilibrium when we increase the number of firms in the industry. Lambson (1984) has shown that under general demand conditions the cooperation continues to hold as long as the demand for the product increases at the same rate as the number of firms. The intuition behind this result is as follows: If the number of firms grows over time but the demand stays constant, then the future profit of each firm would drop, implying that firms would have a stronger incentive to deviate from the collusive output level. Hence, in such a case, collusion is less likely to be sustained. (b) Another natural question to be asked is how booms and recessions affect the possibility of collusion among firms. Rotemberg and Saloner (1986) analyze collusion under stochastic demand. The problem they investigate is whether collusion is more sustainable during booms (a high realization of the demand) than during recessions (a low demand realization).

6.6 International Trade in Homogeneous Products

In this section we analyze two issues related to international trade in homogeneous products. Subsection 6.6.1 demonstrates the possibility that countries sell homogeneous products below cost in other countries. Subsection 6.6.2 evaluates how the formation of customs unions and free trade agreements affect international trade in homogeneous products.

6.6.1 Reciprocal dumping in international trade

An application of the Cournot equilibrium for international trade is given in Brander and Krugman 1983. Suppose that there are two identical trading countries indexed by k, $k = 1, 2$. The demand schedule in each country is given by $p_k(Q_k) = a - bQ_k$, where Q is the sum of local production and import. In each country there is one firm producing a homogeneous product that is sold both at home and abroad. To keep this example simple, assume that production is costless, that is, $c = 0$.

The two countries are separated by an ocean, and therefore, shipping the good across the continents is costly. Also, assume that the transportation cost is paid by the exporting firm.

Let τ denote the per-unit international transportation cost, and let q_k denote the production level of the firm located in country k, $k = 1, 2$. Since each firm sells both at home and abroad, the output of firm k is decomposed into home (local) sales (denoted by q_k^h) and foreign (export) sales (denoted by q_k^f). Therefore, the total output sold in country 1 is $Q_1 = q_1^h + q_2^f$, and the total output sold in country 2 is $Q_2 = q_2^h + q_1^f$.

The profit of each firm is the revenue collected in each country minus the cost of production (assumed to be zero) minus export transportation cost. Formally, the profit of the firm located in country 1 is

$$\pi_1 = p_1(q_1^h + q_2^f)q_1^h + p_2(q_1^f + q_2^h)q_1^f - \tau q_1^f. \tag{6.25}$$

The profit of the firm located in country 2 is

$$\pi_2 = p_2(q_2^h + q_1^f)q_2^h + p_1(q_2^f + q_1^h)q_2^f - \tau q_2^f. \tag{6.26}$$

The first-order conditions for (6.25) are

$$0 = \frac{\partial \pi_1}{\partial q_1^h} = a - 2bq_1^h - bq_2^f \quad \text{and} \quad 0 = \frac{\partial \pi_1}{\partial q_1^f} = a - 2bq_1^f - bq_2^h - \tau.$$

Notice that the two first-order conditions are independent in the sense that q_1^f (foreign sales) does not appear in the first condition and q_1^h (home sales) does not appear in the second. This follows from our particular use of the linear cost structure. In general, when the cost function is nonlinear, the two conditions would not be independent. The first-order conditions for (6.26)

$$0 = \frac{\partial \pi_2}{\partial q_2^h} = a - 2bq_2^h - bq_1^f \quad \text{and} \quad 0 = \frac{\partial \pi_2}{\partial q_2^f} = a - 2bq_2^f - bq_1^h - \tau.$$

Using this special case, we can solve for the Cournot equilibrium output levels for each country separately. In this case (6.5) implies that for firm k, $k = 1, 2$,

$$q_k^h = \frac{a+\tau}{3b}, \quad q_k^f = \frac{a-2\tau}{3b}, \quad Q_k = \frac{2a-\tau}{3b}, \quad \text{and} \quad p_k = \frac{a+\tau}{3}. \tag{6.27}$$

Note that as transportation becomes more costly (τ increases), the share of domestic sales increases in each country, whereas the level of export declines. Also, as τ increases, p_k increases.

Dumping

One of the major rules of GATT (General Agreement on Tariffs and Trade) is that dumping is prohibited. Before we define dumping we need to distinguish between two types of prices used in international transactions: (a) FOB price (free-on-board), meaning the price received by the producer when the product leaves the plant. This price does not include the payments for transportation and insurance. (b) CIF price (cost-insurance-freight), which includes all transportation as well as insurance costs. If we assume away dealers, which would make the CIF price even higher, the consumer pays the CIF price, whereas the exporter receives the FOB price per unit of export.

Brander and Krugman (1983) use the term *dumping* to describe a situation where the FOB export price is lower than the price charged for domestic sales. Formally, in the present model,

$$p_k^{CIF} = p_k = \frac{a+\tau}{3} \quad \text{and} \quad p_k^{FOB} = p_k^{CIF} - \tau = \frac{a-2\tau}{3}. \tag{6.28}$$

Thus, each firm in each country "dumps" the product in the other country by "subsidizing" the transportation cost. Another commonly used definition of dumping is when a firm sells abroad at a price below cost. This does not happen in the present model.

Finally, note that for this problem, the Cournot market structure generates inefficient trade since the world could save the transportation cost if each firm sells only in its home country. However, in general, making each firm a monopoly in its own country would generate the other familiar inefficiencies.

6.6.2 Homogeneous products and preferential trade agreements among countries

There are three general types of trade agreements among countries: (1) the free-trade agreement (FTA), which is an agreement among countries to eliminate trade barriers among the member countries, but under which each country is free to set its own trade restrictions against trade with nonmember countries; (2) the customs union (CU), which is an agreement among countries to eliminate tariffs on goods imported from other member countries of the union and to set a uniform trade policy regarding nonmember countries; and (3) the common market (CM), where, in addition to the elimination of tariffs among member countries and in addition to the common tariff policy toward nonmembers, there is a free movement of factors of production among member countries.

Formal analyses of these agreements were first given by Viner, Meade and Vanek, and the interested reader is referred to surveys of literature

given in Corden 1984 and Vousden 1990, or in almost any elementary book on international trade.

Consider the following world. There are three countries: the European Community (EC), the Far East (FE) and Israel (IL). Assume that IL is a small country, thus it cannot affect the world prices. Only FE and EC produce carpets that are imported by IL. Assume that carpets cannot be produced in IL. We further assume that IL's demand for imported carpets is given by $p^{IL} = a - Q$, where Q denotes the quantity demanded and p^{IL} is the domestic tariff-inclusive price.

Assume that initially (period 0), IL sets a uniform tariff of $\$t$ per carpet irrespective of where the carpets are imported from. Then, in period 1 assume that IL signs a free-trade agreement (FTA) with EC.

Period 0: IL levies a uniform tariff on carpets

We denote by p_{EC} the price of a carpet charged by EC's producers, and by p_{FE} the price charged by FE's producers. Hence, with a uniform tariff of t, the price paid by IL's consumers for carpets imported from EC is $p_{EC}^{IL} = p_{EC} + t$, and the price paid for carpets imported from FE is $p_{FE}^{IL} = p_{FE} + t$. We make the following assumption:

ASSUMPTION 6.1 *The export price of carpets in EC exceeds the export price in FE. Formally,* $p_{EC} > p_{FE}$.

Figure 6.4 illustrates IL's demand for imported carpets and the prices (with and without the tariff) on carpets imported from EC and FE. Figure 6.4 shows that IL will import from the cheapest supplier, which is

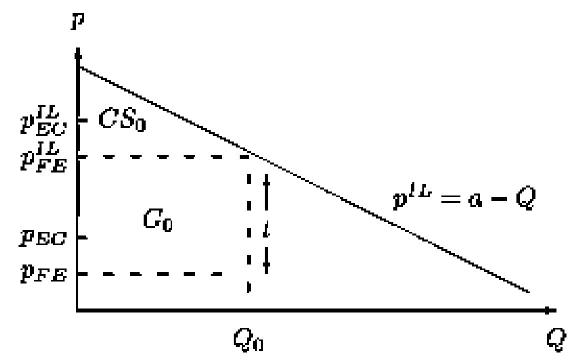

Figure 6.4: IL's import level under a uniform tariff

FE, so that the import level would be Q^0. In this case, the government's revenue from import-tariff collection would be $G^0 = tQ^0$. The IL's consumer surplus (see subsection 3.2.3 for a definition) is given by $CS^0 = (a - p^{IL}_{FE})Q^0/2$. Also, note that $Q^0 = a - p^{IL}_{FE} = a - p_{FE} - t$.

We define IL's social welfare as the sum of consumer surplus plus IL's government revenue from tariff collection. Note that in modeling international trade it is very important not to forget the existence of government's revenue and to assume that the government returns the tariff revenue to consumers in a lump-sum fashion or by other services. Hence,

$$W^0_{IS} \equiv CS^0 + G^0 = (a - p^{IL}_{FE} + 2t)Q^0/2 = (a - p_{FE} + t)Q^0/2$$

implying that

$$W^0_{IS} = \frac{(a - p_{FE} + t)(a - p_{FE} - t)}{2} = \frac{(a - p_{FE})^2 - t^2}{2}. \qquad (6.29)$$

Note that the last step in (6.29) uses the mathematical identity that $(\alpha + \beta)(\alpha - \beta) \equiv \alpha^2 - \beta^2$. Equation (6.29) shows that the welfare of country IL decreases with the tariff rate t and with FE's price of carpets.

Period 1: IL signs a free-trade agreement with the EC

Now suppose that IL signs a FTA with EC, so that the tariff on carpets imported from EC is now set to zero, whereas the tariff on imports from FE remains the same at the level of t per unit. Figure 6.5 illustrates that IL switches from importing from FE to importing from only EC for a price of $p^{IL}_{EC} = p_{EC}$. Given that the price of carpets drops in IL, the

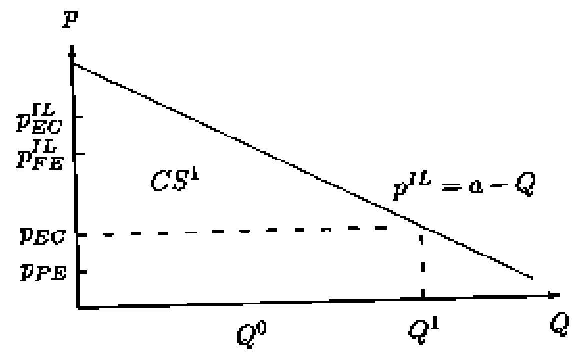

Figure 6.5: IL's import under the FTA

quantity of imported carpets increases to $Q^1 = a - p_{EC} > Q^0$. Notice that although IL's consumer price of carpets has decreased, IL now buys carpets from the more expensive source.

Under the FTA, since all the imports are from EC, the government collects zero revenue, that is $G^1 = 0$. Hence, IL's social welfare equals IL's consumer surplus. That is, $W^1 = CS^1$. The consumers' surplus is illustrated in Figure 6.5 and is calculated to be

$$W^1_{IL} = CS^1 = (a - p_{EC})Q^1/2 = (a - p_{EC})^2/2. \qquad (6.30)$$

Welfare analysis of the free-trade agreement

We now analyze whether IL gains from the FTA with EC. Comparing (6.29) and (6.30), we see that the FTA improves IL's welfare if $W^1 > W^0$. That is,

$$(a - p_{EC})^2 > (a - p_{FE})^2 - t^2$$

or,

$$t > \sqrt{(a - p_{FE})^2 - (a - p_{EC})^2}. \qquad (6.31)$$

Therefore,

Proposition 6.5 *A free-trade agreement between IL and EC is more likely to be welfare improving for IL when (a) the initial uniform tariff is high, and (b) when the difference in prices between the two foreign exporters is small; that is, when p_{EC} is close to p_{FE}.*

We conclude this analysis with a graphic illustration of the gains and loss from the FTA. Figure 6.6 illustrates the welfare implication of IL's signing the FTA with EC. In Figure 6.6, the area denoted by ϕ

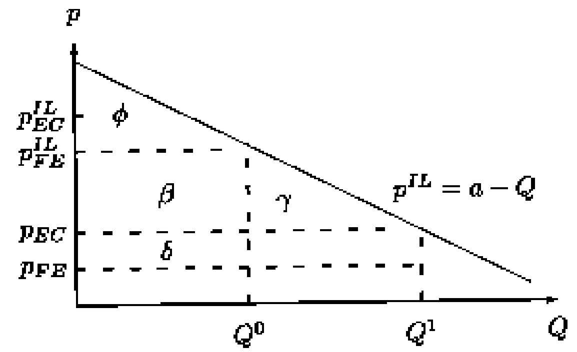

Figure 6.6: The welfare effects of the free-trade agreement

measures IL's consumer surplus prior to signing the FTA. The sum of the areas $\beta + \delta$ measures IL's government tariff revenue prior to signing the agreement. Hence, IL's welfare prior to signing the agreement is $W^0 = \phi + \beta + \delta$.

In Figure 6.6, the sum of the areas $\phi+\beta+\gamma$ measures IL's consumer surplus after the FTA is signed. Since there are no tariff revenues after the FTA (all carpets are imported from the EC), the welfare of IL after the FTA is $W^1 = \phi+\beta+\gamma$.

Altogether, the welfare change resulting from signing the FTA is given by $\Delta W \equiv W^1 - W^0 = \gamma - \delta$.

DEFINITION 6.5 *The change in consumer surplus due to the increase in the consumption of the imported good (area γ in Figure 6.6) is called the* **trade-creation effect** *of the FTA. The change in the importing country's expenditure due to the switch to importing from the more expensive country (area δ in Figure 6.6) is called the* **trade-diversion effect** *of the FTA.*

Thus, the importing country gains from the FTA if the (positive) trade-creation effect associated with the increase in the import level dominates the (negative) trade-diversion effect associated to switching to importing from the more expensive source.

6.7 Appendix: Cournot Market Structure with Heterogeneous Firms

In this appendix we extend the analysis conducted in Subsection 6.1.2, and solve for the Cournot-market-structure equilibrium when there is a large number of firms with different cost functions. Following Bergstrom and Varian (1985), we introduce a method for calculating a Cournot-Nash equilibrium output level without resorting to solving N first-order conditions for the equilibrium N output levels.

In a Cournot market structure with N firms, each with a unit cost of $c_i \geq 0$, $i = 1, \ldots, N$, each firm i chooses its output q_i that solves

$$\max_{q_i} \pi_i(q_i, q^c_{-i}) = \left[a - bq_i - b\left(\sum_{j\neq i} q^c_j\right)\right] q_i - c_i q_i$$

yielding, assuming $q^c_i > 0$ for all i, a first-order condition

$$a - 2bq^c_i - b\left(\sum_{j\neq i} q^c_j\right) = c_i, \quad i = 1, \ldots, N.$$

Now, instead of solving N equations (N first-order conditions) for N output levels, we solve for the aggregate production level by rewriting the first-order conditions in the form of:

$$a - bq^c_i - bQ^c = c_i, \quad i = 1, \ldots, N.$$

Summing over all q_i, $i = 1, \ldots, N$ yields

$$Na - bQ^c - bNQ^c = \sum_{i=1}^{N} c_i.$$

Hence, the Cournot equilibrium aggregate industry output and market price are given by

$$Q^c = \frac{Na}{(N+1)b} - \frac{\sum_{i=1}^N c_i}{(N+1)b} \quad \text{and} \quad p^c = \frac{a}{N+1} + \frac{\sum_{i=1}^N c_i}{N+1}. \tag{6.32}$$

Hence,

Proposition 6.6 *In an industry where firms have constant unit costs, if in a Cournot equilibrium all firms produce strictly positive output levels, then the Cournot aggregate industry equilibrium output and price levels depend only on the sum of the firms' unit costs and not on the distribution of unit costs among the firms.*

The result stated in Proposition 6.6 is important, since it implies that under constant unit costs, industry output, price, and hence, total welfare can be calculated by using the sum of firms' unit costs, without investigating the precise cost distribution among firms. Moreover, the proof of Proposition 6.6 does not rely on linear demand and therefore also applies to nonlinear demand functions.

We conclude this appendix by illustrating a simple application of Proposition 6.6. Consider an industry consisting of two type of firms: high-cost and low-cost firms. Suppose that there are $H \geq 1$ high-cost firms with a unit production cost given by c_H, and $L \geq 1$ low-cost firms with a unit production cost given by c_L, where $c_H \geq c_L \geq 0$. Substituting into (6.32) yields

$$Q^c = \frac{(H+L)a}{(H+L+1)b} - \frac{Hc_H + Lc_L}{(H+L+1)b} \quad \text{and} \quad p^c = \frac{a}{H+L+1} + \frac{Hc_H + Lc_L}{H+L+1}. \tag{6.33}$$

Hence, the Cournot output and price equilibrium levels depend only on $Hc_H + Lc_L$. The advantage of learning this method for calculating Cournot equilibrium outcomes becomes clear in the case where there is an entry (or exit) of some firms. For example, suppose we observe that three additional low-cost firms have joined the industry. Then, the new Cournot equilibrium industry output and price can be immediately calculated by replacing $Hc_H + Lc_L$ with $Hc_H + (L+3)c_L$ in (6.33).

6.8 Exercises

1. Two firms produce a homogeneous product. Let p denote the product's price. The output level of firm 1 is denoted by q_1, and the output level of firm 2 by q_2. The aggregate industry output is denoted by Q, $Q \equiv q_1 + q_2$. The aggregate industry demand curve for this product is given by $p = \alpha - Q$.

   Assume that the unit cost of firm 1 is c_1 and the unit cost of firm 2 is c_2, where $\alpha > c_2 > c_1 > 0$. Perform the following:

   (a) Solve for a competitive equilibrium (see Definition 4.2 on page 65). Make sure that you solve for the output level of each firm and the market price.

   (b) Solve for a Cournot equilibrium (see Definition 6.1 on page 99). Make sure that you solve for the output level of each firm and the market price.

   (c) Solve for a sequential-moves equilibrium (see Section 6.2 on page 104) assuming that firm 1 sets its output level before firm 2 does.

   (d) Solve for a sequential-moves equilibrium, assuming that firm 2 sets its output level before firm 1 does. Is there any difference in market shares and the price level between the present case and the case where firm 1 moves first? Explain!

   (e) Solve for a Bertrand equilibrium (see Definition 6.2 on page 108). Make sure that you solve for the output level of each firm and the market price.

2. In an industry there are N firms producing a homogeneous product. Let q_i denote the output level of firm i, $i = 1, 2, \ldots, N$, and let Q denote the aggregate industry production level. That is, $Q \equiv \sum_{i=1}^{N} q_i$. Assume that the demand curve facing the industry is $p = 100 - Q$. Suppose that the cost function of each firm i is given by

$$TC_i(q_i) \equiv \begin{cases} F + (q_i)^2 & \text{if } q_i > 0 \\ 0 & \text{if } q_i = 0. \end{cases}$$

   Solve the following problems:

   (a) Suppose that the number of firms in the industry N is sufficiently small so that all the N firms make above-normal profits. Calculate the output and profit levels of each firm in a Cournot equilibrium.

   (b) Now, assume that firms are allowed to enter to or exit from the industry. Find the equilibrium number of firms in the industry as a function of F. *Hint:* Equate a firm's profit level that you found earlier to zero and solve for N.

3. Consider a three-period version of the sequential-moves equilibrium analyzed in section 6.2. Assume that the market inverse demand curve is given by $p = 120 - Q$, and suppose that there are three firms that set their output levels sequentially: firm 1 sets q_1 in period 1, firm 2 sets q_2 in period 2, and firm 3 sets q_3 in period 3. Then, firms sell their output and collect their profits. Solve for the sequential-moves equilibrium (assuming that production is costless). Make sure that you solve for the output level of each firm, and the market price.

4. Two firms compete in prices in a market for a homogeneous product. In this market there are $N > 0$ consumers; each buys one unit if the price of the product does not exceed \$10, and nothing otherwise. Consumers buy from the firm selling at a lower price. In case both firms charge the same price, assume that $N/2$ consumers buy from each firm. Assume zero production cost for both firms.

   (a) Find the Bertrand equilibrium prices for a single-shot game, assuming that the firms choose their prices simultaneously.

   (b) Now suppose that the game is repeated infinitely. Let ρ denote the time-discount parameter. Propose trigger price strategies for both firms yielding the collusive prices of (10, 10) each period. Calculate the minimal value of ρ that would enforce the trigger price strategies you proposed.

   (c) Now suppose that the unit production cost of firm 2 is \$4, but the unit cost of firm 1 remained zero. Find the Bertrand equilibrium prices for the single-shot game.

   (d) Assuming the new cost structure, propose trigger price strategies for both firms yielding the collusive prices of (10, 10) each period, and calculate the minimal value of ρ that would enforce the trigger price strategies you propose.

   (e) Conclude whether it is easier for firms to enforce the collusive prices when there is symmetric industry cost structure, or when the firms have different cost structures. Explain!

5. Consider the free-trade agreement model analyzed in subsection 6.6.2. Suppose that the world consists of three countries denoted by A, B, and C. Country A imports shoes from countries B and C and does not have local production of shoes. Let the export shoe prices of countries B and C be given by $p_B = \$60$ and $p_C = \$40$. Also, suppose that initially, country A levies a uniform import tariff of $t = \$10$ per each pair of imported shoes. Answer the following questions:

   (a) Suppose that country A signs a FTA with country B. Does country A gain or lose from this agreement? Explain!

   (b) Suppose now that initially, the export price of shoes in country C is $p_C = \$50.01$. Under this condition, will country A gain or lose from the FTA? Explain!

6. In a market for luxury cars there are two firms competing in prices. Each firm can choose to set a high price given by p_H, or a low price

given by p_L, where $p_H > p_L \geq 0$. The profit levels of the two firms as a function of the prices chosen by both firms is given in Table 6.2. The rules of this two-stage market game are as follows: In the first

		Firm 2 p_H		Firm 2 p_L	
Firm 1	p_H	100	100	0	120
	p_L	120	0	70	70

Table 6.2: Meet the competition clause

stage firm 1 sets its price $p_1 \in \{p_H, p_L\}$. In the second stage firm 1 cannot reverse its decision, whereas firm 2 observes p_1 and then chooses $p_2 \in \{p_H, p_L\}$. Then, the game ends and each firm collects its profit according to Table 6.2.

(a) Formulate the game in extensive form (Definition 2.7 on page 24) by drawing the game tree, and solve for the subgame perfect equilibrium (Definition 2.10 on page 27) for this game.

(b) Suppose now that firm 1 offers its consumers to match its price with the lowest price in the market (the so-called meet the competition clause). Solve for the subgame perfect equilibrium for the modified game. *Hint:* Modify the game to three stages, allowing firm 1 to make a move in the third stage only in the case where it chose p_H in the first stage and firm 2 chose p_L in the second stage.

7. This problem is directed to highly advanced students only: Suppose there are $N > 2$ firms that set their output sequentially, as described in section 6.2. Suppose that all firms have identical unit costs given by c, and suppose that the market inverse demand curve facing this industry is given by $p = a - Q$, where $a > c \geq 0$ and $Q \equiv \sum_{i=1}^{N} q_i$.

(a) Solve for the sequential-moves equilibrium by showing that the output level of the firm that moves in period i, $i = 1, \ldots, N$ is given by

$$q_i^s = \frac{a-c}{2^i}.$$

(b) Show that the aggregate equilibrium-output level is given by

$$Q^s \equiv \sum_{i=1}^{N} q_i = \left[1 - \frac{1}{2^N}\right](a-c).$$

(c) Conclude what happens to the aggregate industry-output level when the number of firms (and periods) increases with no bounds, (i.e., when $N \to \infty$).

6.9 References

Abreu, D. 1986. "Extremal Equilibria of Oligopolistic Supergames." *Journal of Economic Theory* 39: 191–225.

Bergstrom, T., and H. Varian. 1985. "When Are Nash Equilibria Independent of the Distribution of Agents' Characteristics?" *Review of Economics Studies* 52: 715–718.

Bertrand, J. 1883. Reviews of *Théories Mathematique de la Richesse Sociale*, by Léon Walras; and of *Recherches sur les Principles Mathematiques de la Théorie des Richesses*, by Augustin Cournot. *Journal des Savants* 67: 499–508.

Bork, R. 1978. *The Antitrust Paradox.* New York: Basic Books.

Brander, J., and P. Krugman. 1983. "A 'Reciprocal Dumping' Model of International Trade." *Journal of International Economics* 15: 313–321.

Corden, M. 1984. "The Normative Theory of International Trade." In *Handbook of International Economics,* edited by R. Jones and P. Kenen. Amsterdam: North-Holland.

Cournot, A. 1929 [1838]. *Researches into the Mathematical Principles of the Theory of Wealth.* Translated by Nathaniel Bacon. New York: Macmillan.

Edgeworth, F. 1925 [1897]. "The Pure Theory of Monopoly." In *Papers Relating to Political Economy,* edited by F. Edgeworth. London: Macmillan.

Friedman, J. 1971. "A Noncooperative Equilibrium for Supergames." *Review of Economic Studies* 38: 1–12.

Friedman, J. 1977. *Oligopoly and the Theory of Games.* Amsterdam: North-Holland.

Green, E., and R. Porter. 1984. "Non-cooperative Collusion Under Imperfect Price Information." *Econometrica,* 52: 87–100

Konow, J. 1994. "The Political Economy of Heinrich von Stackelberg." *Economic Inquiry* 32: 146–165.

Kreps, D., and J. Scheinkman. 1983. "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes." *Bell Journal of Economics* 14: 326–337.

Lambson, V. 1984. "Self-Enforcing Collusion in Large Dynamic Markets." *Journal of Economic Theory* 34: 282–291.

Rotemberg, J., and G. Saloner. 1986. "Supergame-Theoretic Model of Business Cycles and Price Wars During Booms." *American Economic Review* 76: 390–407.

Segerstrom, P. 1988. "Demons and Repentance." *Journal of Economic Theory* 45: 32–52.

Shubik, M. 1987. "Cournot, Antoine Augustin." In *The New Palgrave Dictionary of Economics,* edited by J. Eatwell, M. Milgate, and P. Newman. New York: The Stockton Press.

von Stackelberg, H. 1934. *Marktform und Gleichgewicht (Market structure and equilibrium)*. Vienna: Springer-Verlag.

Tirole, J. 1988. *The Theory of Industrial Organization*. Cambridge, Mass.: MIT Press.

Vousden, N. 1990. *The Economics of Trade Protection*. Cambridge: Cambridge University Press.

Chapter 7

Markets for Differentiated Products

> You can have it any color you want as long as it's black.
> —Attributed to Henry Ford

In this chapter we analyze oligopolies producing differentiated products. Where in chapter 6 consumers could not recognize or did not bother to learn the producers' names or logos of homogeneous products, here, consumers are able to distinguish among the different producers and to treat the products (brands) as close but imperfect substitutes.

Several important observations make the analysis of differentiated products highly important.

1. Most industries produce a large number of similar but not identical products.

2. Only a small subset of all possible varieties of differentiated products are actually produced. For example, most products are not available in all colors.

3. Most industries producing differentiated products are concentrated, in the sense that it is typical to have two to five firms in an industry.

4. Consumers purchase a small subset of the available product varieties.

This chapter introduces the reader to several approaches to modeling industries producing differentiated products to explain one or more of these observations.

Product differentiation models are divided into two groups: non-address models, and address (location) models. Figure 7.1 illustrates the logical connections among the various approaches. The non-address

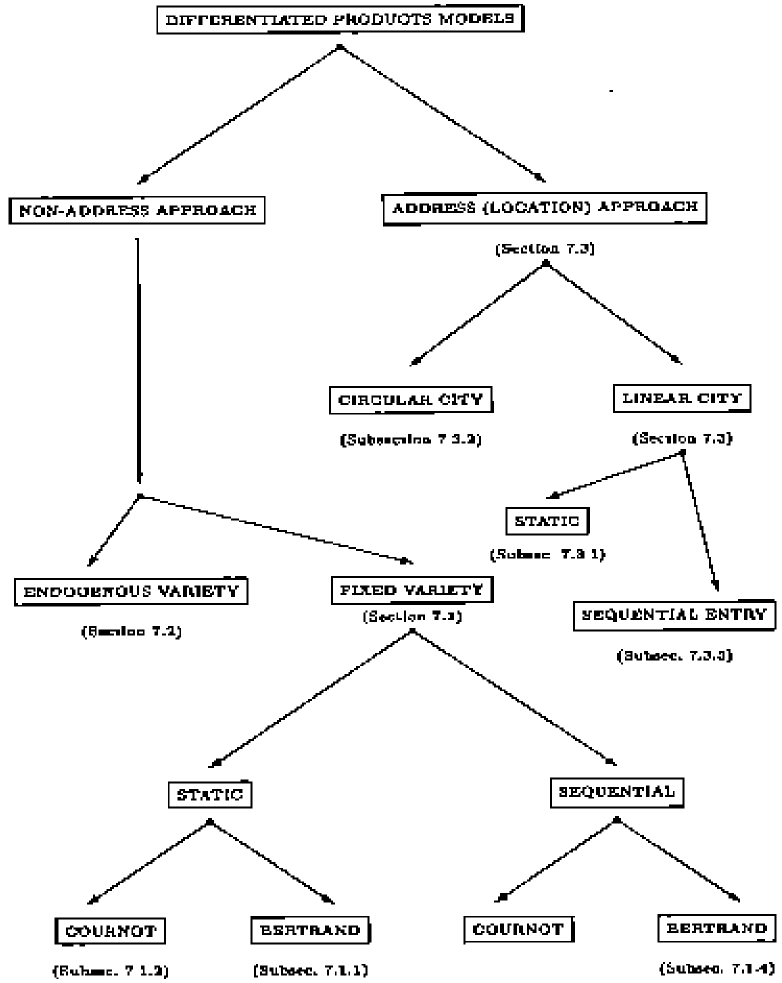

Figure 7.1: Approaches to modeling differentiated-products industries

approach, displayed on the left main branch of Figure 7.1, is divided into two categories: a fixed number of differentiated brands models, and endogenously determined variety models. The fixed number of brands approach is analyzed in section 7.1 (Simple Models for Differentiated

Products), where we analyze and compare quantity and price competition between the two differentiated-brands producers. Basic definitions for the degrees of product differentiation are provided and utilized in the two types of market structures. Section 7.2 (Monopolistic Competition) analyzes a general equilibrium environment where free entry is allowed, so the number of brands in an industry is determined in the model itself. We assume that the economy is represented by a single consumer whose preferences exhibit love for variety of differentiated brands, and that firms' technologies exhibit returns to scale together with fixed cost of production. Assuming free entry of firms enables us to compute the equilibrium variety of differentiated brands. The monopolistic competition approach proves to be extremely useful in analyzing international markets, which is discussed in subsection 7.2.2.

The address (location) approach, displayed on the right main branch of Figure 7.1, is analyzed in section 7.3 (Location Models). This approach provides an alternative method for modeling product differentiation by introducing location, or addresses, into consumers' preferences that measure how close the brands actually produced are to the consumers' ideal brands. This approach is useful to model heterogeneous consumers who have different tastes for the different brands.

Together, sections 7.2 and 7.3 discuss the two major approaches to product differentiation: the non-address approach and the address approach, respectively (see a discussion in Eaton and Lipsey 1989). The major difference between the approaches is that in the non-address approach all consumers gain utility from consuming a variety of products and therefore buy a variety of brands (such as a variety of music records, of movies, of software, of food, etc.). In contrast, the address (location) approach, each consumer buys only one brand (such as one computer, one car, or one house), but consumers have different preferences for their most preferred brand. A third approach to product differentiation, not discussed in this chapter, is found in Lancaster 1971. Lancaster's "characteristics" approach assumes that each product consists of many characteristics (such as color, durability, safety, strength); in choosing a specific brand, the consumer looks for the brand that would yield the most suitable combinations of the product's characteristics. Finally, a reader interested in applications of product differentiation to the ready-to-eat cereals industry is referred to Scherer 1979 and Schmalensee 1978.

7.1 Simple Models for Two Differentiated Products

Consider a two-firm industry producing two differentiated products indexed by $i = 1, 2$. To simplify the exposition, we assume that production is costless. Following Dixit (1979) and Singh and Vives (1984), we assume the following (inverse) demand structure for the two products:

$$p_1 = \alpha - \beta q_1 - \gamma q_2 \quad \text{and} \quad p_2 = \alpha - \gamma q_1 - \beta q_2, \quad \text{where} \quad \beta > 0,\ \beta^2 > \gamma^2. \tag{7.1}$$

Thus, we assume that that there is a fixed number of two brands and that each is produced by a different firm facing an inverse demand curve given in (7.1). The assumption of $\beta^2 > \gamma^2$ is very important since it implies that the effect of increasing q_1 on p_1 is larger than the effect of the same increase in q_2. That is, the price of a brand is more sensitive to a change in the quantity of this brand than to a change in the quantity of the competing brand. A common terminology used to describe this assumption is to say that the *own-price effect* dominates the *cross-price effect*.

The demand structure exhibited in (7.1) is formulated as a system of inverse demand functions where prices are functions of quantity purchased. In order to find the direct demand functions, (quantity demanded as functions of brands' prices) we need to invert the system given in (7.1). The appendix (section 7.4) shows that

$$q_1 = a - bp_1 + cp_2 \quad \text{and} \quad q_2 = a + cp_1 - bp_2, \quad \text{where} \tag{7.2}$$

$$a \equiv \frac{\alpha(\beta-\gamma)}{\beta^2-\gamma^2}, \quad b \equiv \frac{\beta}{\beta^2-\gamma^2} > 0, \quad c \equiv \frac{\gamma}{\beta^2-\gamma^2}.$$

How to measure the degree of brand differentiation

We would now like to define a measure for the degree of product differentiation.

DEFINITION 7.1 *The brands'* **measure of differentiation,** *denoted by* δ, *is*

$$\delta \equiv \frac{\gamma^2}{\beta^2}.$$

1. *The brands are said to be* **highly differentiated** *if consumers find the products to be very different, so a change in the price of brand j will have a small or negligible effect on the demand for brand i. Formally, brands are highly differentiated if δ is close to 0. That is, when $\gamma^2 \to 0$, (hence $c \to 0$).*

2. *The brands are said to be* **almost homogeneous** *if the cross-price effect is close or equal to the own-price effect. In this case, prices of all brands will have strong effects on the demand for each brand, more precisely, if an increase in the price brand j will increase the demand for brand i by the same magnitude as a decrease in the price of brand i, that is, when δ is close to 1, or equivalently when $\gamma^2 \to \beta^2$, (hence $c \to b$).*

Figure 7.2 illustrates the relationships between the the demand parameters β and γ as described in Definition 7.1. In Figure 7.2 a hori-

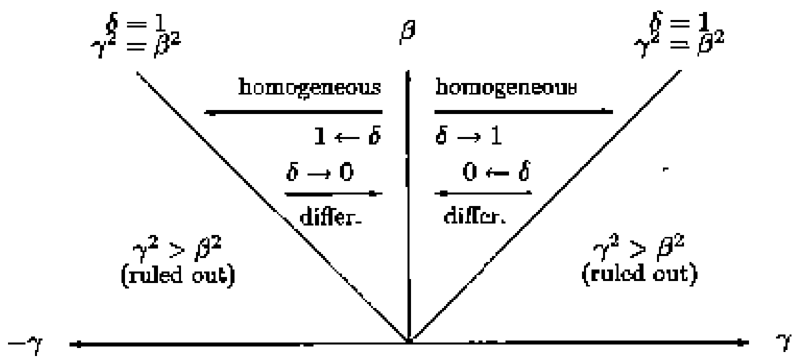

Figure 7.2: Measuring the degree of product differentiation

zontal movement toward the diagonals implies that the products are becoming more homogeneous, $(\gamma^2 \to \beta^2)$. In contrast, a movement toward the center is associated with the products becoming more differentiated, $(\gamma \to 0)$.

7.1.1 Quantity game with differentiated products

We now solve for the prices and quantity produced under the Cournot market structure, where firms choose quantity produced as actions. Just as we did in solving a Cournot equilibrium for the homogeneous products case, we look for a Nash equilibrium in firms' output levels, as defined in Definition 6.1 on page 99.

Assuming zero production cost, using the inverse demand functions given in (7.1), we note that each firm i takes q_j as given and chooses q_i to

$$\max_{q_i} \pi_i(q_1, q_2) = (\alpha - \beta q_i - \gamma q_j)q_i \quad i,j = 1,2,\ i \neq j. \tag{7.3}$$

The first-order conditions are given by $0 = \frac{\partial \pi_i}{\partial q_i} = \alpha - 2\beta q_i - \gamma q_j$, yielding best response functions given by

$$q_i = R_i(q_j) = \frac{\alpha - \gamma q_j}{2\beta} \quad i,j = 1,2,\ i \neq j. \tag{7.4}$$

Figure 7.3 illustrates the best-response functions in the $(q_1 - q_2)$ space. Notice that these functions are similar to the ones obtained for the Cournot game with homogeneous products illustrated in Figure 6.1. Notice that as $\gamma \nearrow \beta$ (the products are more homogeneous), the best-

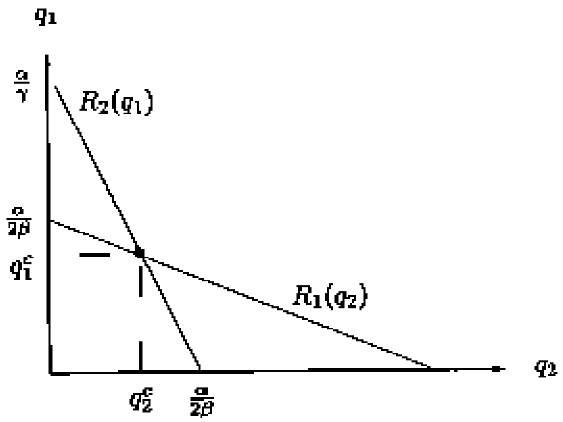

Figure 7.3: Best-response functions for quantity competition in differentiated products

response function becomes steeper, thereby making the profit-maximizing output level of firm i more sensitive to changes in the output level of firm j (due to stiffer competition). In contrast, as $\gamma \searrow 0$, the best-response function becomes constant (zero sloped), since the products become completely differentiated.

Solving the best-response functions (7.4), using symmetry, we have that

$$q_i^c = \frac{\alpha}{2\beta + \gamma}, \quad p_i^c = \frac{\alpha\beta}{2\beta + \gamma}, \quad \pi_i^c = \frac{\alpha^2\beta}{(2\beta + \gamma)^2} \quad i = 1,2. \tag{7.5}$$

Clearly, as γ increases (the products are less differentiated), the individual and aggregate quantity produced, the prices, and the profits all decline. Hence,

Proposition 7.1 *In a Cournot game with differentiated products, the profits of firms increase when the products become more differentiated.*

The importance of Proposition 7.1 is that it can explain why firms tend to spend large sums of money to advertise their brands: because firms would like the consumers to believe that the brands are highly differentiated from the competing brands for the purpose of increasing their profits. In other words, differentiation increases the monopoly power of brand-producing firms.

7.1.2 Price game with differentiated products

We now solve for the prices and quantity produced under the Bertrand market structure, where firms choose prices as their actions. Just as we did in solving for a Bertrand equilibrium for the homogeneous products case, we look for a Nash equilibrium in firms' prices, as defined in Definition 6.2 on page 108 for the homogeneous product case.

Using the direct demand functions given in (7.2), each firm i takes p_j as given and chooses p_i to

$$\max_{p_i} \pi_i(p_1, p_2) = (a - bp_i + cp_j)p_i \quad i,j = 1,2,\ i \neq j. \tag{7.6}$$

The first-order conditions are given by $0 = \frac{\partial \pi_i}{\partial p_i} = a - 2bp_i + cp_j$, yielding best-response functions given by

$$p_i = R_i(p_j) = \frac{a + cp_j}{2b} \quad i,j = 1,2,\ i \neq j. \tag{7.7}$$

The best-response functions are drawn in Figure 7.4. You have probably

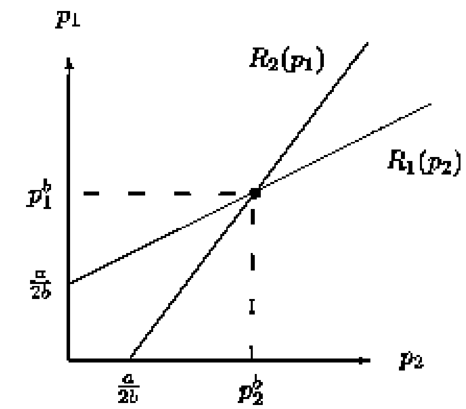

Figure 7.4: Best-response functions for price competition in differentiated products

noticed that there is something different in Figure 7.4 compared with what is in Figure 7.3: In price games, the best-response functions are upward sloping, meaning that if one firm raises its price, the other would respond by raising its price as well. Well, this "discovery" deserves a definition (Bulow, Geanakoplos, and Klemperer 1985):

DEFINITION 7.2

1. *Players' strategies are said to be* **strategic substitutes** *if the best-response functions are downward sloping.*
2. *Players' strategies are said to be* **strategic complements** *if the best-response functions are upward sloping.*

Note that this terminology may be misleading, since there is no relationship between this definition and whether goods are substitutes or complements in consumption. Definition 7.2 implies that in a quantity game the quantities are strategic substitutes, whereas in a price game prices are strategic complements.

Solving (7.7) yields for $i = 1, 2$,

$$p_i^b = \frac{a}{2b-c} = \frac{\alpha(\beta-\gamma)}{2\beta-\gamma},\ q_i^b = \frac{ab}{2b-c},\ \pi_i^b = \frac{a^2 b}{(2b-c)^2} = \frac{\alpha^2\beta(\beta-\gamma)}{(2\beta-\gamma)^2(\beta+\gamma)} \tag{7.8}$$

The profit levels decline when the products become less differentiated (γ increases). In the limit, when $\gamma = \beta$, the products become homogeneous, and the profits drop to zero as in the Bertrand equilibrium for homogeneous products analyzed in section 6.2. Hence,

Proposition 7.2 *In a Bertrand game with differentiated products, the profits of firms increase when the products become more differentiated.*

As with the Cournot case, product differentiation increases the monopoly power of brand-producing firms by loosening up price competition among the brand-producing firms.

7.1.3 Cournot versus Bertrand in differentiated products

Which market structure, a Cournot or a Bertrand, would yield a higher market price? How would changing the degree of product differentiation affect the relative difference between the two market-structure outcomes? As you may expect, the price under Bertrand is indeed lower than it is under the Cournot market structure. Formally, comparing (7.5) with (7.8) yields

$$p_i^c - p_i^b = \frac{\alpha\beta}{2\beta+\gamma} - \frac{a}{2b-c} = \frac{\alpha\beta}{2\beta+\gamma} - \frac{\alpha(\beta-\gamma)}{2\beta-\gamma} = \frac{\alpha}{4\frac{\beta^2}{\gamma^2}-1}. \tag{7.9}$$

Thus,

Proposition 7.3 *In a differentiated products industry:*

1. *The market price under Cournot is higher than it is under Bertrand. Formally $p_i^c > p_i^b$.*
2. *The more differentiated the products are, the smaller the difference between the Cournot and Bertrand prices. Formally, $\frac{\partial[p_i^c - p_i^b]}{\partial\gamma} > 0$.*
3. *This difference in prices is zero when the products become independent. Formally, $\lim_{\gamma\to 0}[p_i^c - p_i^b] = 0$.*

The intuition behind Proposition 7.3, given in Vives 1985, is as follows: Under Cournot market structure each firm expects the other firm to hold its output level constant. Hence, each firm would maintain a low output level since it is aware that a unilateral output expansion would result in a drop in the market price. In contrast, under the Bertrand market structure each firm assumes that the rival firm holds its price constant, hence output expansion will not result in a price reduction. Therefore, more output is produced under the Bertrand market structure than under the Cournot market structure. Cheng 1985 provides some additional graphical intuition for the differences between the market outcomes obtained under the two market structures.

7.1.4 Sequential-moves price game

Consider a two-period, price-setting sequential game that is similar to the sequential-moves quantity game described in section 6.2; but here, we let firms set prices rather than quantity produced. In order to have some fun, let us take a specific numerical example for the demand system given in (7.2):

$$q_1 = 168 - 2p_1 + p_2 \quad \text{and} \quad q_2 = 168 + p_1 - 2p_2. \tag{7.10}$$

For this particular example, (7.8) implies that the single-period game Bertrand prices and profit levels are $p_i^b = 56$ and $\pi_i^b = 6272$.

Following the same logical steps as those in section 6.2, we look for a SPE in prices where firm 1 sets its price before firm 2. Thus, in the first period, firm 1 takes firm 2's best-response function (7.7) as given, and chooses p_1 that solves

$$\max_{p_1} \pi_1(p_1, R_2(p_1)) = \left(168 - 2p_1 + \frac{168 + p_1}{4}\right) p_1. \tag{7.11}$$

The first-order condition is $0 = \frac{\partial \pi_1}{\partial p_1} = 210 - \frac{7}{2}p_1$. Therefore, $p_1^s = 60$, hence, $p_2^s = 57$. Substituting into (7.10) yields that $q_1^s = 105$ and $q_2 = 114$. Hence, $\pi_1^s = 60 \times 105 = 6300 > \pi_1^b$, and $\pi_2^s = 57 \times 114 = 6498 > \pi_2^b$.

Why do we bother to go over this exercise under a price game? Well, the following proposition yields a rather surprising result concerning the relationship between firms' profit levels and the order of moves.

Proposition 7.4 *Under a sequential-moves price game (or more generally, under any game where actions are strategically complements):*

1. *Both firms collect a higher profit under a sequential-moves game than under the single-period Bertrand game. Formally, $\pi_i^s > \pi_i^b$ for $i = 1, 2$.*
2. *The firm that sets its price first (the leader) makes a lower profit than the firm that sets its price second (the follower).*
3. *Compared to the Bertrand profit levels, the increase in profit to the first mover (the leader) is smaller than the increase in profit to the second mover (the follower). Formally, $\pi_1^s - \pi_1^b < \pi_2^s - \pi_2^b$.*

It this amazing? What we have learned from this example is that being the first to move is not always an advantage. Here, each firm would want the other firm to make the first move. The intuition behind this result is as follows. When firm 1 sets its price in period 1, it calculates that firm 2 will slightly undercut p_1 in order to obtain a larger market share than firm 1. This calculation puts pressure on firm 1 to maintain a high price to avoid having firm 2 set a very low market price. Hence, both firms set prices above the static Bertrand price levels. Now, firm 1 always makes a lower profit than firm 2, since firm 2 slightly undercuts firm 1 and captures a larger market share.

Finally, note that we could have predicted that the profit of firm 1 will increase beyond the static Bertrand profit level even without resorting to the precise calculations. Using a *revealed profitability* argument, we can see clearly that firm 1 can always set $p_1 = p_1^b$ and make the same profit as under the static Bertrand game. However, given that firm 1 chooses a different price, its profit can only increase.

Finally, part 1 of Proposition 7.4 reveals the major difference between the price sequential-moves game and the quantity sequential-moves game analyzed in section 6.2. Here, the profit of firm 2 (the follower's) is higher under the sequential-moves price game than its profit under the static Bertrand game. In contrast, under the sequential-moves quantity game the followers' profit is lower than it is under the static Cournot game.

7.2 Monopolistic Competition in Differentiated Products

In this section, we analyze a monopolistic-competition environment (Chamberlin 1933). Our major goal is to calculate the equilibrium number of differentiated brands produced by the industry.

The main features of this environment are that: (1) consumers are homogeneous (have identical preferences) or can be represented by a single consumer who loves to consume a variety of brands. Thus, this model better describes markets in which consumers like to consume a large variety of brands—such as a variety of music records, of video, of clothes, and of movies—rather than markets for cars where most individuals consume, at most, one unit; (2) there is an unlimited number of potentially produced brands; and (3) free entry of new brand-producing firms.

It should be pointed out that this model is a general equilibrium one. Unlike the partial equilibrium models, the general equilibrium model is one where consumers' demand is derived from a utility maximization where the consumers' income is generated from selling labor to firms and from owning the firms. Subsection 7.2.1 analyzes a single-economy monopolistic competition, and subsection 7.2.2 extends the model to two open economies.

7.2.1 The basic model

We analyze here a simplified version of Dixit and Stiglitz 1977. Consider an industry producing differentiated brands indexed by $i = 1, 2, 3, \ldots, N$, where N is an endogenously determined number of produced brands. We denote by $q_i \geq 0$ the quantity produced/consumed of brand i, and by p_i the price of one unit of brand i.

Consumers

In this economy, there is a single (representative) consumer whose preferences exhibit the love-for-variety property. Formally, the utility function of the representative consumer is given by a constant-elasticity-of-substitution (CES) utility function:

$$u(q_1, q_2, \ldots, q_N) \equiv \sum_{i=1}^{N} \sqrt{q_i}. \tag{7.12}$$

This type of utility function exhibits love for variety since the marginal utility of each brand at a zero consumption level is infinite. That is, $\lim_{q_i \to 0} \frac{\partial u}{\partial q_i} = \lim_{q_i \to 0} \frac{1}{2\sqrt{q_i}} = +\infty$.

In addition, Figure 7.5 illustrates that the indifference curves are convex to the origin, indicating that the consumers like to mix the brands in their consumption bundle. Also, note that the indifference curves touch

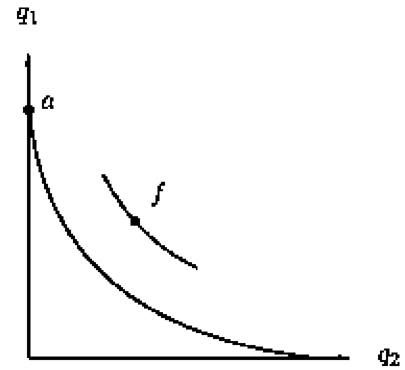

Figure 7.5: CES indifference curves for $N = 2$

the axes, therefore making it possible for the consumers to gain utility even when some brands are not produced (hence not consumed). We use the word *representative* consumer for this utility function since, in reality, individual consumers do not purchase the entire variety of products. Sattinger (1984) proposed a method for aggregating individuals who purchase a single brand into aggregate market demand facing all the brand-producing firms.

Finally, the consumer's income (denoted by I) is composed of the total wages paid by the producing firms plus the sum of their profits (if any). We denote by $\pi_i(q_i)$ the profit of the firm producing brand i. We also normalize the wage rate to equal 1, so all "monetary" values (p_i, I, and π_i) are all denominated in units of labor. Hence, the consumers maximize their utility (7.12) subject to a budget constraint given by

$$\sum_{i=1}^{N} p_i q_i \leq I \equiv L + \sum_{i=1}^{N} \pi_i(q_i). \qquad (7.13)$$

We form the Lagrangian

$$\mathcal{L}(q_i, p_i, \lambda) \equiv \sum_{i=1}^{N} \sqrt{q_i} + \lambda \left[I - \sum_{i=1}^{N} p_i q_i \right].$$

The first-order condition for every brand i is

$$0 = \frac{\partial \mathcal{L}}{\partial q_i} = \frac{1}{2\sqrt{q_i}} - \lambda p_i, \quad \text{for } i = 1, 2, \ldots, N.$$

Therefore, the direct demand, inverse demand, and the resulting price elasticity (η_i) for each brand i, are given by

$$q_i(p_i) = \frac{1}{4\lambda^2 (p_i)^2}, \quad p_i(q_i) = \frac{1}{2\lambda\sqrt{q_i}} \quad \text{and} \quad \eta \equiv \frac{\partial q_i}{\partial p_i}\frac{p_i}{q_i} = -2. \quad (7.14)$$

Finally, note that we assumed that λ is a constant. However, λ is not really a constant but a function of all prices and N. This procedure would be right had we assumed a continuum of brands indexed on the interval $[0, \infty)$. In this case, a rise in the price of a single brand would not have an effect on consumers' expenditure and hence on λ. The continuum version of (7.12) should be written as $u = \int_0^\infty \sqrt{q(i)}di$. However, in an attempt to avoid using integrals in this book, we provide the present approach as a good approximation for the continuous case.

Brand-producing firms

Each brand is produced by a single firm. All (potential) firms have identical technologies (identical cost structure) with increasing returns to scale (IRS) technologies. Formally, the total cost of a firm producing q_i units of brand i is given by

$$TC_i(q_i) = \begin{cases} F + cq_i & \text{if } q_i > 0 \\ 0 & \text{if } q_i = 0. \end{cases} \qquad (7.15)$$

Defining a monopolistic-competition market structure

DEFINITION 7.3 *The triplet* $\{N^{mc}, p_i^{mc}, q_i^{mc}, i = 1, \ldots, N^{mc}\}$ *is called a Chamberlinian* **monopolistic-competition equilibrium** *if*

1. *Firms: Each firm behaves as a monopoly over its brand; that is, given the demand for brand i (7.14), each firm i chooses* q_i^{mc} *to* $\max_{q_i} \pi_i = p_i(q_i)q_i - (F + cq_i)$.

2. *Consumers: Each consumer takes his income and prices as given and maximizes (7.12) subject to (7.13), yielding a system of demand functions (7.14).*

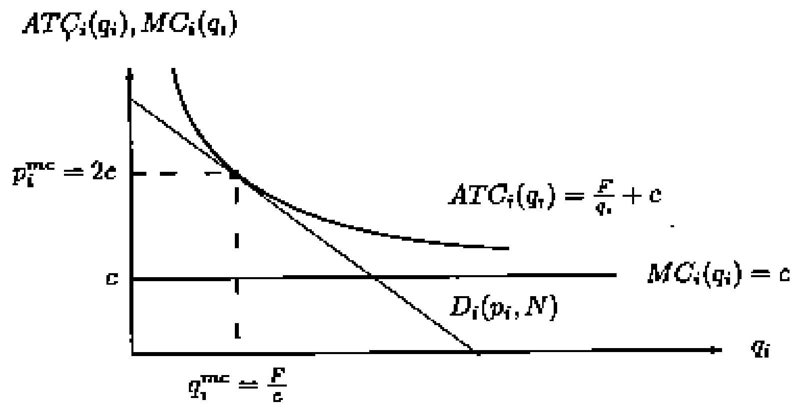

Figure 7.6: Decreasing average-cost technology

3. *Free entry: Free entry of firms (brands) will result in each firm making zero profits;* $\pi_i(q_i^{mc}) = 0$ *for all* $i = 1, 2, \ldots, N$.

4. *Resource constraint: Labor demanded for production equals the total labor supply;* $\sum_{i=1}^{N}(F + cq_i) = L$.

Definition 7.3 can be easily interpreted using Figure 7.6. The demand facing each (existing) brand-producing firm depends on the total number of brands in the industry, N. When N increases, the demand facing each brand-producing firm shifts downward, reflecting the fact that consumers partially substitute higher consumption levels of each brand with a lower consumption spread over a large number of brands. Therefore, free entry increases the number of brands until the demand facing each firm becomes tangent to the firm's average cost function. At this point, each existing brand-producing firm makes zero profit, and entry stops. The equilibrium condition in which demand becomes tangent to the average cost of each firm is known as *Chamberlin's tangency condition.*

Two important observations follow from the tangency condition displayed in Figure 7.6. First, in equilibrium the price of each brand equals average cost. Second, in equilibrium all brand-producing firms produce on the downward sloping part of the average cost curve. Thus, firms do not minimize average cost under a monopolistic-competition market structure.

Solving for a monopolistic-competition equilibrium

A firm's profit-maximization problem (item 1 of Definition 7.3) is the already familiar monopoly's problem analyzed in chapter 5. In that chapter we showed that if a monopoly produces a strictly positive amount of output, then the monopoly's price would satisfy

$$MR_i(q_i) = p_i\left(1 + \frac{1}{\eta}\right) = p_i\left(1 + \frac{1}{-2}\right) = \frac{p_i}{2} = c = MC(q_i).$$

Hence, the equilibrium price of each brand is given by $p_i^{mc} = 2c$ (twice the marginal cost).

The zero-profit condition (item 3 of Definition 7.3) implies that $0 = \pi_i(q_i^{mc}) = (p_i^{mc} - c)q_i^{mc} - F = cq_i^{mc} - F$. Hence, $q_i^{mc} = F/c$.

We are left to find how many brands will be produced in this economy. The resource-constraint condition (item 4 of Definition 7.3) implies that $N[F + c(F/c)] = L$. Hence, $N = L/(2F)$. Altogether, we have it that

Proposition 7.5

1. *In a monopolistic competition equilibrium with strictly positive fixed and marginal cost, only a finite number of brands will be produced. The equilibrium is given by*

$$p_i^{mc} = 2c;\ \ q_i^{mc} = \frac{F}{c};\ \ N^{mc} = \frac{L}{2F}.$$

2. *When the fixed cost is large, there will be a low variety of brands, but each brand will be produced/consumed in a large quantity. When the fixed cost is low, there will be a large variety of brands, and each will be produced/consumed in a small quantity.*

7.2.2 Monopolistic competition in international markets

In the late 1970s trade theorists began applying the theory of monopolistic competition to international trade (see Helpman and Krugman 1985). The major motivation was that the neoclassical international trade theory failed to explain the data showing that most international trade consists of trade with similar products (intraindustry trade) rather of very different products (interindustry trade) as predicted by the traditional factor-proportion theory. That is, the application of monopolistic competition was needed in order to explain why countries trade in similar products. There are two (mutually dependent) ways for explaining gains from trade under increasing-returns production technologies: (a) trade increases specialization, thereby enabling firms to produce at a higher scale and therefore at a lower average cost; and (b) trade increases the world variety of brands facing each consumer in each country.

Consider a two-country world economy, in which each country is identical to the one analyzed above. Under autarky (no trade), each country is described by Proposition 7.5. Our first question is what would happen to the patterns of production and consumption when the two countries start trading (move to a free-trade regime)?

When the world is integrated into a single large economy, the labor resource and the number of consumers basically doubles. In view of the equilibrium described in Proposition 7.5, there will be no change in brand prices and the level of production of each brand. However, the number of brands under free trade will double and become $N^f = L/F = 2N^a$, where f and a denote equilibrium values under free trade and under autarky, respectively. Also, note that since the quantity produced of each brand remains unchanged ($q_i^f = q_i^a = F/c$), but the entire population has doubled, under free trade each consumer (country) consumes one-half of the world production ($F/(2c)$).

Our second question is whether there are gains from trade, given that we found that the consumption level of each brand has decreased to one-half the autarky level while the number of brands has doubled. In order to answer that, we should calculate the equilibrium utility levels under autarky and under free trade. Thus,

$$\begin{aligned} u^f &= N^f\sqrt{q_i^f} = \frac{L}{F}\sqrt{\frac{F}{2c}} = \frac{L}{\sqrt{2}\sqrt{cF}} \qquad (7.16)\\ &> \frac{L}{2\sqrt{cF}} = \frac{L}{2F}\sqrt{\frac{F}{c}} = N^a\sqrt{q_i^a} = u^a. \end{aligned}$$

Hence, each consumer in each country gains from trade. The intuition is quite simple. Comparing point a with point f in Figure 7.5 shows that a consumer is always better off if the variety doubles, despite the decline in the consumption level of each brand.

We conclude our analysis of the gains from trade with two remarks. First, we have shown that, under monopolistic competition, free trade yields a higher welfare level than autarky. However, Gros (1987) has shown that countries may benefit from imposing some import tariff on foreign-produced brands. Second, let us note that we have shown there are gains from trade when there is only one industry producing differentiated brands. Chou and Shy (1991) have shown that the gains from trade in monopolistic competition extend to the case where some industries produce nontraded brands; however, the remote possibility that trade may reduce the welfare of all countries (Pareto inferior trade) remains.

7.3 "Location" Models

In this section we present models in which consumers are heterogeneous. That is, due to different tastes or location, each consumer has a different preference for the brands sold in the market.

There could be two interpretations of "location" for the environment modeled in this section: Location can mean the physical location of a particular consumer, in which case the consumer observes the prices charged by all stores and then chooses to purchase from the store at which the price plus the transportation cost is minimized. Or, location can mean a distance between the brand characteristic that a particular consumer views as ideal and the characteristics of the brand actually purchased. That is, we can view a space (say, a line interval) as measuring the degree of sweetness in a candy bar. Consumers located toward the left are those who prefer low-sugar bars, whereas those who are located toward the right prefer high-sugar bars. In this case, the distance between a consumer and a firm can measure the consumers' disutility from buying a less-than-ideal brand. This disutility is equivalent to the transportation cost in the previous interpretation.

We analyze only horizontally differentiated products. That is, we analyze brands that are not uniformly utility ranked by all consumers. More precisely, horizontally differentiated brands are ones that, if sold for identical prices, elicit from different consumers choices of different brands (called ideal brands). The analysis of vertically differentiated brands, that is, brands that are uniformly ranked by all consumers, is postponed to section 12.2, where we discuss product differentiation with respect to quality (see more on these issues in Beath and Katsoulacos [1991] and Anderson, Palma, and Thisse [1992]; for a survey see Gabszewicz and Thisse [1992]).

7.3.1 The linear approach

Hotelling (1929) considers consumers who reside on a linear street with a length of $L > 0$. Suppose that the consumers are uniformly distributed on this interval, so at each point lies a single consumer. Hence, the total number of consumers in the economy is L. Each consumer is indexed by $x \in [0, L]$, so x is just a name of a consumer (located at point x from the origin).

Price game with fixed location

Suppose that there are two firms selling a product that is identical in all respects except one characteristic, which is the location where it is sold. That is, Figure 7.7 shows that firm A is located a units of distance from point 0. Firm B is located to the right of firm A, b units of distance from point L. Assume that production is costless.

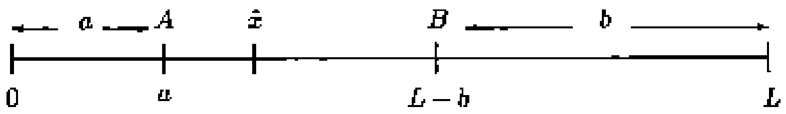

Figure 7.7: Hotelling's linear city with two firms

Each consumer buys one unit of the product. To go to a store, a consumer has to pay transportation cost of τ per unit of distance. Thus, a consumer located at some point x has to pay transportation cost of $\tau|x-a|$ for shopping at firm A, or $\tau|x-(L-b)|$ for shopping at firm B. The reader should note that distance here can have a different interpretation. We can think of a candy bar that can be produced with different degrees of sweetness. Thus, if we let x measure the percentage of sugar put into a candy bar, firm B produces a sweeter candy than firm A. A consumer located at x desires x degree of sweetness more than any other degree of sweetness. However the firms offer most consumers degrees of sweetness that differ from the most preferred one. With this interpretation, the equivalent of transportation costs is the monetary equivalent loss to a consumer who desires x degree of sweetness but instead has to purchase a candy bar with a different degree of sweetness.

Let us define the utility function of a consumer located at point x by

$$U_x \equiv \begin{cases} -p_A - \tau|x-a| & \text{if he buys from } A \\ -p_B - \tau|x-(L-b)| & \text{if she buys from } B. \end{cases} \tag{7.17}$$

Let $\hat{x}$ denote the consumer who is indifferent to whether he or she purchases from A or B. Formally, if $a < \hat{x} < L-b$, then

$$-p_A - \tau(\hat{x}-a) = -p_B - \tau(L-b-\hat{x}).$$

Hence,

$$\hat{x} = \frac{p_B - p_A}{2\tau} + \frac{L-b+a}{2},$$

which is the demand function faced by firm A. The demand function faced by firm B is

$$L - \hat{x} = \frac{p_A - p_B}{2\tau} + \frac{L+b-a}{2}.$$

We now look for a Bertrand-Nash equilibrium in price strategies. That is, Firm A takes p_B as given and chooses p_A to

$$\max_{p_A} \pi_A = \frac{p_B p_A - (p_A)^2}{2\tau} + \frac{(L-b+a)p_A}{2}. \tag{7.18}$$

The first-order condition is given by

$$0 = \frac{\partial \pi_A}{\partial p_A} = \frac{p_B - 2p_A}{2\tau} + \frac{(L-b+a)}{2}. \tag{7.19}$$

Firm B takes p_A as given and chooses p_B to

$$\max_{p_B} \pi_B = \frac{p_B p_A - (p_B)^2}{2\tau} + \frac{(L+b-a)p_B}{2}. \tag{7.20}$$

The first-order condition is given by

$$0 = \frac{\partial \pi_B}{\partial p_B} = \frac{p_A - 2p_B}{2\tau} + \frac{L+b-a}{2}.$$

Hence, the equilibrium prices are given by

$$p_A^h = \frac{\tau(3L-b+a)}{3} \text{ and } p_B^h = \frac{\tau(3L+b-a)}{3}. \tag{7.21}$$

The equilibrium market share of firm A is given by

$$\hat{x}^h = \frac{3L-b+a}{6}. \tag{7.22}$$

Note that if $a = b$, then the market is equally divided between the two firms. The profit of firm A is given by

$$\pi_A^h = \hat{x}^h p_A^h = \frac{\tau(3L-b+a)^2}{18}, \tag{7.23}$$

which shows that the profit of each brand-producing firm increases with the transportation cost parameter, τ. This is not surprising in view of the fact that Propositions 7.1 and 7.2 showed firms reach higher profit levels when the brands they produce are more differentiated. In fact, Hotelling (1929, 50) states

> These particular merchants would do well, instead of organizing improvement clubs and booster associations to better the roads, to make transportation as difficult as possible.

We leave it to the reader to determine whether such a behavior is observed or unobserved.

The above calculations were performed under the assumption that an equilibrium where firms charge strictly positive prices always exists. The following proposition describes the equilibria and provides precise conditions for existence. The proof of the proposition is given in the appendix (section 7.5).

Proposition 7.6

1. *If both firms are located at the same point ($a + b = L$, meaning that the products are homogeneous), then $p_A = p_B = 0$ is a unique equilibrium.*
2. *A unique equilibrium exists and is described by (7.21) and (7.22) if and only if the two firms are not too close to each other; formally if and only if*

$$\left(L + \frac{a-b}{3}\right)^2 \geq \frac{4L(a+2b)}{3} \quad \text{and} \quad \left(L + \frac{b-a}{3}\right)^2 \geq \frac{4L(b+2a)}{3}$$

the unique equilibrium is given by (7.21), (7.22), and (7.23).

When the two firms are located too closely, they start undercutting each other's prices, resulting in a process of price cuts that does not converge to an equilibrium. Proposition 7.6 shows that in order for an equilibrium to exist, the firms cannot be too closely located.

Location and price game

So far, we have assumed that the location of the firms is fixed, say, by the regulating (license-issuing) authority. It would be nice to have a theory under which firms can choose price and location. Unfortunately, we now show that there is no solution for this two-dimensional strategy game.

To show that, we ask what would firm A do if, given the price and location of its opponent, it would be allowed to relocate. To answer that, (7.23) implies that

$$\frac{\partial \pi_A}{\partial a} > 0,$$

meaning that for any locations a and b, firm A could increase its profit by moving toward firm B (obviously, to gain a higher market share). This case, where firms tend to move toward the center, is called in the literature the *principle of minimum differentiation* since by moving toward the center the firms produce less-differentiated products. However,

Proposition 7.6 shows that if firm A gets too close to firm B, an equilibrium will not exist. Also, if firm A locates at the same point where firm B locates, its profit will drop to zero, implying that it is better off to move back to the left. Hence

Proposition 7.7 *In the Hotelling linear-city game, there is no equilibrium for the game where firms use both prices and location as strategies.*

Quadratic transportation cost

Proposition 7.21 shows that even when the location is fixed, the linear-location model does not have an equilibrium in a price game when the firms are too close to each other. We also showed that there is no equilibrium in a game when firms choose both prices and location.

However, it is important to observe that so far, we have assumed linear transportation costs. The existence problem can be solved if we assume quadratic transportation costs. That is, let (7.17) be written as

$$U_x \equiv \begin{cases} -p_A - \tau(x-a)^2 & \text{if he buys from } A \\ -p_B - \tau[x-(L-b)]^2 & \text{if she buys from } B. \end{cases} \tag{7.24}$$

To have even more fun, using the quadratic-transportation-cost setup, we can formulate a two-period game in which firms decide where to locate in the first period, and set prices in the second period. Since we look for a SPE (Definition 2.10), the reader who is eager to solve this game should follow the following steps:

Second period:

1. For given location parameters a and b, find the Nash-Bertrand equilibrium prices, following the same steps we used in order to derive (7.21).
2. Substitute the equilibrium prices into the profit functions (7.18) and (7.20) to obtain the firms' profits as functions of the location parameters a and b.

First period: Maximize the firms' profit functions which you calculated for the second period with respect to a for firm A and with respect to b for firm B. Prove that for a given b, $\frac{\partial \pi_A}{\partial a} < 0$, meaning that firm A would choose $a = 0$. Similarly, show that firm B would locate at point L.

This exercise shows that when there are quadratic transportation costs, firms will choose maximum differentiation. This result is consistent with Propositions 7.1 and 7.2, showing that profits increase with differentiation.

7.3.2 The circular approach

Proposition 7.7 shows that an equilibrium in games in which firms jointly decide on prices and location does not exist in the Hotelling model. One way to solve this problem is to let the city be the unit-circumference circle, where the consumers are uniformly distributed on the circumference.

As with the Hotelling model, this location model can also be given an interpretation for describing differentiated products that differs from the physical-location interpretation. Consider for example airline, bus, and train firms which can provide a round-the-clock service. If we treat the circle as twenty-four hours, each brand can be interpreted as the time where an airline firm schedules a departure.

Firms

This model does not explicitly model how firms choose where to locate. However, it assumes a monopolistic-competition market structure, in which the number of firms N is endogenously determined. All (infinitely many) potential firms have the same technology. Denoting by F the fixed cost, by c the marginal cost, and by q_i and $\pi_i(q_i)$ the output and profit levels of the firm-producing brand i, we assume that

$$\pi_i(q_i) = \begin{cases} (p_i - c)q_i - F & \text{if } q_i > 0 \\ 0 & \text{if } q_i = 0. \end{cases} \tag{7.25}$$

Consumers

Consumers are uniformly distributed on the unit circle. We denote by τ the consumers' transportation cost per unit of distance. Each consumer buys one unit of the brand that minimizes the sum of the price and transportation cost.

Assuming that the N firms are located at an equal distance from one another yields that the distance between any two firms is $1/N$. Figure 7.8 illustrates the position of firm 1 relative to the positions of firm 2 and firm N. Then, assuming that firms 2 and N charge a uniform price p, the consumer who is indifferent to whether he or she buys from firm 1 or firm 2 (similarly, firm N) is located at $\hat{x}$ determined by $p_1 + \tau\hat{x} = p + \tau(1/N - \hat{x})$. Hence,

$$\hat{x} = \frac{p - p_1}{2\tau} + \frac{1}{2N}. \tag{7.26}$$

Since firm 1 has customers on its left and on its right, the demand function facing firm 1 is

$$q_1(p_1, p) = 2\hat{x} = \frac{p - p_1}{\tau} + \frac{1}{N}. \tag{7.27}$$

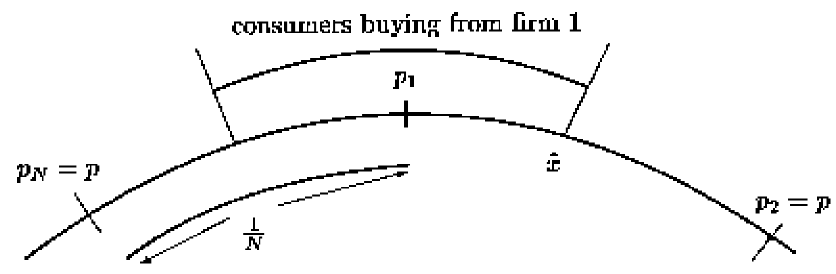

Figure 7.8: The position of firms on the unit circle

Defining and solving for the monopolistic-competition equilibrium

Let us begin with a definition:

DEFINITION 7.4 *The triplet $\{N^\circ, p^\circ, q^\circ\}$ is an equilibrium if*

1. *Firms: Each firm behaves as a monopoly on its brand; that is, given the demand for brand i (7.27) and given that all other firms charge $p_j = p^\circ$, $j \neq i$, each firm i chooses p° to*

$$\max_{p_i} \pi_i(p_i, p^\circ) = p_i q_i(p_i) - (F + cq_i) = (p_i - c)\left(\frac{p^\circ - p_i}{\tau} + \frac{1}{N}\right) - F.$$

2. *Free entry: Free entry of firms (brands) will result in zero profits; $\pi_i(p^\circ, p^\circ) = 0$ for all $i = 1, 2, \ldots, N^\circ$.*

The first-order condition for firm i's maximization problem is

$$0 = \frac{\partial \pi_i(p_i, p^\circ)}{\partial p_i} = \frac{p^\circ - 2p_i + c}{\tau} + \frac{1}{N}.$$

Therefore, in a symmetric equilibrium, $p_i = p^\circ = c + \tau/N$.

To find the equilibrium number of brands N, we set

$$0 = \pi_i(p^\circ, p^\circ) = (p^\circ - c)\frac{1}{N} - F = \frac{\tau}{N^2} - F.$$

Hence

$$N^\circ = \sqrt{\frac{\tau}{F}}, \quad p^\circ = c + \frac{\tau}{N^\circ} = c + \sqrt{\tau F}, \quad q^\circ = \frac{1}{N^\circ}. \tag{7.28}$$

Welfare

We would like to investigate whether the "free market" produces a larger or a smaller variety than the optimal variety level. Before defining the economy's welfare function, we calculate the economy's aggregate transportation costs, denoted by T. Figure 7.8 shows that in equilibrium, all consumers purchasing from firm 1, say, are located between 0 and $1/(2N)$ units of distance from the firm (on each side). Since there are $2N$ such intervals, the economy's total transportation cost is given by

$$T(N) = 2N\tau \left(\int_0^{\frac{1}{2N}} x\,dx \right) = 2N\tau \left[\frac{x^2}{2} \right]_0^{\frac{1}{2N}} = \frac{\tau}{4N}. \tag{7.29}$$

An alternative way to find the aggregate transportation cost without using integration is to look at the cost of the average consumer who is located half way between $\hat{x} = 1/(2N)$ and a firm. That is, the average consumer has to travel $1/(4N)$, which yields (7.29).

We define the economy's loss function, $L(F, \tau, N)$, as the sum of the fixed cost paid by the producing firms and the economy's aggregate transportation cost. Formally, the "Social Planner" chooses the optimal number of brands N^* to

$$\min_N L(F, \tau, N) \equiv NF + T(N) = NF + \frac{\tau}{4N}. \tag{7.30}$$

The first-order condition is $0 = \frac{\partial L}{\partial N} = F - \tau/(4N^2)$. Hence,

$$N^* = \frac{1}{2}\sqrt{\frac{\tau}{F}} < N^\circ. \tag{7.31}$$

Therefore, in a free-entry location model, too many brands are produced. Notice, that there is a welfare tradeoff between the economies of scale and the aggregate transportation cost. That is, a small number of brands is associated with lower average production costs but higher aggregate transportation costs (because of fewer firms). A large number of brands means a lower scale of production (higher average cost) but with a lower aggregate transportation cost. Equation (7.31) shows that it is possible to raise the economy's welfare by reducing the number of brands.

7.3.3 Sequential entry to the linear city

So far, we have not discussed any model in which firms strategically choose where to locate. In subsection 7.3.1 we have shown that the basic linear-street model does not have an equilibrium where firms choose both prices and location.

In this subsection, we discuss an example set forth by Prescott and Visscher (1977) in which prices are fixed at a uniform level set by the regulator. For example, in many countries, prices of milk, bread, and basic cheese products are regulated by the government. Thus, the only choice variable left to firms is where to locate (what characteristics - degree of sweetness in our example - the product should have).

Consider the unit interval (street) where there are three firms entering sequentially. In this three-period model, firm 1 enters in period 1, firm 2 in period 2, and firm 3 in period 3. We look for a SPE (see Definition 2.10) in location strategies, where each firm maximizes its market share.

We denote by $0 \le x_i \le 1$ the location strategy chosen by firm i (in period i), $i = 1, 2, 3$. Let ϵ denote a "very small" number, representing the smallest possible measurable unit of distance. Solving the entire three-period game is rather complicated. Instead, we shall assume that firm 1 has already moved and located itself at the point $x_1 = 1/4$. Figure 7.9 illustrates the location of firm 1.

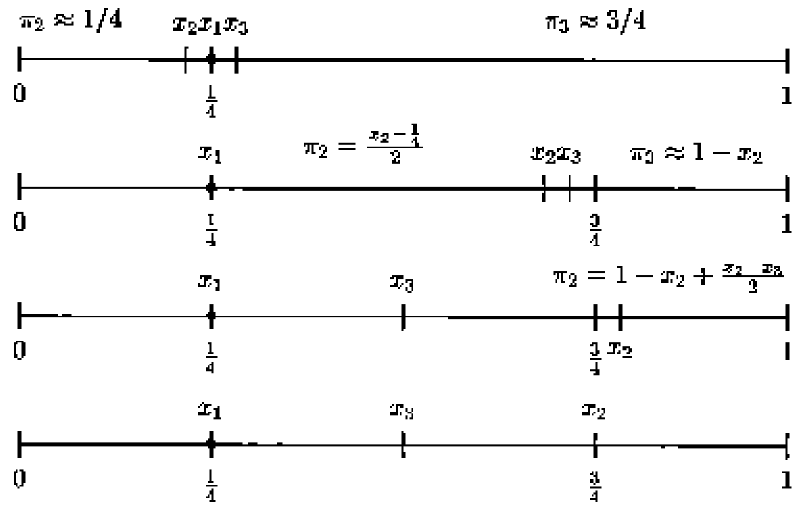

Figure 7.9: Sequential-location game

The third-period subgame

Firm 3 decides on its location x_3 after firm 1 and firm 2 are already located. There are three possible locations of firm 2 corresponding to the three upper parts of Figure 7.9.

$x_2 = \frac{1}{4} - \epsilon$: In this case firm 3 would locate at $x_3 = \frac{1}{4} + \epsilon$. Here, $\pi_3 \approx \frac{3}{4}$ while $\pi_2 = x_2 + \frac{1}{2}(\frac{1}{4} - x_2) < \frac{1}{4}$.

$\frac{1}{4} < x_2 < \frac{3}{4}$: In this case firm 3 would locate to the right of firm 2, at $x_3 = x_2 + \epsilon$. Here, $\pi_3 \approx 1 - x_2$ while $\pi_2 \approx \frac{x_2 - \frac{1}{4}}{2} < \frac{1}{4}$. That is, firm 2 shares the $[x_1, x_2]$ interval with firm 1.

$x_2 \geq \frac{3}{4}$: In this case firm 3 would locate between firm 1 and firm 2, at any point $x_1 < x_3 < x_2$. With no loss of generality, assume that $x_3 = \frac{\frac{1}{4} + x_2}{2}$. Here, $\pi_3 = \frac{x_2 - \frac{1}{4}}{2}$ and

$$\pi_2 = 1 - x_2 + \frac{x_2 - x_3}{2} = 1 - x_2 + \frac{1}{2}\left[x_2 - \frac{\frac{1}{4} + x_2}{2}\right] = \frac{15 - 12x_2}{16}.$$

The second-period subgame

Firm 2 knows that in the third period, the location decision of firm 3 will be influenced by its own choice of location. Thus, firm 2 calculates the best-response function of firm 3 (which we calculated above). Hence, firm 2 takes the *decision rule* of firm 3 as given and chooses x_2 that would maximize its profit. Clearly, firm 2 will not locate at $x_2 = \frac{1}{4} - \epsilon$ since this location yields a maximum profit of $\pi_2 \approx \frac{1}{4}$ (it will collect a higher profit by locating elsewhere, as described below).

If firm 2 locates at $\frac{1}{4} < x_2 < \frac{3}{4}$, we have shown that $x_3 = x_2 + \epsilon$ and $\pi_2 = \frac{x_2 - \frac{1}{4}}{2} < \frac{1}{4}$.

However, if firm 2 locates at $x_2 \geq \frac{3}{4}$ we have shown that $\pi_2 = \frac{15 - 12x_2}{16}$, which is maximized at $x_2 = \frac{3}{4}$. Located at $x_2 = \frac{3}{4}$, the profit of firm 2 is $\pi_2 = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$.

In summary, the SPE is reached where

$$x_2 = \frac{3}{4}, \quad \pi_2 = \frac{1}{4} + \frac{1}{8} \quad \text{and} \quad x_3 = \frac{1}{2}, \quad \pi_3 = \frac{1}{4}. \tag{7.32}$$

The bottom part of Figure 7.9 illustrates the location of the firms in a SPE.

7.3.4 Calculus-free location model

In this subsection we develop a calculus-free version of the Hotelling linear-city model analyzed in subsection 7.3.1.

Consider a city where consumers and producers are located only at the city's edges. Suppose that the city consists of N_0 consumers located at point $x = 0$ and N_L consumers located at the point $x = L$. There are two firms; firm A is located also at $x = 0$ and firm B is located at $x = L$. Assume that production is costless. Figure 7.10 illustrates the location of firms and consumers in this city.

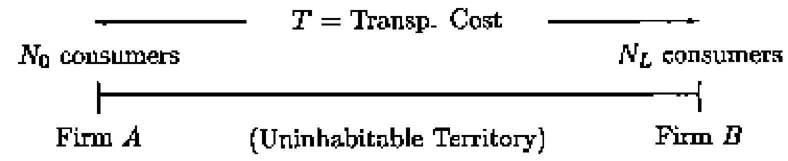

Figure 7.10: Discrete-location model

Each consumer buys one unit either from the firm located where the consumer is, or from the firm located on the other side of town. Shopping nearby does not involve transportation cost, whereas shopping on the other side of town involves paying a fixed transportation cost of $T \geq 0$. Let p_A denote the price charged by firm A, and p_B the price charged by firm B. Thus, we assume that the utility of the consumer located at point $x = 0$ is given by

$$U_0 \equiv \begin{cases} -p_A & \text{buying from } A \\ -p_B - T & \text{buying from } B. \end{cases} \tag{7.33}$$

Similarly, the utility of the consumer located at point $x = L$ is given by

$$U_L \equiv \begin{cases} -p_A - T & \text{buying from } A \\ -p_B & \text{buying from } B. \end{cases} \tag{7.34}$$

Let n_A denote the number of consumers buying from firm A, and n_B denote the number of consumers buying from firm B. Then, (7.33) and (7.34) imply that

$$n_A = \begin{cases} 0 & \text{if } p_A > p_B + T \\ N_0 & \text{if } p_B - T \leq p_A \leq p_B + T \\ N_0 + N_L & \text{if } p_A < p_B - T \end{cases} \tag{7.35}$$

$$n_B = \begin{cases} 0 & \text{if } p_B > p_A + T \\ N_L & \text{if } p_A - T \leq p_B \leq p_A + T \\ N_0 + N_L & \text{if } p_B < p_A - T. \end{cases}$$

Non-existence of a Nash-Bertrand equilibrium

A Nash-Bertrand equilibrium is the nonnegative pair $\{p_A^N, p_B^N\}$, such that for a given p_B^N, firm A chooses p_A^N to maximize $\pi_A \equiv p_A n_A$; and for a given p_A^N, firm B chooses p_B^N to maximize $\pi_B \equiv p_B n_B$, where n_A and n_B are given in (7.35).

Proposition 7.8 *There does not exist a Nash-Bertrand equilibrium in prices for the discrete version of Hotelling's location model.*

Proof. By way of contradiction, suppose that $\{p_A^N, p_B^N\}$ constitute a Nash equilibrium. Then, there are three cases: (i) $|p_A^N - p_B^N| > T$; (ii) $|p_A^N - p_B^N| < T$; and (iii) $|p_A^N - p_B^N| = T$.

(i) With no loss of generality, suppose that $p_A^N - p_B^N > T$. Then, (7.35) implies that $n_A^N = 0$, and hence $\pi_A^N = 0$. However, firm A can deviate and increase its profit by reducing its price to $\tilde{p}_A = p_B^N + T$ and by having $\tilde{n}_A = N_0$, thereby earning a profit of $\tilde{\pi}_A = N_0(p_B^N + T)$. A contradiction.

(ii) With no loss of generality, suppose that $p_A^N < p_B^N + T$. Then, firm A can deviate and increase its profit by slightly increasing its price to $\tilde{p}_A$, satisfying $p_A^N < \tilde{p}_A < p_B^N + T$ and obtaining a profit level of $\tilde{\pi}_A = N_0\tilde{p}_A > \pi_A^N$. A contradiction.

(iii) With no loss of generality, suppose that $p_A^N - p_B^N = T$. Then, $p_B^N = p_A^N - T < p_A^N + T$. Hence, as firm A did in case (ii), firm B can increase its profit by slightly raising p_B^N. A contradiction. ∎

Undercutproof equilibrium

Since a Nash equilibrium in prices for the discrete-location model does not exist, in this subsection we define, motivate, and solve for the undercutproof equilibrium.

In an undercutproof equilibrium, each firm chooses the highest possible price, subject to the constraint that the price is sufficiently low so that the rival firm would not find it profitable to set a sufficiently lower price in order to grab the entire market. That is, in an undercutproof equilibrium, firms set prices at the levels that ensure that competing firms would not find it profitable to completely undercut these prices. Thus, unlike behavior in a Nash-Bertrand environment, where each firm assumes that the rival firm does not alter its price, in an undercutproof equilibrium environment, firms assume that rival firms are ready to reduce their prices whenever undercutting prices and grabbing their rival's market are profitable to them. This behavior is reasonable for firms competing in differentiated products.

Definition 7.5 *An undercutproof equilibrium for this economy is nonnegative n_A^U, n_B^U, and p_A^U, p_B^U such that*

1. *For given p_B^U and n_B^U, firm A chooses the highest price p_A^U subject to*
$$\pi_B^U \equiv p_B^U n_B^U \ge (N_0 + N_L)(p_A^U - T).$$
2. *For given p_A^U and n_A^U, firm B chooses the highest price p_B^U subject to*
$$\pi_A^U \equiv p_A^U n_A^U \ge (N_0 + N_L)(p_B^U - T).$$
3. *The distribution of consumers between the firms is determined in (7.35).*

Part 1 of Definition 7.5 states that in an undercutproof equilibrium, firm A sets the highest price under the constraint that the price is sufficiently low to prevent firm B from undercutting p_A^U and grabbing the entire market. More precisely, firm A sets p_A^N sufficiently low so that B's equilibrium profit level exceeds B's profit level when it undercuts by setting $\tilde{p}_B = p_A^U - T$, and grabbing the entire market ($n_B = N_0 + N_L$). Part 2 is similar to part 1 but describes how firm B sets its price. We proceed with solving for the equilibrium prices.

Proposition 7.9 *There exists a unique undercutproof equilibrium for the discrete-location problem given by $n_A^U = N_0$, $n_B^U = N_L$, and*

$$p_A^U = \frac{(N_0 + N_L)(N_0 + 2N_L)T}{(N_0)^2 + N_0N_L + (N_L)^2} \quad \textit{and} \quad p_B^U = \frac{(N_0 + N_L)(2N_0 + N_L)T}{(N_0)^2 + N_0N_L + (N_L)^2}. \tag{7.36}$$

Proof. First note that by setting $p_i \le T$, each firm can secure a strictly positive market share without being undercut. Hence, in an undercutproof equilibrium both firms maintain a strictly positive market share. From (7.35), we have it that $n_A^U = N_0$ and $n_B^U = N_L$. Substituting $n_A^U = N_0$ and $n_B^U = N_L$ into the two constraints in Definition 7.5 and then verifying (7.35) yields the unique undercutproof equilibrium. ∎

Figure 7.11 illustrates how the undercutproof equilibrium is determined. The left side of Figure 7.11 shows how firm A is constrained in setting p_A to fall into the region where firm B would not benefit from undercutting p_A^U (compare with part 1 in Definition 7.5). The center of Figure 7.11 shows how firm B is constrained in setting p_B to fall into the region where firm A would not benefit from undercutting p_B^U (compare with part 2 in Definition 7.5). The right side of Figure 7.11 illustrates the region where neither firm finds it profitable to undercut the rival firm and the undercutproof equilibrium prices. It should be emphasized that the curves drawn in Figure 7.11 are not best-response (reaction) functions. The curves simply divide the regions into prices that make undercutting profitable or unprofitable for one firm.

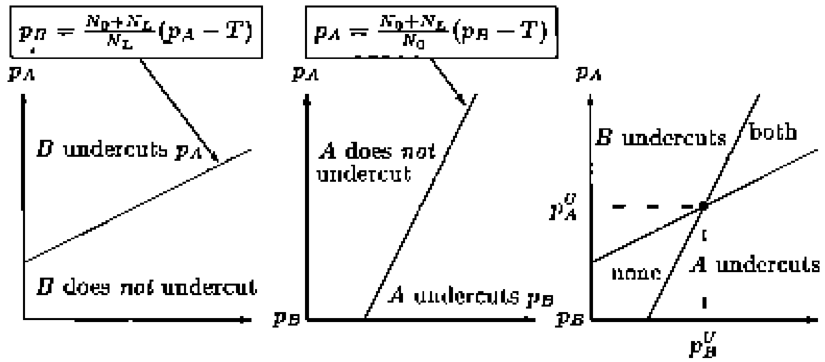

Figure 7.11: Undercutproof equilibrium for the discrete-location model

Properties of the undercutproof equilibrium

Clearly, prices rise with transportation costs and monotonically decline to zero as transportation costs approach zero, reflecting a situation in which the products become homogeneous. More interestingly,

$$\Delta p \equiv p_B - p_A = \frac{(N_0 + N_L)(N_0 - N_L)T}{(N_0)^2 + N_0 N_L + (N_L)^2}. \tag{7.37}$$

Hence, $\Delta p \geq 0$ if and only if $N_0 \geq N_L$. Thus, in an undercutproof equilibrium, the firm selling to the larger number of consumers charges a lower price. This lower price is needed to secure the firm from being totally undercut.

Finally, under symmetric distribution of consumers ($N_0 = N_L$), the equilibrium prices are given by $p_A^U = p_B^U = 2T$. That is, each firm can mark up its price to twice the level of the transportation cost without being undercut.

7.4 Appendix: Inverting Demand Systems

The demand system (7.1) can be written as

$$\begin{bmatrix} \beta & \gamma \\ \gamma & \beta \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} \alpha - p_1 \\ \alpha - p_2 \end{bmatrix}.$$

Define Δ to be the determinant of

$$\Delta \equiv \det \begin{bmatrix} \beta & \gamma \\ \gamma & \beta \end{bmatrix} = \beta^2 - \gamma^2.$$

Then, using Cramer's Law we have it that

$$q_1 = \frac{1}{\Delta} \det \begin{bmatrix} \alpha - p_1 & \gamma \\ \alpha - p_2 & \beta \end{bmatrix} = \frac{\alpha(\beta - \gamma) - \beta p_1 + \gamma p_2}{\beta^2 - \gamma^2}$$

$$q_2 = \frac{1}{\Delta} \det \begin{bmatrix} \beta & \alpha - p_1 \\ \gamma & \alpha - p_2 \end{bmatrix} = \frac{\alpha(\beta - \gamma) - \beta p_2 + \gamma p_1}{\beta^2 - \gamma^2}.$$

This establishes equation (7.2).

7.5 Appendix: Existence of an Equilibrium in the Linear City

We now prove Proposition 7.6. (1) When $a + b = 1$, the products are homogeneous, so the undercutting procedure described in section 6.3 applies.

(2) For the general proof see d'Aspremont, Gabszewicz, and Thisse 1979. Here, we illustrate the argument made in their proof for the simple case where firms are located at equal distances along the edges. That is, assume that $a = b$, $a < L/2$. Then, we are left to show that the equilibrium exists if and only if $L^2 \geq 4La$, or if and only if $a \leq L/4$.

When $a = b$, the distance between the two firms is $L - 2a$. Also, if equilibrium exists, (7.21) is now given by $p_A = p_B = \tau L$. The profit level of firm A as a function of its own price p_A and a *given* B's price $\bar{p}_B = \tau L$ for the case of $a = b$ is drawn in Figure 7.12.

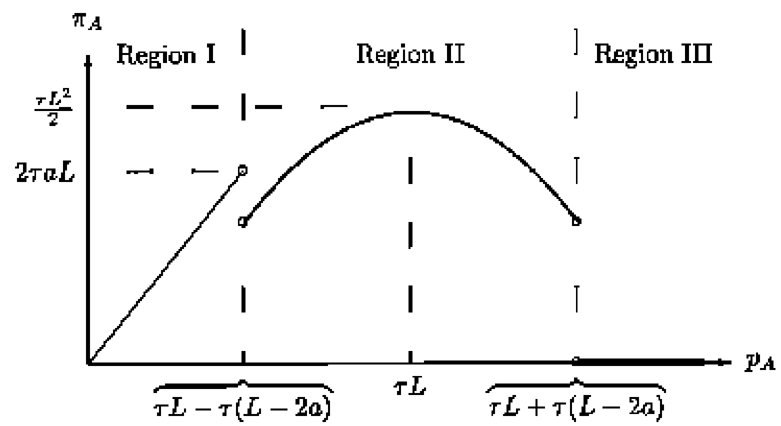

Figure 7.12: Existence of equilibrium in the linear city: The profit of firm A for a given $\bar{p}_B = \tau L$

Figure 7.12 has three regions:

Region I: Here, $p_A < \tau L - \tau(L-2a)$. In this case, p_A is very low, so that even the consumer located at the same point where firm B is located would purchase from firm A. Thus, firm A has the entire market, and its profit is given by $\pi_A = p_A L$.

Region II: Here, both firms sell a strictly positive amount, so the profit of firm A as a function of p_A is given in equation (7.18). Substituting the equilibrium $p_B = \tau L$ into (7.18) yields

$$\pi_A = p_A L - \frac{(p_A)^2}{2\tau}, \tag{7.38}$$

which is drawn in Region II of Figure 7.12. Maximizing (7.38) with respect to p_A yields $\pi_A = \tau L^2/2$, which corresponds to the peak drawn in Figure 7.12.

Region III: Here, p_A is high, so all consumers purchase from firm B. This is the polar case of Region I.

Now, for a given $p_B = \tau L$, Figure 7.12 shows that π_A has two local maxima. In one it has the entire market share ($p_A = \tau L - \tau(L-2a) - \epsilon$), whereas in the other it shares the market with firm B ($p_A = \tau L$). For (7.21) to constitute the equilibrium prices, we must have it that in equilibrium, the globally profit-maximizing price for firm A would lie in Region II (and not Region I). Or, that for the equilibrium $p_B = \tau L$,

$$\pi_A^{II} = \frac{\tau L^2}{2} \geq \pi_A^{I} = [\tau L - \tau(L-2a)]L = 2\tau a L,$$

implying that $a \leq L/4$. ■

7.6 Exercises

1. Suppose that there are only two firms selling coffee, called firms 1 and 2. Let α_i denote the advertising level of firm i, $i = 1, 2$. Assume that the profits of the firms are affected by the advertising levels taken by the firms. Formally, assume that

$$\pi_1(\alpha_1, \alpha_2) \equiv 4\alpha_1 + 3\alpha_1\alpha_2 - (\alpha_1)^2 \text{ and } \pi_2(\alpha_1, \alpha_2) \equiv 2\alpha_2 + \alpha_1\alpha_2 - (\alpha_2)^2.$$

Answer the following questions:

(a) Calculate and draw the best-response function of each firm. That is, for any given advertising level of firm j, find the profit-maximizing advertising level of firm i.

(b) Infer whether the strategies are strategically complements or strategically substitutes (see Definition 7.2).

(c) Find the Nash equilibrium advertising levels. Also, calculate the firms' Nash equilibrium profit levels.

2. Consider the Hotelling linear-city model analyzed in Subsection 7.3.1. Suppose that in the linear city there is only one restaurant, located at the center of the street with a length of 1 km. Assume that the restaurant's cost is zero. Consumers are uniformly distributed on the street, which is the interval [0, 1], where at each point on the interval lives one consumer. Suppose that the transportation cost for each consumers is \$1 for each unit of distance (each kilometer of travel). The utility of a consumer who lives a units of distance from the restaurant is given by $U \equiv B - a - p$, where p is the price of a meal, and B is a constant. However, if the consumer does not eat at the restaurant, her utility is $U = 0$. Answer the following questions:

(a) Suppose that the parameter B satisfies $0 < B < 1$. Find the number of consumers eating at this restaurant. Calculate the monopoly restaurant's price and profit levels.

(b) Answer the previous question assuming that $B > 1$.

3. University Road is best described as the interval [0, 1]. Two fast-food restaurants serving identical food are located at the edges of the road, so that restaurant 1 is located on the most left-hand side, and restaurant 2 is located on the most right-hand side of the road. Consumers are uniformly distributed on the interval [0, 1], where at each point on the interval lives one consumer. Each consumer buys one meal from the restaurant in which the price plus the transportation cost is the lowest.

In University Road, the wind blows from right to left, hence the transportation cost for a consumer who travels to the right is \$$R$ per unit of distance, and only \$1 per unit of distance for a consumer who travels to the left. Answer the following questions.

(a) Let p_i denote the price of a meal at restaurant i, $i = 1, 2$. Assume that p_1 and p_2 are given and satisfy

$$0 < p_1 - R < p_2 < 1 + p_1.$$

Denote by $\hat{x}$ the location of the consumer who is indifferent to whether he or she eats at restaurant 1 or restaurant 2 and calculate $\hat{x}$ as a function of p_1, p_2, and R.

(b) Suppose that the given prices satisfy $p_1 = p_2$. What is the minimal value of the parameter R such that all consumers will go to eat only at restaurant 1?

4. Consider the Hotelling model with quadratic transportation cost described in equation (7.24) and assume that both firms are located at the same distances from the edges of the unit interval (i.e., $a = b \geq 0$ in Figure 7.7).

(a) Assuming that firms produce the product with zero cost, calculate the (symmetric) Nash equilibrium in prices.

(b) Assuming that firm A is allowed to make a small adjustment in its location before both firms choose their prices; would firm A move inward or outward? Prove your answer!

7.7 References

Anderson, S., A. Palma, and J. Thisse. 1992. *Discrete Choice Theory of Product Differentiation*. Cambridge, Mass.: MIT Press.

Beath, J., and Y. Katsoulacos. 1991. *The Economic Theory of Product Differentiation*. Cambridge: Cambridge University Press.

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

Chamberlin, E. 1933. *The Theory of Monopolistic Competition*. Cambridge, Mass.: Harvard University Press.

Cheng, L. 1985. "Comparing Bertrand and Cournot Equilibria: A Geometric Approach." *Rand Journal of Economics* 16: 146–152.

Chou, C., and O. Shy. 1991. "Intraindustry Trade and the Variety of Home Products." *Canadian Journal of Economics* 24: 405–416.

d'Aspremont, C., J. Gabszewicz, and J. Thisse. 1979. "On Hotelling's Stability in Competition." *Econometrica* 17: 1145–1151.

Dixit, A. 1979. "A Model of Duopoly Suggesting a Theory of Entry Barriers." *Bell Journal of Economics* 10: 20–32.

Dixit, A., and J. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review* 67: 297–308.

Eaton, B. C., and R. Lipsey. 1989. "Product Differentiation." In *Handbook of Industrial Organization*, edited by R. Schmalensee and R. Willig. Amsterdam: North-Holland.

Gabszewicz J., and J. Thisse. 1992. "Location." In *Handbook of Game Theory*, edited by R. Aumann and S. Hart. Amsterdam: North-Holland.

Gros, D. 1987. "A Note on the Optimal Tariff, Retaliation, and the Welfare Loss from Tariff Wars in a Model with Intra-Industry Trade." *Journal of International Economics* 23: 457–367.

Hotelling, H. 1929. "Stability in Competition." *Economic Journal* 39: 41–57.

Helpman, E., and P. Krugman. 1985. *Market Structure and Foreign Trade*. Cambridge, Mass.: MIT Press.

Lancaster, K. 1979. *Variety Equity, and Efficiency*. New York: Columbia University Press.

Prescott, E., and M. Visscher. 1977. "Sequential Location among Firms with Foresight." *The Bell Journal of Economics* 8: 378–393.

Salop, S. 1979. "Monopolistic Competition with Outside Goods." *Bell Journal of Economics* 10: 141–156.

Sattinger, M. 1984. "Value of an Additional Firm in Monopolistic Competition." *Review of Economic Studies* 51: 321–332.

Scherer, F. M. 1979. "The Welfare Economics of Product Variety: An Application to the Ready-To-Eat Cereals Industry." *Journal of Industrial Economics* 28: 113–133.

Schmalensee, R. 1978. "Entry Deterrence in the Ready-To-Eat Breakfast Cereal Industry." *Bell Journal of Economics* 9: 305–327.

Singh, N., and X. Vives. 1984. "Price and Quantity Competition in a Differentiated Duopoly." *Rand Journal of Economics* 15: 546–554.

Vives, X. 1985. "Efficiency of Bertrand and Cournot Equilibria with Product Differentiation." *Journal of Economic Theory* 36: 166–175.

Chapter 8

Concentration, Mergers, and Entry Barriers

> A prime reason for studying industrial organization is for understanding why concentration is observed very often.
> —Common statement

As we discussed in the introduction, the study of industrial organization is motivated mainly by the failure of the competitive market structure model, analyzed in chapter 4, to explain the commonly observed high concentration of firms in the same industry. Therefore, in this chapter we attempt to address the following questions:

1. Why do firms in some industries make pure profits?
2. When oligopolies make pure profits, how come entry of new firms does not always occur, thereby eliminating all pure profits?
3. What can explain mergers among firms in a given industry?
4. What is and what should be the regulators' attitudes towards concentrated industries? More precisely,
   (a) Should the regulator limit and control mergers among firms in the same industry?
   (b) Even if mergers do not occur, should the regulator attempt to control the degree of concentration in industries?

Section 8.1 (Concentration Measures) discusses and defines methods for measuring the degree of concentration in an industry. That is, we define indexes for measuring the distribution of market shares across firms in a given industry. Section 8.2 (Mergers) analyzes merger activities among firms and how those activities affect the industry's level of concentration. This section investigates the incentives of firms within various industries to merge with other firms in the same industry.

Section 8.3 (Entry Barriers) and section 8.4 (Entry Deterrence) provide a wide variety of explanations, classified into two related groups, for why entry does not always occur despite the fact that existing firms in the industry make strictly positive profits. By *entry barriers* we will refer to a long list of conditions that explain why entry does not occur. These conditions could be technological, such as economies of scale or sunk entry costs; legal, such as patent protection or exclusive rights given by other firms or regulators; or the result of market organization conditions, such as distribution channels, marketing networks, or consumer loyalty and goodwill. All these conditions are discussed in section 8.3.

By *entry deterrence* we will refer to strategic actions taken by incumbent firms when faced with a threat of actual entry into their industry. By *strategic actions* we mean actions that the incumbent firm would not find profitable to take in the absence of entry threats. Analyzing all possible such actions is the subject of section 8.4.

The distinction between entry-barrier arguments and entry-deterrence arguments is not without troubles, for several reasons: In many cases it is hard to find whether the conditions leading to no entry are external to the firms or are created by the incumbent firms. This in most cases makes antitrust litigation against monopoly firms very difficult because the monopoly firm can claim that the conditions that prevent entry are external to the firms. Furthermore, some of the conditions preventing entry can be augmented by the incumbent's behavior. More precisely, we will show that the existence of sunk (irreversible) costs may be sufficient to sustain one monopoly firm in the industry. Now, note that some sunk costs are external to the firms, such as entry taxes paid to the local authorities, initial market surveys required by the investors and so on. However, there are many sunk costs that are firm dependent. For example, the incumbent firm may spend on R&D to improve its product for the purpose of forcing R&D costs on the potential entrant. In addition, the incumbent may spend large sums of money on advertising for the purpose of forcing advertising sunk cost on the potential entrant.

In most of our analysis, sunk cost is either explicitly assumed or implicitly assumed to prevail as a consequence of having firms committing to certain capacity/output levels. Section 8.5 (Contestable Markets) introduces a *contestable market structure* which describes the behavior of an incumbent firm when potential entrants can enter without having to bear any sunk cost (generally called *hit-and-run entry*).

Finally, an appendix, section 8.6, provides an overview on how the Department of Justice and the Federal Trade Commission decide whether to challenge a merger and the corresponding operating guidelines. Appendix section 8.7 discusses the legal approach to entry-deterrence behavior.

8.1 Concentration Measures

So far, our discussion of industry concentration regarded concentrated industry as one where there are few firms, and each firm maintains a high market share. In this section we eliminate the vagueness behind the concept of concentration and propose precise measures of concentration. There are two reasons why there is a need for these precise measures. First, to be able to compare concentration among different industries in the same or different countries. The compared industries need not share anything in common, but a proper concentration measure should be able to compare concentration despite the fact that different industries have different numbers of firms and different distributions of market shares. Second, in case the regulating authority would like to intervene and to prevent a change in concentration of a certain industry, the regulator must specify a general measure by which it decides that a certain industry is concentrated. These measures can then be used by the legal system that arbitrates conflicts between the firms and the regulator about mergers.

What is a concentrated industry? Clearly, the most concentrated industry is a monopoly which sells 100% of the industry's output. When the number of firms is greater than one, there are two factors that influence concentration: (a) the number of firms in the industry, and (b) the distribution of output among the firms in the industry. Thus, a measure of concentration should be sensitive to both the distribution of the industry's output across firms as well as the number of firms in the industry.

Let N be the number of firms in the industry, let Q denote the aggregate industry-output level (aggregate amount sold to consumers), and let q_i denote the output of firm i, $i = 1, 2, \ldots, N$. Thus, $Q = \sum_{i=1}^{N} q_i$. Obviously, there may be a problem with this summation if the industry is composed of firms producing differentiated products. In other words, can we add red cars and purple cars? What about adding large cars and small cars? We ignore these aggregation problems, which come up in almost every empirical work in industrial organization and international trade.

Let $s_i \equiv (100 q_i/Q)$ denote the percentage of the industry's total output sold by firm i. We call s_i the *market share* of firm i. Observe that $0 \le s_i \le 100$ and that

$$\sum_{i=1}^{N} s_i = \frac{100 \sum_{i=1}^{N} q_i}{Q} = 100.$$

In what follows, we discuss two commonly used measures of concentration which, among other indicators, are used by the U.S. Federal Trade Commission for determining whether to approve a merger. For more measures and their interpretation, see discussions in Jacquemin 1987 and Tirole 1988, Chapter 5.

8.1.1 The four-firm concentration ratio

The four-firm concentration ratio was used for merger guidelines (see an appendix, section 8.6.2) purposes from 1968 to around 1982. It simply sums up the market shares of the four largest firms in the industry. Let us order all the firms in the industry (rename them) so that firm 1 would have the largest market share, firm 2 the second largest, and so on. That is, $s_1 \ge s_2 \ge s_3 \ge \ldots \ge s_N$. We define the four-firm concentration ratio by

$$I_4 \equiv \sum_{i=1}^{4} s_i. \tag{8.1}$$

Table 8.1 demonstrates the value of I_4 for four imaginary industries.

% share	s_1	s_2	s_3	s_4, s_5	$s_6 \ldots s_8$	s_9, s_{10}	I_4	I_{HH}
Industry 1	60	10	5	5	5	0	80	3,850
Industry 2	20	20	20	20	0	0	80	2,000
Industry 3	$\frac{100}{3}$	$\frac{100}{3}$	$\frac{100}{3}$	0	0	0	100	3,333
Industry 4	49	49	0.25	0.25	0.25	0.25	98.5	4,802

Table 8.1: Measures for industry's concentration (s_i in percentage)

You probably notice that there is something unsatisfactory about the four-firm concentration ratio. In industry 1, firm 1 has 60% of the market. Industry 2 has five firms, all have equal market shares of 20%. However, the four-firm concentration ratio yields $I_4 = 80\%$ for both industries. We conclude that since the four-firm measure is linear, it does not differentiate between different firm sizes as long as the largest four firms maintain "most" of the market shares. Comparing industries 3 and 4 demonstrates the same problem where an industry equally shared by three firms is measured to be more concentrated than an industry dominated by only two firms.

8.1.2 The Herfindahl-Hirshman index

The Herfindahl-Hirshman index (denoted by I_{HH}) is a convex function of firms' market shares, hence it is sensitive to unequal market shares. We define this measure to be the sum of the squares of the firms' market shares. Formally,

$$I_{HH} \equiv \sum_{i=1}^{N} (s_i)^2. \tag{8.2}$$

Table 8.1 shows that the I_{HH} for industry 1 is almost twice the I_{HH} for industry 2. This follows from the fact that squaring the market shares of the large firms increases this index to a large value for industries with significantly unequal market shares. Comparing industries 3 and 4 shows that while the I_4 measure indicates that industry 3 is more concentrated than industry 4, the I_{HH} measure indicates that industry 4 is more concentrated than industry 3. For this reason, the I_{HH} is found to be the preferred concentration measure for regulation purposes.

8.2 Mergers

The terms *mergers, takeovers, acquisitions,* and *integration* describe a situation where independently owned firms join under the same ownership. We will use the term *merger* to refer to any type of joining ownership and disregard the question of whether the merger is initiated by both firms, or whether one firm was taken over by another. Instead, we investigate the gains and incentives to merge and the consequences of mergers for the subsequent performance and productivity of the firms involved, for consumers' welfare, and for social welfare.

The Federal Trade Commission classifies mergers into three general categories:

Horizontal merger: This occurs when firms in the same industry, producing identical or similar products and selling in the same geographical market, merge.

Vertical merger: This occurs when a firm producing an intermediate good (or a factor of production) merges with a firm producing the final good that uses this intermediate good, or when two companies who have a potential buyer-seller relationship prior to a merger merge.

Conglomerate merger: This occurs when firms producing less related products merge under the same ownership. More precisely, conglomerate mergers are classified into three subclasses:

- *Product extension:* The acquiring and acquired firms are functionally related in production or distribution.
- *Market extension:* The firms produce the same products but sell them in different geographic markets.
- *Other conglomerate:* The firms are essentially unrelated in the products they produce and distribute.

Ravenscraft and Scherer (1987) provide a comprehensive study of merger activities in the United States and report four great merger "waves" that have marked American industrial history: one peaking in 1901; a milder one during the late 1920s; a third, with its peak in 1968; and the most recent one, a resurgence in the early 1980s. Looking at the types of mergers, we note that the data show a significant decline in horizontal and vertical activity and a rise in "pure" conglomerate mergers from the 1960s. The merger wave of the turn of the century was preponderantly horizontal. The wave of the 1920s saw extensive activity in the public-utility sector, in vertical and product-line extension, and in horizontal mergers that created oligopolies rather than monopolies. The wave of the 1960s was preponderantly conglomerate, reflecting a much more stringent antitrust policy against horizontal mergers.

Why do mergers occur? First, a merger may reduce market competition between the merged firms and other firms in the industry, thereby increasing the profit of the merged firms. However, note that section 8.8, exercise 2 demonstrates in a Cournot market structure that when there are more than two firms in the industry, the aggregate profit of the merged firms can be lower than the profit of the two firms separately before the merger occurs. Second, if the merger involves merging capital, assets, and other fixed factors of production, then the merged firms would be able to increase their size, possibly reduce cost, and thereby increase their market share, hence profit. Third, mergers and takeovers occur when there is a disparity of valuation judgments, given uncertainty about future business conditions: the buyer is for some reason more optimistic about the firm's future than the seller, or the buyer believes it can run the acquired entity more profitably as a part of this organization than the seller could by remaining independent. Fourth, those who control the acquiring entity seek the prestige and monetary rewards associated with managing a large corporate empire, whether or not the consolidation adds to the profits.

8.2.1 Horizontal merger

In subsection 6.1.3 we saw some theoretical basis for the presumption that under a Cournot market structure, a decrease in the number of firms in an industry (via, say, a merger) reduces social welfare. That is, we have shown that under a Cournot market structure, in the case of identical firms with no fixed costs, an increase in the number of firms increases the sum of consumer surplus and producers' profits despite the fact that profits decline.

However, there is still a question of whether a regulator should refuse to permit a merger to take place only on the basis of the associated sharp increase in concentration. The answer to this question is no! That is, in what follows, we construct an example where a merger of a high-cost firm with a low-cost firm increases overall welfare despite the increase in concentration (for a comprehensive analysis of mergers under the Cournot market structure for the case of n firms, see Salant, Switzer, and Reynolds 1983).

Consider the Cournot duopoly case, that of two firms producing a homogeneous product, analyzed in subsection 6.1.1 on page 98. Let the unit costs be $c_1 = 1$ and $c_2 = 4$ and the demand be $p = 10 - Q$. Equations (6.5), (6.6), and (6.7) imply that under the Cournot duopoly market structure: $q_1^c = 4$, $q_2^c = 1$, $p^c = 10 - (4+1) = 5$, $\pi_1^c = 16$, $\pi_2^c = 1$. Hence, in view of (3.3) (see subsection 3.2.3), the consumer surplus is $CS(5) = \frac{1}{2}(a - p^c)^2 = 25/2$. Hence, in view of (6.13), $W^c = CS(5) + \pi_1^c + \pi_2^c = 29.5$.

Now, allow a merger between the two firms. The new firm is a multiplant monopoly, and as shown in section 5.4, the newly merged firm would shut down plant number 2. Hence, the merged firm solves a simple single-plant monopoly problem analyzed in section 5.1, yielding an output level of $Q^m = 4.5$, and $p^m = 10 - 4.5 = 5.5$; hence, $\pi^m = (5.5 - 1)4.5 = 81/4$. Also, $CS(4.5) = \frac{1}{2}(10 - 5.5)^2 = 81/8$. Altogether, $W^m = CS(4.5) + \pi^m = 30.375$.

Comparing the premerger concentration level with the postmerger monopoly yields that

$$I_{HH}^c = (80\%)^2 + (20\%)^2 = 6,800 < 10,000 = (100\%)^2 = I_{HH}^m. \quad (8.3)$$

Observing that $W^m > W^c$, we can state the following:

Proposition 8.1 *Under a Cournot market structure, a merger among firms leading to an increase in concentration does not necessarily imply an overall welfare reduction.*

The intuition behind Proposition 8.1 is that when firms have different production costs, there exists a tradeoff between production efficiency and the degree of monopolization. In other words, a merger between a high-cost and a low-cost firm increases production efficiency since it eliminates the high-cost producer. However, the increase in concentration increases the market price and therefore reduces consumer welfare. Now, when the difference in production costs between the two firms is significant, the increase in production efficiency effect dominates the reduction in consumer welfare.

In view of the merger guidelines described in subsection 8.6.2 such a merger will not be approved, despite this example's demonstration that the merger would improve overall welfare. However, the reader is advised not to take this example too seriously for the following reason: It is possible that our methodology is wrong in the sense that we are making welfare judgments based on the Cournot market structure. Had the firms played Bertrand, the inefficient firm (firm 2) would not be producing in the duopoly case. In summary, conclusions about welfare that are based only on the Cournot market structure should be checked to determine whether they also hold under different market structures. Otherwise, such a welfare analysis is not robust.

The analysis of this subsection has a major shortcoming in that it is done without accounting for firms' size and therefore for the effects of changes in size associated with every merger. That is, under a Cournot market structure, when two firms with the same unit costs merge, their actual size merges into a single firm. Davidson and Deneckere (1984) develop a model that overcomes this shortcoming by introducing capacity to the analysis. In their model, when two firms with invested capacity merge, they merge with their entire stock of capacity, so the joint firm maintains a larger capacity level than each individual firm.

8.2.2 Vertical merger

A vertical merger is defined as a merger between a supplier (producer) of an intermediate good and a producer of a final good who uses this intermediate good as a factor of production. The common terminology used to describe these firms is to call the intermediate-good suppliers as *upstream firms*, and the final-good producers as *downstream firms*. Figure 8.1 illustrates an industry structure in which there are two upstream firms selling an input to two downstream firms. In Figure 8.1 the two input suppliers denoted by A and B sell identical inputs to both downstream firms denoted by 1 and 2. The left-hand side of Figure 8.1 shows the initial situation in which all firms are disjoint. The right-hand side illustrates the case in which the upstream firm A merges with downstream firm 1. We denote the merged firm by $A1$. There are several ways in which competition in the upstream and downstream markets could

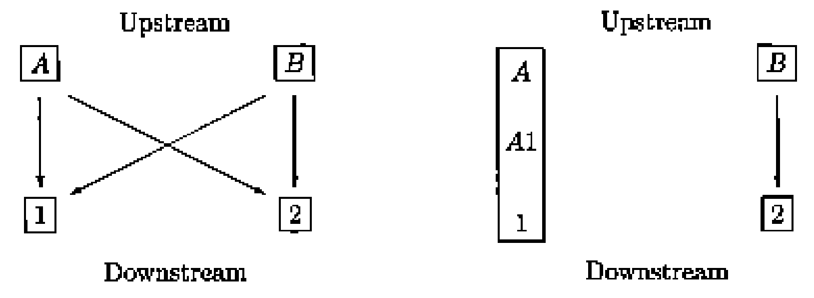

Figure 8.1: Upstream factor suppliers and downstream producers

be modeled (See for example Ordover, Saloner, and Salop 1990; Perry 1989; Salinger 1988; and Tirole 1988, Chap. 4). Clearly, if both the upstream and the downstream markets are characterized by a Bertrand price competition, then it is easy to show that profits of all firms are identically zero before and after vertical integration occurs. In order to solve this modeling problem we could assume that the downstream firms produce differentiated products (such as the Hotelling spatial competition analyzed in subsection 7.3.1) so that firms would make positive profits. Instead, we take an approach yielding similar results by assuming that the upstream market is characterized by a Bertrand price competition (section 6.3), whereas the downstream market is characterized by a Cournot quantity competition (section 6.1).

Downstream competition

We assume that the demand for the good marketed in the downstream market is given by the linear demand $p = \alpha - q_1 - q_2$, where $\alpha > 0$, and q_1 and q_2 are the output levels sold by downstream firms 1 and 2. Let the technology be such that one unit of input produces one unit of output, and denote by c_1 and c_2 the price of the input paid by firms 1 and 2, respectively. Hence, the firms' unit costs are given by c_1 and c_2, respectively. In section 6.1 we showed that under this demand and cost structure, a Cournot quantity competition yields the output and profit for each firm i given by

$$q_i = \frac{\alpha - 2c_i + c_j}{3} \quad \text{and} \quad \pi_i = \frac{(\alpha - 2c_i + c_j)^2}{9}. \tag{8.4}$$

Hence, the aggregate downstream production and price levels are

$$Q \equiv q_1 + q_2 = \frac{2\alpha - c_1 - c_2}{3} \quad \text{and} \quad p = \alpha - Q = \frac{\alpha + c_1 + c_2}{3}. \tag{8.5}$$

Upstream competition before the merger

The upstream firms A and B sell the intermediate product to the downstream firms 1 and 2. Since the two upstream firms engage in a Bertrand price competition, prices fall to their unit production cost which is assumed to be zero. Hence, $c_1 = c_2 = 0$, so the downstream firms have zero production costs. Thus, substituting into (8.4) yields

$$q_1 = q_2 = \frac{\alpha}{3}, \quad \pi_1 = \pi_2 = \frac{\alpha^2}{9} \quad \text{and} \quad \pi_A = \pi_B = 0. \tag{8.6}$$

Upstream and downstream merge

Suppose now that upstream firm A merges with downstream firm 1. We denote the merged firm by $A1$. Hence, the input cost of the merged firm $A1$ is zero. We assume that the merged firm $A1$ does not sell the intermediate good to firm 2; therefore, the upstream firm B is now a monopoly in the factor market and maximizes its profit by choosing the price for its intermediate product c_2 that equals the cost of production of downstream firm 2. Thus, the profit of upstream firm B is its price c_2 times the output level of downstream firm 2 given in (8.4). Formally, the upstream firm B chooses c_2 that solves

$$\max_{c_2} \pi_B = c_2 q_2 = \frac{c_2(\alpha - 2c_2 + c_1)}{3}. \tag{8.7}$$

The first-order condition yields $0 = \alpha - 4c_2 + c_1$, yielding that $c_2 = \alpha/4$. Clearly, the second-order condition is satisfied, so substituting $c_1 = 0$ and $c_2 = \alpha/4$ into (8.4) and (8.5) yields

$$q_1 = \frac{5\alpha}{12}, \quad q_2 = \frac{\alpha}{6}, \quad Q = \frac{7\alpha}{12}, \quad \text{and } p = \frac{5\alpha}{12}. \tag{8.8}$$

Hence, the profit of the two downstream firms is given by

$$\pi_{A1} = p q_{A1} = \frac{25\alpha^2}{144} \quad \text{and} \quad \pi_2 = (p - c_2) q_2 = \frac{\alpha^2}{36}. \tag{8.9}$$

Equation (8.8) yields the following proposition:

Proposition 8.2 *A merger between an upstream and a downstream firm increases the output level of the merged firm and reduces the output level of the downstream firm that does not merge.*

Proposition 8.2 is rather intuitive. The downstream firm that does not merge faces an increase in its input cost resulting from having to buy its input from a single monopoly firm B. Hence, the increase in firm 2's production cost and the reduction in firm 1's production cost would increase the output of firm 1 and reduce the output of firm 2.

We wish to investigate whether this vertical merger is profitable to the vertically merging firms. To see that, we need to compare the sum of profits of firms A and 1 prior to the merger to the profit of the merged firm $A1$. However, prior to the merger, firm A made zero profit, hence prior to merger their joint profit was $\pi_1 = \alpha^2/9$. Comparing this sum to π_{A1} in (8.9) implies that

Proposition 8.3

1. *The combined profit of the merging upstream and downstream firms increase after they merge.*
2. *A merger between the upstream and the downstream firms will not foreclose the market of the disjoint downstream firm but will only reduce its profit.*

Proposition 8.3 is important, since it is often argued that vertical mergers lead to a foreclosure of the disjoint downstream firms, which in our example means that firm B or firm 1 or both would go out of business. Note that this cannot happen in the present model since the upstream firm B will reduce the input price to prevent firm 2 from leaving the market (firm B sells only to firm 2 after the merger). Since vertical integration does not necessarily imply foreclosure, the FTC seems to be more forgiving to vertical mergers than to horizontal mergers. Moreover, many economists believe that vertical integration should be viewed as an increase in efficiency since most firms carry on several stages of production under a single plant anyway, with or without vertical integration. Thus, a firm is by definition a vertically merged entity and is believed to be an efficient form of organization.

Finally, the sum of the profits of the disjoint upstream firm B and downstream firm 2 is given by

$$\pi_B + \pi_2 = \frac{\alpha^2}{24} + \frac{\alpha^2}{36} = \frac{10\alpha^2}{144} < 0 + \frac{\alpha^2}{9}, \tag{8.10}$$

which is the sum of profits of firm B and 2 prior to the merger between firm A and firm 1. Thus, despite the fact that the profit of the nonmerging upstream firm B increases with the merger of firm A with firm 1, the decline in the profit of the nonmerging final-good-producer firm 2 is larger than the increase in π_B, which is caused by the sharp drop in market share of firm 2.

8.2.3 Horizontal merger among firms producing complementary products

It was Cournot who realized that horizontal merger need not increase the equilibrium price level when two firms producing complementary products merge. The reader is probably familiar with the definition and examples of complementary products. Examples include coffee and milk or sugar, audio receivers and speakers, video players and cassettes, cameras and film, computers and monitors, computers and software, cars and tires, transportation and hotel services, and more. The reader is referred to section 10.3 for further analyses of the economics of systems that are composed of complementary components.

In this subsection we analyze an industry where firms produce two complementary products. Economides and Salop 1992, provide a more extensive analysis of complementary systems by considering several producers of each product.

Demand for systems

Consider a market for computer systems. A computer system is defined as a combination of two complementary products called computers (denoted by X), and monitors (denoted by Y). We denote by p_X the price of one computer and by p_Y the price of a monitor. Therefore, since a system consists of one computer and one monitor, the price of a system is given by $p_S = p_X + p_Y$. Let Q denote the quantity of systems purchased by all consumers, and assume that the aggregate consumer demand is given by

$$Q = \alpha - p_S = \alpha - (p_X + p_Y) \quad \text{or} \quad p_S = p_X + p_Y = \alpha - Q, \quad \alpha > 0. \tag{8.11}$$

We denote by x the amount of computers sold to consumers and by y, the amount of monitors sold. Since the two components are perfect complements, $x = y = Q$.

Independently owned producing firms

Suppose that computers and monitors are produced by different firms whose strategic variables are prices, and suppose that production of either product is costless. Consider the problem solved by the computer firm (X-producer). For a given p_Y, firm X chooses p_X that solves

$$\max_{p_X} \pi_X = p_X X(p_X) = p_X[\alpha - (p_X + p_Y)]. \tag{8.12}$$

The first-order condition yields $0 = \frac{\partial \pi_X}{\partial p_X} = \alpha - 2p_X - p_Y$. Clearly, the second-order condition is satisfied. Hence, firm X's price-best-response

function to Y's price is $p_X = (\alpha - p_Y)/2$. Similarly, we can show that Y's price best response with respect to X's price is $p_Y = (\alpha - p_X)/2$. Altogether, when the complementary components are produced by independent firms, their prices, quantities, and firms' profit levels are given by

$$p_X = p_Y = \frac{\alpha}{3}, \quad Q = x = y = \alpha - (p_X + p_Y) = \frac{\alpha}{3} \quad \text{and} \quad \pi_X = \pi_Y = \frac{\alpha^2}{9}. \tag{8.13}$$

Monopoly producing all components

Now suppose that firms X and Y merge under a single ownership. Thus, computers are now sold as systems composed of a single monitor bundled with a single computer. Therefore, the monopoly systems producer chooses a system price p_S that solves

$$\max_{p_S} \pi_{XY} = p_S(\alpha - p_S)$$

yielding a first-order condition given by $0 = \frac{\partial \pi_{XY}}{\partial p_S} = \alpha - 2p_S$. Clearly, the second-order condition is satisfied. Hence, the price of a system under monopoly and the monopoly's profit are given by

$$p_S^M = \frac{\alpha}{2}, \quad Q^M = x^M = y^M = \alpha - p_S^M = \frac{\alpha}{2}, \quad \text{and} \quad \pi_{XY}^M = \frac{\alpha^2}{4}. \tag{8.14}$$

We conclude the discussion on mergers with the following proposition, which follows from the comparison of (8.13) and (8.14).

Proposition 8.4 *A merger into a single monopoly firm between firms producing complementary products would*

1. *reduce the price of systems (i.e., $p_S^M < p^S = p_X + p_Y$);*
2. *increase the number of systems sold (i.e., $Q^M > Q$); and*
3. *increase the sum of profits of the two firms (i.e., $\pi_{XY}^M > \pi_X + \pi_Y$).*

The significance of Proposition 8.4 is that a merger between two firms producing complementary products can increase social welfare, since consumers face lower prices, and firms gain a higher profit. The intuition behind Proposition 8.4 is as follows. Given that the two components are perfect complements, a rise in the price of one component reduces the demand for both components. Under price competition among independent component-producing firms, each firm overprices its component since each firm is affected by the reduced demand for its component and not the entire system. Thus, the negative externality on the other firm's demand is not internalized. However, when the firms merge, the joint ownership takes into consideration how the demand for both components is affected by an increase in the price of one component, and the negative demand externality is internalized.

We conclude our discussion of merger of firms producing complementary products with two remarks: First, Sonnenschein (1968) has shown that the Nash equilibrium where firms compete in price and sell perfect complements is isomorphic to the case where firms compete in quantity and sell perfect substitutes. One simply has to interchange the roles of price in the network case with the industry quantity in the perfect substitutes case. For example, Proposition 8.4 can be reinterpreted as showing that under quantity competition among firms selling perfect substitutes, a merger to monopoly would (1) reduce the aggregate quantity produced, (2) increase the price, and (3) result in strictly larger industry profits. Second, Gaudet and Salant (1992) show that the merger of firms producing complements and setting prices may be unprofitable if some members of the industry are not parties to the merger. Given Sonnenschein's observation, their result implies that mergers to less than monopoly may also be unprofitable if firms produce perfect substitutes and engage in Cournot competition, a point first noted in Salant, Switzer, and Reynolds 1983.

8.3 Entry Barriers

Why do we frequently observe that firms do not enter an industry despite the fact that the existing firms in the industry make above normal profits? In this section we investigate the following question: If oligopolies make pure profits, why does free entry not occur until competition brings down the price so that existing firms will no longer make above normal profits? Barriers to entry are considered an important structural characteristic of an industry. The competitiveness and the performance of an industry are generally assumed to be strongly influenced by its entry conditions.

There can be many reasons why entry may not occur. The primary explanation for entry barriers is the existence of entry cost. Bain's pioneering work (1956) specified three sources of entry barriers: absolute cost advantages of incumbent firms, economies of scale, and product-differentiation advantages of incumbent firms, such as reputation and goodwill. In addition politicians and all levels of governments may explicitly or implicitly support the existing firms (and the existing firms may in return support and contribute to the campaigns of politicians). Maintaining such connections seems impossible for new investors. Other

reasons include the learning experience possessed by the existing firms, consumers' loyalty to brands already consumed, and availability of financing (banks are less eager to lend to new investors) (see also Geroski, Gilbert, and Jacquemin 1990).

In this section we briefly discuss entry barriers. As we mentioned earlier, we regard entry barriers as the conditions that are not controlled by the incumbent firms that explain why entry does not occur. Section 8.4 below will address issues of entry deterrence, which we regard as the strategic actions taken by incumbent firms when facing the entry into an industry of potential competitors. Subsection 8.3.1 demonstrates a technological explanation for entry barriers and shows how the degree of concentration is related to the fixed production costs. Subsection 8.3.2 demonstrates the role that the existence of sunk costs play in generating the conditions for entry barriers.

8.3.1 Concentration and fixed costs in a noncompetitive market structure: an example

Let us demonstrate the relationship between fixed costs and concentration by means of an example. Consider the monopolistic competition in the differentiated-products environment analyzed in section 7.2 on page 143. In that environment, firms have to bear a fixed cost, implying that in equilibrium there will be entry of a finite number of firms. More precisely, recall from Proposition 7.5 on page 147 that the number of firms is $N^{mc} = L/(2F)$, where L is the economy's resource endowment and F is the fixed cost of each firm, $(L > 2F)$. Hence, the industry described in Section 7.2 yields a concentration level given by

$$I_{HH} = N^{mc}\left(\frac{100}{N^{mc}}\right)^2 = \frac{100^2}{N^{mc}} = \frac{2F}{L}10{,}000. \tag{8.15}$$

Consequently, in a monopolistic-competition environment, the I_{HH} concentration ratio increases with the fixed cost. A similar calculation can be performed in a Cournot market structure, where firms have fixed cost, and therefore, only a finite number of firms would enter.

For the case of an industry producing a homogeneous product, von Weizsäcker (1980) demonstrates that if production technologies exhibit increasing returns to scale at low output levels (U-shaped average-cost functions), then the equilibrium number of firms is larger than the social optimum.

8.3.2 Sunk costs generate entry barriers

By *sunk costs* we mean costs that cannot be reversed or for which the investment associated with paying them cannot be converted to other causes, or resold in order to recapture part of the investment cost. Examples include legal (lawyers') fees and taxes that an entering firm must bear prior to the actual entry. If after paying this cost a firm reverses its decision to enter, the firm cannot recover these fees. Other forms of sunk costs include market surveys (almost always mandated by the investors), advertising costs, and expenditures on nontransportable, nonconvertible plant and equipment, such as the site preparation work for any plant.

Following Stiglitz (1987), we now demonstrate how in a market for a homogeneous product, the existence of even small sunk costs can serve as an entry barrier so that entry will not occur even if the incumbent continues to make a monopoly profit. There are two firms, A and B, both capable of producing an identical product with identical constant marginal costs. Firm B is the potential entrant. If firm B enters, it has to sink ϵ dollars into the process. Firm A is the incumbent monopoly firm earning a profit of $\pi^A = \pi^M - \epsilon$, where π^M denotes the monopoly's profit level, not including the entry cost it has already sunk in. This extensive-form game is illustrated in Figure 8.2.

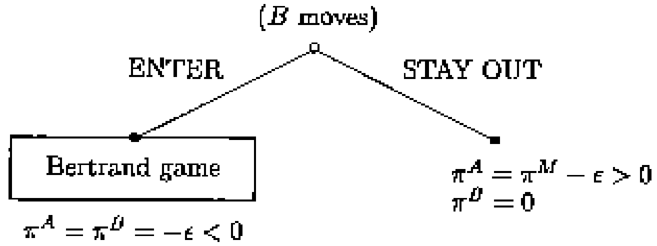

Figure 8.2: Sunk costs and entry barriers

In the game illustrated in Figure 8.2 the potential entrant (firm B) moves first by choosing whether to enter or not. In case firm B chooses not to enter, it saves the entry cost ϵ and therefore earns zero profit. In this case, firm A remains a monopoly and makes the monopoly profit less the entry cost it sunk earlier. In contrast, if firm B enters, the firms are assumed to set their prices simultaneously, yielding a Bertrand equilibrium (see Definition 6.2 on page 108) where price equals marginal cost. In this case, both firms make a loss equal to the sunk cost. It is straightforward to establish the following proposition:

Proposition 8.5 *For any level of sunk entry cost satisfying* $0 < \epsilon < \pi^M$, *there exists a unique subgame perfect equilibrium where firm A is a monopoly earning* $\pi^A = \pi^M - \epsilon$ *and firm B stays out.*

That is, in a SPE, the entrant foresees that after entry occurs (the second stage of the game), the incumbent will switch from being a monopoly to being in an aggressive price competition and leading the marginal-cost pricing. Hence, in the first stage the potential entrant will choose not to enter since staying out yields zero profit.

Proposition 8.5 is rather disturbing because it means that entry will never occur as long as there are some (even infinitesimal) sunk costs associated with entry. However, the reader should notice that Proposition 8.5 applies only to homogeneous products. In fact, under these circumstances, it is likely that the entrant will engage itself in further investments (higher sunk costs) in order to develop a differentiated brand, in which case price competition need not yield zero or negative profits. However, Proposition 8.5 makes a point by stating that even small sunk cost can create all the conditions for entry barriers. In fact, the incumbent does not need to do anything to deter this entry and simply continues producing the monopoly output level. Proposition 8.5 highlights the role ex-post competition plays in creating entry barriers. What generates the entry barriers even for negligible sunk cost is the intensity of the postentry price competition. Had we assumed that the firms play Cournot after entry occurs, low sunk cost would not generate entry barriers. Assuming Bertrand price competition generates the postentry intense competition that makes entry unprofitable for even low entry costs.

We conclude this analysis by considering a situation where a firm could receive an amount of $\phi > 0$ upon exit. For example, if $\phi \leq \epsilon$, then we can view ϕ as the amount of its original expenditure the firm can recover upon exit. Figure 8.3 illustrates the modified game.

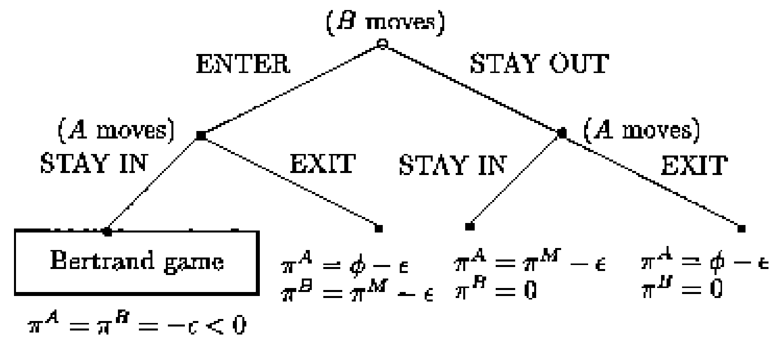

Figure 8.3: Sunk-cost entry barriers with partial cost recovery

In Figure 8.3 we added an additional stage enabling the incumbent (firm A) to exit after firm B makes its choice whether to enter or not. In fact, firm A's exit choice could have been included in the original game (Figure 8.2); however, in that game A's exit action was clearly dominated by other actions and was therefore ignored.

We now look for a subgame perfect equilibrium for this game. The subgame on the right (B does not enter) has a unique Nash equilibrium where the incumbent stays in and firm B earns zero profits. The subgame on the left (starting with the node where the incumbent makes a move) has a unique Nash equilibrium where the incumbent exits the industry and collects a profit of $\phi - \epsilon > -\epsilon$ it would collect if it stays in. In this case, the entrant becomes a monopoly. Therefore,

Proposition 8.6 *There exists a unique SPE for the game described in Figure 8.3, where firm B enters and firm A (incumbent) exits the industry. Formally, ((EXIT, STAY-IN), ENTER) is a unique SPE.*

Proposition 8.6 states that the market for this product will remain dominated by a monopoly market structure despite the fact that entry (and exit) occur. One monopoly replaces another monopoly. Hence, from the consumers' point of view, this particular market will be regarded as one that has substantial entry barriers.

Finally, we can further modify the game described in Figure 8.3 by adding an initial stage in which firm A makes a choice whether to enter the game and become the incumbent firm. Clearly, if firm A would be able to recover only part of its sunk cost ($\phi < \epsilon$), then it would not enter at all, and no other firm would ever find it profitable to enter. This result makes our argument even stronger since in this case the entry barriers are so strong that entry is not profitable to any firm, because any entering firm would have to exit when another firm enters.

8.4 Entry Deterrence

We now turn to the strategic approach for explaining entry barriers. We assume that initially there is one firm, called the *incumbent* or the established firm, that is a monopoly in a certain market. In the second stage, we assume that another firm, called the *potential entrant* is entering the market if entry results in above normal profit.

Modifying Bain's classifications of entry deterrence, we use the following terminology:

Blockaded entry: The incumbent is not threatened by entry; no firm would find it profitable to enter, even if the incumbent produces the monopoly output level.

Deterred entry: The incumbent modifies its behavior (say, by lowering price or expanding capacity) in order to deter entry; if prices are lowered, then we say that the incumbent exercises *limit pricing.*

Accommodated entry: Entry occurs, and the incumbent firm modifies its action to take into account of entry that occurs.

Thus, blockaded entry corresponds to what we called entry barrier in section 8.3, where we discussed several conditions yielding entry barriers other than the behavior of incumbent firms. In contrast, we refer to entry deterrence and entry accommodation as actions taken by incumbent firms when faced with a threat of entry.

Earlier authors held that an incumbent firm may be able to deter entry by overproducing and selling at lower prices prior to the date at which entry is expected. These types of models relied on the *Bain-Sylos postulate*, under which the prospective entrant was assumed to believe that the established firm would maintain the same output after entry that it did before entry. Then the established firm naturally acquired a leadership role as described in the Leader-Follower model (section 6.2). In addition, some of the earlier models assumed that entrants have to sink (output-independent) costs in order to begin their operation, whereas incumbents do not.

Presently, most economists disregard these arguments for the following reasons: First, note that this cost asymmetry could be reversed, considering the fact that established firms may have to pay some costs that the entrant does not have to bear. For example, established firms may operate according to long-term contracts. Most notably, wage contractees and unions are hard to negotiate with, and the downward adjustment of wages needed to meet the competition with the entrant would invoke tough resistance from workers and unions. Yet in some instances, the potential entrant is free to choose workers and can decide on wages without having any prior obligation. The same argument holds for subcontracting and binding contracts with suppliers of raw material and parts. In addition, assuming asymmetric cost structure turns the problem of entry deterrence into an ad hoc problem since there always exists a level of entry cost that would prevent firms from entering the market. Moreover, even if the above asymmetry holds true in reality, it is likely that in the long run the entrant would be able to collect a high enough (duopoly) profit to more than cover the entry cost. In addition, banks observing that the entering firm would make such a profit would be willing to lend the entrant the entry cost since the firm would be able to pay back the loan (and interest) with its future profits.

Second, Friedman (1979) and Dixit (1980) question the validity of the Bain-Sylos postulate by raising some doubt regarding the logic behind the above entry-deterrence argument. They point out that the preentry price choice (or quantity in our case) of the established firm is irrelevant for the entry decision of the potential entrant. The only thing that should matter to the potential entrant is what the postentry market structure would be. After entry occurs and the entry cost is already paid, there is no reason to assume that the firms would play the Leader-Follower game. It would be more reasonable to assume that the firms would play Cournot or Bertrand where the firms have equal power and knowledge. Now, given that the entrant knows that the market structure would change after entry occurs, all the first-period entry-deterrence strategies (limit pricing) or overproduction are irrelevant to the postentry profits collected by all firms. Third, in modeling entry deterrence it is not clear why one firm gets to be the first to choose and commit itself to a certain production level, thereby obtaining what is commonly called a *first-mover advantage*?

The approach to modeling entry deterrence based on the Bain-Sylos postulate is given in subsection 8.4.1, where we sketch an analog to Spence (1977) and demonstrate that entry can be deterred if an incumbent firm builds an *irreversible capacity* prior to the period when entry is allowed, so that a potential entrant faces a saturated market if it decides to enter. Subsection 8.4.2 relaxes the Bain-Sylos postulate and assumes that the incumbent is aware of the possibility that the entrant may find it profitable to alter its actions after entry occurs.

Subsection 8.4.3 (Investment in capital replacement) introduces a dynamic entry-deterrence model showing how in the face of entry threats an incumbent with depreciating capital is forced to invest more frequently than what is needed to simply replace depreciated capital. Subsection 8.4.4 (Judo economics) focuses on the strategic choices of a potential entrant when an incumbent firm may find it more profitable to allow a small-scale entry rather than fighting it. Subsection 8.4.5 (Credible spatial preemption) analyzes an incumbent differentiated-good producer facing entry in one of its markets. We conclude our analysis of entry barriers with subsection 8.4.6, where we demonstrate that limit pricing can serve as an entry-deterring strategy when the entrant does not know the production cost of the incumbent.

8.4.1 Capacity commitment under the Bain-Sylos postulate

Earlier models analyzing entry deterrence adopted the Bain-Sylos postulate, under which the prospective entrant was assumed to believe that the incumbent firm would maintain the same output after entry as before. Spence (1977) explicitly distinguishes between capacity and quantity produced. In his model, the quantity produced is constrained by

the amount of capacity firm 1 invests in the first period. Thus, as long as entry does not occur, the capacity is underutilized. However, in the event of a threat of entry, the incumbent can expand its output level and use all the capacity, thereby reducing the price to the level that makes entry unprofitable. In this subsection we refrain from making the distinction between capacity and output level and concentrate on analyzing how the incumbent determines how much capital to invest under the threat of entry.

Consider the two-period Leader-Follower game described in section 6.2. However, instead of assuming that firms decide how much to produce, let us assume that the firms' actions are confined to how much capacity (or capital) to accumulate (invest). Although this distinction is only a semantic one, it makes our story somewhat more convincing since capacity bears the sense of irreversibility (one is unable to discard it and to collect the costs already paid), thereby making capacity accumulation a credible strategic variable. Thus, in period 1, firm 1 has to choose its capacity-output investment, $k_1 \in [0, \infty)$; in period 2, firm 2 chooses whether to enter (choosing $k_2 > 0$) or to stay out ($k_2 = 0$).

We assume that the firms are identical in all respects, except that the potential entrant (firm 2) has to pay an entry cost. Such costs include an investment in new equipment, payments to lobbyists for facilitating the industry's control regulations, and so on. We denote the entry cost by $E, E \geq 0$. To completely describe the game, we define the profit of the firms (collected at the end of the second period) to be:

$$\pi_1(k_1, k_2) \equiv k_1(1 - k_1 - k_2) \quad \text{and} \tag{8.16}$$

$$\pi_2(k_1, k_2) \equiv \begin{cases} k_2(1 - k_1 - k_2) - E & \text{if entry occurs} \\ 0 & \text{otherwise.} \end{cases}$$

We solve this game backwards by first analyzing the last period, given the action taken in the preceding period.

The second period

In the second period, firm 2 takes $k_1 = \bar{k}_1$ as given and chooses k_2 to maximize its profit given in (8.16). There can be two cases: Firm 2 enters and pays the entry cost E, or it does not enter. Suppose for a moment that it enters. Then, firm 2 chooses k_2 to satisfy

$$0 = \frac{\partial \pi_2(\bar{k}_1, k_2)}{\partial k_2} = 1 - 2k_2 - \bar{k}_1, \quad \text{hence } k_2 = \frac{1 - \bar{k}_1}{2}. \tag{8.17}$$

Substituting into the profit function of firm 2 (8.16), we have it that if firm 2 enters, then

$$\pi_2 = \frac{1-\bar{k}_1}{2}\left(1-\bar{k}_1-\frac{1-\bar{k}_1}{2}\right)-E,$$

which is greater than zero if and only if $\bar{k}_1 < 1-2\sqrt{E}$.

We summarize the analysis for the second period by the best-response function of firm 2:

$$k_2 = R_2(\bar{k}_1, E) = \begin{cases} \frac{1-\bar{k}_1}{2} & \text{if } \bar{k}_1 < 1-2\sqrt{E} \\ 0 & \text{otherwise.} \end{cases} \tag{8.18}$$

The first period

In the first period firm 1 has to set k_1 knowing how it will affect the capacity choice of firm 2. That is, firm 1 calculates (8.18). Firm 1 also knows that the best-response function of firm 2 is discontinuous when it sets $k_1 = 1-2\sqrt{E}$. Thus, firm 1 would take into consideration that small changes in its capacity around $k_1 = 1-2\sqrt{E}$ may induce firm 2 to alter its entry decision.

With this discontinuity in mind, our search for the profit-maximizing strategy for firm 1 would involve comparing the profit of firm 1 when firm 2 enters (the leader's profit level, denoted by π_1^s) with the profit of firm 1 when firm 2 does not enter (the monopoly profit level, denoted by π_1^m). Formally, these profit levels are given by

$$\pi_1^s = k_1\left(1-k_1-\frac{1-k_1}{2}\right) = k_1\left(\frac{1-k_1}{2}\right) \text{ and } \pi_1^m = k_1(1-k_1). \tag{8.19}$$

Thus, for a given k_1, the monopoly's profit level is twice the leader's profit levels in the present formulation. The two profit functions are drawn in Figure 8.4. In Figure 8.4, the upper bell-shaped curves are the incumbent's monopoly profit (when entry does not occur). The lower bell-shaped curves are the leader's profit level (when entry occurs). Also, the entry-deterring capacity level of firm 1 (given by $k_1 = 1-2\sqrt{E}$) is marked by the vertical solid line with a rightward pointing arrow, indicating that for $k_1 \geq 1-2\sqrt{E}$ firm 1 is a monopoly (hence the upper bell-shaped profit curves apply).

Figure 8.4 is divided into three parts, indicating how firm 1 reacts for different levels of firm 2's entry cost.

1. *Blockaded entry:* This case is not displayed in Figure 8.4 but applies when $1-2\sqrt{E} < 1/2$ (high entry cost). In this case, choosing the monopoly capacity level is sufficient for deterring entry. That

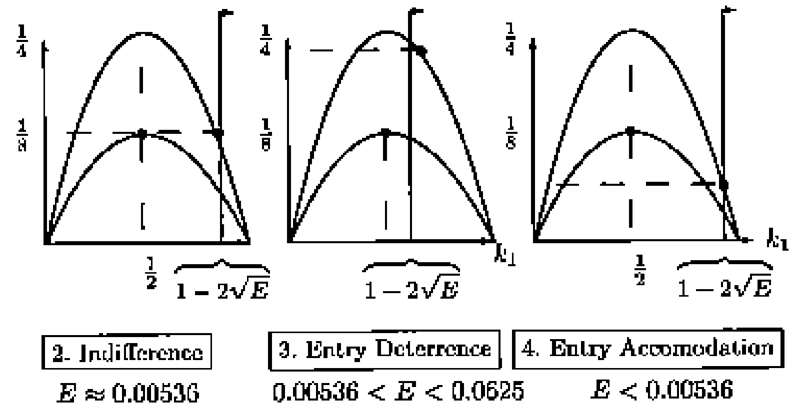

Figure 8.4: Incumbent's profit levels and capacity choices for different levels of entry cost.

is, when the entry cost is high, firm 2 will not enter when firm 1 plays its monopoly capacity level. Thus, substituting $k_2 = 0$ into (8.16), firm 1 chooses k_1 to maximize its monopoly profit. The first-order condition satisfies

$$0 = \frac{\partial \pi_1(k_1, 0)}{\partial k_1} = 1-2k_1.$$

Hence, $k_1 = 1/2$. Now, for having this output level deterring the entry of firm 2, (8.18) implies that E has to satisfy $k_1 = 1/2 \geq 1-2\sqrt{E}$, implying that $E \geq 1/16 = 0.0625$.

2. *Indifference between deterrence and accommodation:* We need to find the magnitude of the entry cost parameter E that would make firm 1 indifferent to whether it chooses to set $k_1 = 1-2\sqrt{E}$ to deter entry or to set $k_1 = 1/2$ and accommodate entry (it is clear that $k_1 = 1/2$ is the profit-maximizing capacity under monopoly as well as when entry occurs, since both profit curves peak at $k_1 = 1/2$).

   Thus, we need to compare the leader's profit level under $k_1 = 1/2$ when entry occurs given in (8.19) to the profit level when firm 1 deters entry by setting $k_1 = 1-2\sqrt{E}$, denoted by π_1^d. Hence,

$$\pi_1^d = (1-2\sqrt{E})2\sqrt{E} = \frac{1}{8} = \pi_1^s. \tag{8.20}$$

Thus, we need to solve $4E - 2\sqrt{E} + 1/8 = 0$, yielding

$$\sqrt{E} = \frac{16 - \sqrt{16^2 - 4 \times 32}}{64} \approx 0.07322,$$

implying that $E \approx 0.00536$.

3. *Entry deterrence:* From case 2 and case 3 of Figure 8.4, we have it that entry deterrence is profitable for firm 1 when the entry cost is at an intermediate level. That is, when $0.00536 < E < 0.0625$.

4. *Entry accommodation:* When the entry cost is very low, firm 1 would have to increase k_1 to a very high level in order to deter entry. Case 4 of Figure 8.4 shows that if $E < 0.00536$, deterring entry is not profitable, and that entry accommodation yields a higher profit level for firm 1.

8.4.2 Relaxing the Bain-Sylos postulate

So far, our analysis has relied on the Bain-Sylos postulate, under which the potential entrant is assumed to believe that the incumbent firm will maintain the same action after entry as before. Thus, under this postulate, the potential entrant is assumed to believe that upon entry, the incumbent will utilize its entire capacity to produce the highest possible output level in order to make entry unprofitable for the entrant. In this section, following Dixit 1980, we demonstrate that such an assumption is inconsistent with a strategic behavior under a subgame perfect equilibrium (Definition 2.10 on page 27). More precisely, we demonstrate that under a subgame perfect equilibrium, the incumbent firm will not find it profitable to utilize its entire capacity even when entry does occur. Thus, a rational potential entrant should be able to predict that a profit-maximizing incumbent will not find it profitable to utilize all its entire capacity. Therefore, we show that in a subgame perfect equilibrium, a profit-maximizing incumbent will not invest in excess capacity for the purpose of entry deterrence. In other words, overaccumulation of capacity will not occur.

Consider the following two-stage game. In the first stage firm 1 (incumbent) chooses a capacity level $\bar{k}$ that would enable firm 1 to produce without cost $q_1 \le \bar{k}$ units of output in the second stage of the game. If, however, the incumbent chooses to expand capacity beyond $\bar{k}$ in the second stage, then the incumbent incurs a unit cost of c per each unit of output exceeding $\bar{k}$. Figure 8.5 illustrates the marginal-cost function facing the incumbent in the second stage of the game.

Intuitively speaking, we can say that any amount produced above the firm's capacity will require special inputs that are costly to the firm

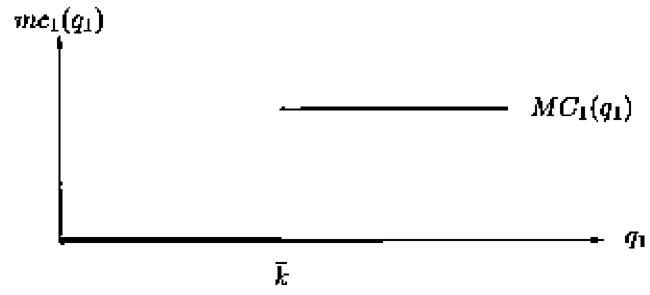

Figure 8.5: Capacity accumulation and marginal cost

when purchased at the last minute. Finally, to make our argument even stronger, we assume that capacity accumulation in the first stage is costless to the incumbent.

The entrant is assumed to make its entry decision in the second stage of the game. More precisely, in the second stage, both firms jointly choose their output levels and play a Cournot game (see section 6.1). We assume that firm 2 does not have any capacity and thus bears a unit cost of c, which is the same unit cost of the incumbent for producing beyond its capacity. If firm 2 chooses $q_2 = 0$, we say that entry does not occur. The game is illustrated in Figure 8.6.

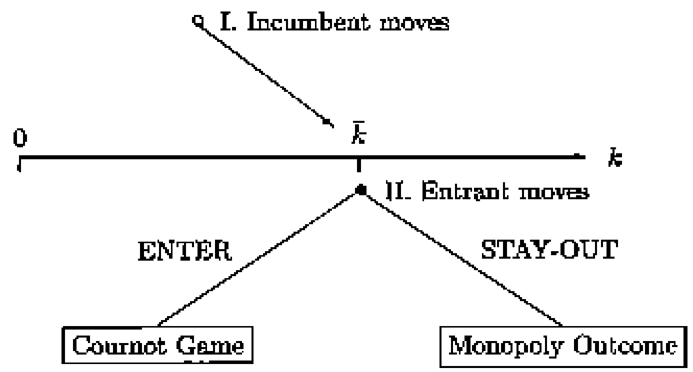

Figure 8.6: Relaxing the Bain-Sylos postulate

We now turn to the second stage after firm 1 has chosen its irrevocable capacity level given by $\bar{k}$. Figure 8.7 illustrates Cournot output best-response functions for three given choices of $\bar{k}$ by firm 1 in the first stage. The best-response functions drawn in Figure 8.7 are derived in the same way as that under the conventional Cournot market structure,

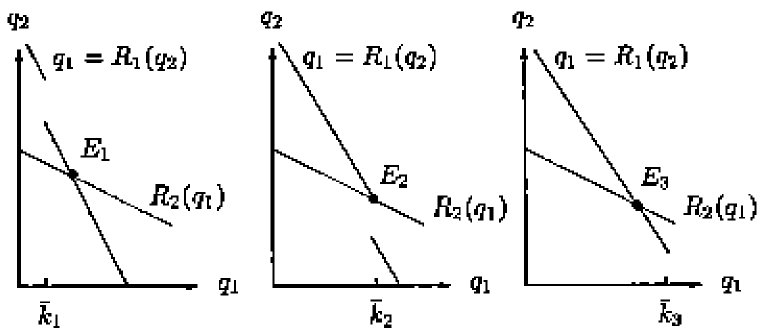

Figure 8.7: Best-response functions with fixed capacity: *Left:* low capacity; *Middle:* medium capacity; *Right:* High capacity

(see section 6.1, in particular Figure 6.1 on page 100). The only difference between the present case and the conventional Cournot case is that the incumbent's best-response function is discontinuous at an output level $q_1 = \bar{k}$, reflecting a jump in the unit cost associated with a production level beyond the firm's planned capacity.

Figure 8.7 has three drawings associated with having the incumbent investing in low, medium, and high capacity in the first period, thereby determining three Cournot equilibria denoted by E_1, E_2, and E_3, respectively. The most important observation coming from Figure 8.7 is that the equilibrium marked by E_2 is identical to the equilibrium marked by E_3, despite the fact that E_3 is associated with a higher capacity level invested in by firm 1 in the first stage. This proves our main proposition.

Proposition 8.7 *The incumbent cannot deter entry by investing in a large capacity. More generally, investing in excess capacity cannot serve as a tool for deterring entry.*

More interestingly, in our example the first-period cost of capital (capacity) is zero. Despite that cost, firm 1 cannot benefit by investing in $\bar{k}_3$ units of capital since after entry occurs, the incumbent's best response is to produce $q_1 = \bar{k}_2 < \bar{k}_3$. That is, the entrant can calculate that in the subgame of the second period, in a Cournot equilibrium, firm 1 will limit its production for the same reason that any firm limits its production under a Cournot market structure (preventing a price fall) and will therefore enter.

The main message conveyed by Proposition 8.7 is that investing in excess capital cannot provide the incumbent with a credible threat by which convince the potential entrant that entry is unprofitable. Thus, the Bain-Sylos postulate imposes an unrealistic belief on the potential

entrant, namely, the belief that the incumbent will utilize all its capacity after entry occurs, despite the fact that this action does not maximize the incumbent's profit.

8.4.3 Investment in capital replacement

So far we have assumed that investment in capacity is sufficient to produce output for the desired period of production. However, plants and equipment are of finite duration. If investment in capital deters entry, then entry is unavoidable if capital depreciates and the incumbent does not invest in capital replacement. In what follows we construct a discrete-time version of the analysis found in Eaton and Lipsey 1980, and investigate how the threat of entry affects the frequency of capital investment by an incumbent firm in the presence of depreciating capital.

Consider an industry with two firms, firm 1 (incumbent) and firm 2 (potential entrant). Each firm can produce only if it has capital. The profit of each firm is as follows. If only firm 1 has capital in a certain period, then firm 1 earns a monopoly profit, given by H, in this particular period. If both firms have capital in a certain period, then each earns a duopoly profit, given by L, in this period.

Suppose that in each period t, $t = 0, 1, 2, \ldots$, each firm can invest $\$F$ in capital with finite duration, and that during the time period(s) of this capital, the firm can produce any amount of a homogeneous product. We denote the action taken by firm i in period t by a_t^i where $a_t^i \in \{INV, NI\}$ (Invest or Not Invest). Figure 8.8 illustrates the time path and the timing of actions taken by the two firms.

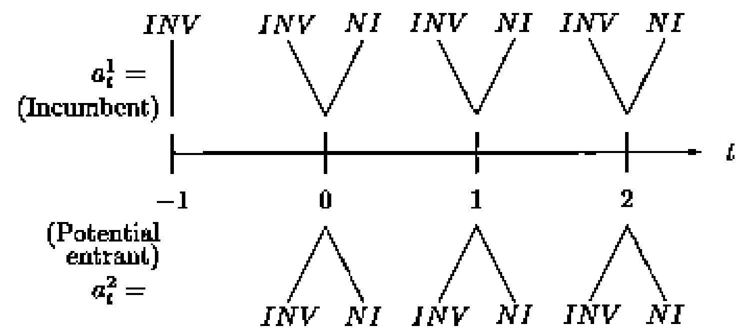

Figure 8.8: Capital replacement and entry deterrence

In Figure 8.8, firm 1 (the incumbent) is assumed to invest in capital in period $t = -1$, and then the game starts at $t = 0$, where both firms can invest in capital in any period $t = 0, 1, 2, 3 \ldots$. We make the following assumption on the parameters of the model:

ASSUMPTION 8.1

1. *Capital lasts for exactly two periods only. At the end of the second period the capital completely disintegrates, cannot be resold, and has a scrap value of zero.*
2. *The duopoly profit is insufficient to sustain two firms in the industry, whereas the monopoly profit level is sufficiently high relative to the capital investment cost. Formally, $2L < F < H$.*

Assumption 8.1 implies that if firm 1 invests in capital in period t, then using this capital the firm can produce in periods t and $t+1$ where the capital completely disintegrates at the end of the second period after production is undertaken.

The game proceeds as follows: In period 0, if firm 2 (potential entrant) invests in capital, then each firm earns L in period 0. If firm 2 does not enter (does not invest in capital), then firm 1 earns H in period 0.

Let $0 < \rho < 1$ denote the discount parameter, and assume that each firm maximizes the sum of its discounted profit given by

$$\Pi^i \equiv \sum_{t=0}^{\infty} \rho^t \left[R_t^i - C_t^i \right],$$

where $R_t^i = H$ if only firm i has capital in period t, and $R_t^i = L$ if both firms have capital in period t; and $C_t^i = F$ if firm i invests in capital in period t, and $C_t^i = 0$ if no investment is undertaken by firm i in period t.

Our purpose is to demonstrate the following:

Proposition 8.8 *Under Assumption 8.1:*

1. *If firm 2 is not allowed to enter, then firm 1 invests in capital in odd periods only. That is,*

$$a_t^1 = \begin{cases} INV & \text{for } t = 1, 3, 5, \ldots \\ NI & \text{for } t = 0, 2, 4, 6, \ldots. \end{cases}$$

2. *If firm 2 is allowed to enter, and if the time discount parameter is sufficiently small and satisfies*

$$\frac{F}{H} < \rho < \sqrt{\frac{F-L}{H-L}}, \tag{8.21}$$

then the following strategies constitute a subgame perfect equilibrium (Definition 2.10) for this game:

$$a_t^i = \begin{cases} INV & \text{if } a_{t-1}^j = NI \\ NI & \text{otherwise} \end{cases} \quad i,j = 1,2,\ i \neq j,\ t = 0,1,2\ldots. \tag{8.22}$$

Hence, in this equilibrium entry is deterred by having firm 1 (incumbent) investing in each period.

Proof. We look at equilibrium strategies where firm 1 invests in every t and firm 2 does not invest.

First, observe that since firm 1 invests at t and still has capacity at $t+1$, if firm 2 deviates and invests at t, it will earn $L-F$ at t, $L-F$ at $t+1$, and $H-F$ in each period thereafter. Firm 2 will not deviate, i.e., will not invest at t, if

$$\Pi^2 = (1+\rho)(L-F) + \rho^2 \frac{H-F}{1-\rho} < 0, \quad \text{or } \rho^2 < \frac{F-L}{H-L}.$$

Secondly, if firm 1 deviates, i.e., ceases investing at $t-1$, then it has no capacity at t and firm 2 will earn $H-F$ at t. Hence, firm 2 will enter.

Thirdly, if firm 1 stops investing at $t-1$, it will earn a profit of H in period $t-1$ and zero thereafter. Thus, in order for having firm 1 engaging in continuous investment, it must be that

$$H < \frac{H-F}{1-\rho}, \quad \text{or } \rho > \frac{F}{H}.$$

Therefore, the strategies specified in (8.22) constitute a Nash equilibrium when condition (8.21) holds. ■

Proposition 8.8 conveys the very idea that in order to deter entry the incumbent must carry out a costly activity, which is investing in extra capital (capital that is not needed for production purposes). This idea was suggested earlier by Schelling (1960), where he argued that in games involving such conflicts, a threat that is costly to carry out can be made credible by entering into an advanced commitment. That is, we showed that despite the fact that capital lasts for two periods, an incumbent monopoly must invest in each period in order to make entry unprofitable for potential entrants. If the incumbent neglects to invest in even one period, the entrant can credibly cause the exit of the incumbent by investing in capital. Thus, the fact that capital lasts for more than one period makes investing in capital a credible entry-deterring strategy because it ensures the existence of a firm in a subsequent period.

8.4.4 Judo economics

So far, our discussion of entry deterrence has focused mainly on the incumbent firms. In this subsection, we analyze the strategic options available to the potential entrant prior to the time of entry into the industry. In particular, we analyze the entrant's choice of capacity when facing a large dominant incumbent firm that has the option to expand capacity and deter entry. We show that the potential entrant may profit by adopting a strategy of *judo economics* (Gelman and Salop 1983), which refers to having the entrant invest in only limited capacity—which would restrict the entrant's scale of entry and therefore its market share. We show that when the potential entrant limits its capacity sufficiently, it is the incumbent's best interest to accommodate entry rather than to fight it.

Consider a two-stage game in which in the first stage a potentially entering firm chooses: (a) whether to enter, (b) its capacity (maximum output) level, denoted by k and, (c) its price, denoted by p^e. In the second stage, the incumbent firm chooses its price, denoted by p^I. We assume that the incumbent firm is large in the sense that it has an unlimited capacity. Assume that production is costless and that the firms produce a homogeneous product for a single market with a demand curve given by $p = 100 - Q$. Also, assume that all consumers prefer the less expensive brand; however, consumers prefer the incumbent's brand at equal prices. Formally, let q^I denote the quantity demanded from the incumbent firm and q^e denote the quantity demanded from the entrant (if entering). Then, for a given sufficiently low capacity invested by the entrant, k, the demand facing each firm is given by

$$q^I = \begin{cases} 100 - p^I & \text{if } p^I \le p^e \\ 100 - k - p^I & \text{if } p^I > p^e \end{cases} \quad \text{and} \quad q^e = \begin{cases} k & \text{if } p^e < p^I \\ 0 & \text{if } p^e \ge p^I. \end{cases} \tag{8.23}$$

That is, after the entrant sets p^e, the incumbent can always undercut the entrant by setting $p^I = p^e$. However, if the incumbent sets a price slightly above the entrant's price, the entrant gets to sell the first k units and then the incumbent faces the residual demand given by $q^I = 100 - k - p^I$.

Suppose now that in the first stage the entrant enters and sets a capacity k and a price p^e. Then, in the second stage, the incumbent can deter entry by setting $p^I = p^e$ or accommodate entry by setting $p^I > p^e$. If entry is deterred, then the incumbent's profit is given by $\pi^I_D = p^e(100 - p^e)$. In contrast, if the incumbent accommodates entry, then the incumbent's profit is $\pi^I_A = p^I(100 - k - p^I)$. Thus, under entry

accommodation, the incumbent chooses $p^I > p^e$ to

$$\max_{p^I > p^e} \pi^I = p^I(100 - k - p^I),$$

yielding a first-order condition given by $0 = 100 - k - 2p^I$. Therefore, $p^I_A = (100 - k)/2$, hence $q^I_A = (100 - k)/2$ and $\pi^I_A = (100 - k)^2/4$. Comparing the incumbent's entry-deterring profit level to its profit under entry accommodation yields that

$$\pi^I_A \ge \pi^I_D \quad \text{and} \quad \frac{(100 - k)^2}{4} \ge p^e(100 - p^e). \tag{8.24}$$

Under entry accommodation, the entrant earns $\pi^e = p^e k > 0$.

We now turn to the first stage, where the entrant sets its capacity level and its price. Figure 8.9, derived from (8.24), illustrates the range of k and p^e that would induce the incumbent to accommodate entry.

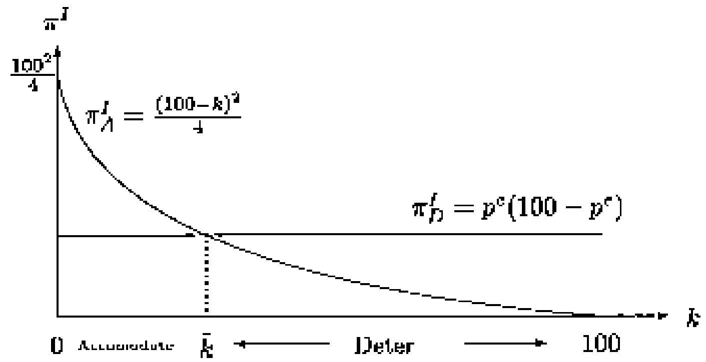

Figure 8.9: Judo economics: How an entrant secures entry accommodation

Figure 8.9 demonstrates that for a sufficiently low p^e, there always exists k small enough to induce the incumbent to accommodate entry according to the condition given in (8.24). More precisely, when the entrant reduces the price p^e, the horizontal line converges to the horizontal axis thereby increasing the area in which the incumbent accommodates the entry. Thus,

Proposition 8.9 *There exist a sufficiently limited capacity level k and a price p^e set by the entrant that ensure that the incumbent will find it profitable to accommodate the entry.*

The intuition behind this result is as follows. When the incumbent accommodates the entrant it does not match the entrant's price, but rather maintains an "umbrella" under which the entrant can prosper as long as it remains satisfied with its modest market share. In this case, the incumbent can maintain a higher price than the entrant and still sell since the entrant has a limited capacity that leaves a sufficiently profitable market share to the incumbent. Thus, when the entrant sets a sufficiently low capacity and price, entry deterrence (setting p^I to a very low level) yields a lower profit than entry accommodation to the incumbent firm.

The model presented in this subsection applies only to those situations in which the entrant can make credible capacity-limitation commitments. Such credibility can be enhanced by the use of contracts. For example, entry accommodation is sometimes observed in the airline industry, where large, established airline firms accommodate small carriers on some routes after observing that the entrant purchased a limited number of airport gates, a limited aircraft fleet, and low-capacity aircraft. Of course, as happens from time to time, some of these small entrants grow to become major carriers.

8.4.5 Credible spatial preemption

Our entry-deterrence analysis has concentrated so far on entry in a single market for a homogeneous product. In reality, firms produce differentiated, substitutable brands, so entry is likely to cause a head-to-head competition only on a subset of the incumbent's already produced brands. For example, in the airline industry where a monopoly airline is threatened by entry, it is likely to occur on a subset of the routes operated by the incumbent airline. The question raised in Judd 1985 is how would the incumbent firm react to partial entry, when entry into one market would affect the demand in a market for a substitute good, hence the incumbent's profit from the substitute good?

We demonstrate this entry problem by considering a monopoly firm (firm 1) which owns two restaurants, one Chinese (denoted by C) and one Japanese (denoted by J). Suppose that there are two consumers in town who are slightly differentiated with respect to the utility the receive from Chinese and Japanese food. More precisely, the utility of the consumer who is oriented toward Chinese food (U^C) and the utility of the consumer who is oriented toward Japanese food (U^J) are given by

$$U^C \equiv \begin{cases} \beta - p^C & \text{if eats Chinese food} \\ \beta - \lambda - p^J & \text{if eats Japanese food} \end{cases} \quad (8.25)$$

$$U^J \equiv \begin{cases} \beta - \lambda - p^C & \text{if eats Chinese food} \\ \beta - p^J & \text{if eats Japanese food} \end{cases}$$

where β reflects the satisfaction from eating, and $\lambda > 0$ denotes the slight

disutility a consumer has from buying his less preferred food. We assume that $\lambda < \beta < 2\lambda$, and normalize the restaurants' costs of operation to zero.

Suppose first that both restaurants are owned by a single firm (firm 1). Then, (8.25) implies that the monopoly owner would charge prices $p^C = p^J = \beta$ in each restaurant, and the monopoly's total profit would be $\pi_1 = 2\beta$.

Entry into the market for Chinese food

Suppose that a new restaurant (firm 2) with a different owner opens a new Chinese restaurant that serves food identical to the already existing Chinese restaurant owned by the monopoly. Assuming price competition, we see the price of Chinese food drop to zero (the assumed unit-production cost). Thus, $p_1^C = p_2^C = 0$. How would entry into the Chinese food market affect the price of a Japanese dinner? Well, clearly if the monopoly does reduce its price of a Chinese dinner to zero, all consumers including the one oriented toward Japanese food would purchase only Chinese food. Therefore, the maximum price the monopoly could charge for a Japanese dinner would be $p^J = \lambda$. Clearly, for this price the consumer oriented toward Japanese food would purchase Japanese since

$$U^J(J) = \beta - p^J = \beta - \lambda \geq \beta - \lambda - p^C = U^J(C).$$

That is, at $p^J = \lambda$ the Japanese-food-oriented consumer is indifferent to whether he or she buys Japanese (gaining a utility of $U^J(J)$) or Chinese (gaining $U^J(C)$). In this case the profit earned by the monopoly after the entry into the Chinese-food market occurs is $\pi_1 = \lambda$.

Incumbent withdraws from the Chinese restaurant

Now suppose that firm 1 (the initial monopoly on oriental food) shuts down its Chinese restaurant and keeps only the Japanese restaurant. In this event, after entry occurs, there are two restaurants, one serving Chinese food and the other serving Japanese food. Thus, the market structure is now a duopoly with firms selling differentiated products.

Lemma 8.1 *The unique duopoly price game between the Chinese and the Japanese restaurants results in the consumer oriented toward Japanese food buying from the Japanese restaurant, the consumer oriented toward Chinese food buying from the Chinese restaurant, and equilibrium prices given by* $p_1^J = p_2^C = \beta$.

Proof. We have to show that no restaurant can increase its profit by undercutting the price of the competing restaurant. If the Japanese restaurant would like to attract the consumer oriented toward Chinese food it has to set $p^J = p^G - \lambda = \beta - \lambda$. In this case, $\pi_2 = 2(\beta - \lambda)$. However, when it does not undercut, $\pi_2 = \beta > 2(\beta - \lambda)$ since we assumed that $\beta < 2\lambda$. A similar argument reveals why the Chinese restaurant would not undercut the Japanese restaurant. ■

We can now state our major proposition.

Proposition 8.10 *When faced with entry into the Chinese restaurant's market, the incumbent monopoly firm would maximize its profit by completely withdrawing from the Chinese restaurant's market.*

Proof. The profit of the incumbent when it operates the two restaurants after the entry occurs is $\pi_1 = \lambda$. If the incumbent withdraws from the Chinese restaurant and operates only the Japanese restaurants, Lemma 8.1 implies that $\pi_1 = \beta > \lambda$. ■

The intuition behind Proposition 8.10 is as follows. When entry occurs in one market, the price falls to unit cost. Given the reduction in this price, consumers buying a substitute good (Japanese food) would switch to buying Chinese food. Hence, the incumbent would have to reduce the price in its other market despite the fact that no entry occurred in the other market. Consequently, the incumbent would suffer a profit reduction in both markets. To avoid the latter, the incumbent would benefit from withdrawing and letting the entrant charge a higher price in the competing market. This would enable the incumbent to maintain the monopoly price in the remaining monopolized market (Japanese food). Thus, by withdrawing from competition, the incumbent differentiates itself from the entrant, so both firms could maintain a high price.

8.4.6 Limit pricing as cost signaling

Friedman's argument concerning the irrelevance of limit pricing raises the question whether incumbent firms would ever find it useful to exercise limit pricing during the preentry period. Milgrom and Roberts (1982) came up with an argument that limit pricing (or, expanded capacity or quantity produced) can serve as a cost-signaling device to the potential entrant who may not know the cost structure of the incumbent firm. We discuss here a simplified version of their model.

Demand, firms, and timing

There are two periods denoted by $t = 1, 2$. The market demand curve in each period is given by $p = 10 - Q$, where Q is the aggregate amount sold to consumers. Firm 1 is the incumbent and has to choose an output level in period 1 denoted by q_1^1. Firm 2 does not exist in $t = 1$ and chooses whether (or not) to enter only in the second period. Thus, firm 1 earns profits in the preentry period ($t = 1$) and in $t = 2$.

What about the output levels in the second period? Following Friedman's argument, we assume the following:

ASSUMPTION 8.2 *In the second period ($t = 2$), if entry occurs, then both firms play the Cournot game. If entry does not occur at $t = 2$, firm 1 produces the monopoly output level.*

This assumption highlights Friedman's argument in the sense that the incumbent's action at $t = 1$ has no influence on the market structure at $t = 2$, and therefore, we assume the most commonly used market structure for $t = 2$, which is Cournot if entry occurs and monopoly in the case of no entry.

Cost and information

Firm 2's unit-production cost is given by $c_2 = 1$. In addition, firm 2 has to pay an entry cost of $F_2 = 9$ if it enters at $t = 2$. The cost structure of firm 2 is assumed to be common knowledge.

In contrast, the cost structure of firm 1 (the incumbent) is known only to firm 1. The potential entrant does not exactly know the cost structure of the incumbent, but it knows the *probability distribution* of cost functions. Formally, firm 2 knows that the unit cost of firm 1 satisfies:

$$c_1 = \begin{cases} 0 & \text{with probability } 0.5 \\ 4 & \text{with probability } 0.5. \end{cases} \tag{8.26}$$

That is, firm 2 bases its decisions on the assumption that with 50% probability the incumbent is a low-cost firm ($c_1 = 0$), and with a 50% probability the incumbent is a high-cost firm ($c_1 = 4$).

Profits

The incumbent collects profits in periods 1 and 2 and maximizes the sum of the two periods' profits. The entrant collects profit only in the second period. In section 6.1, you have learned how to calculate the Cournot profit levels, so we avoid performing these simple calculations. These profit calculations are summarized in Table 8.2.

The two-period game

In the preentry era (period 1) firm 1 chooses its output level q_1^1. Thus, the profit of firm 1 in $t = 1$ is $\pi_1(c_1, q_1^1) = (10 - q_1^1)q_1^1 - c_1 q_1^1$.

In period 2, firm 2 observes q_1^1 and decides whether or not to enter. Its decision is based on the value of q_1^1 and on the estimated cost structure of firm 1, given in (8.26). Figure 8.10 illustrates this game.

Incumbent's cost:	Firm 2 (potential entrant)			
	ENTER		DO NOT ENTER	
Low ($c_1 = 0$)	$\pi_1^c(0) = 13$	$\pi_2^c(0) = -1.9$	$\pi_1^m(0) = 25$	$\pi_2 = 0$
High($c_1 = 4$)	$\pi_1^c(4) = 1$	$\pi_2^c(4) = 7$	$\pi_1^m(4) = 9$	$\pi_2 = 0$

Table 8.2: Profit levels for $t = 2$ (depending on the entry decision of firm 2). *Note:* All profits are functions of the cost of firm 1 (c_1); π_1^m is the monopoly profit of firm 1; π_i^c is the Cournot profit of firm i, $i = 1, 2$.

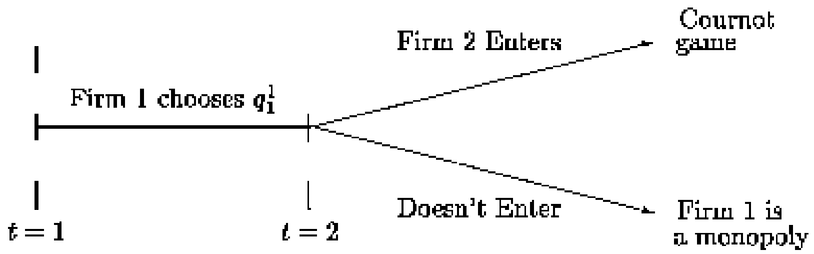

Figure 8.10: Two-period, signaling, entry-deterrence game

Solving the game assuming a high-cost incumbent

Without any further knowledge and assuming that firm 2 maximizes expected profit, we learn from (8.26) and Table 8.2 that upon entry firm 2's expected profit is

$$E\pi_2^c = \frac{1}{2}\pi_2^c(0) + \frac{1}{2}\pi_2^c(4) = \frac{1}{2}(-1.9) + \frac{1}{2}7 > 0,$$

hence, with no additional knowledge, firm 2 would enter. But why shouldn't the incumbent always state that it is a low-cost firm (rather than a high-cost firm)? Well, the incumbent can state whatever it wants, but firm 2 has no reason to believe the incumbent's statements.

Hence, given that entry occurs and the firms play Cournot in $t = 2$, the best firm 1 could do is to maximize the first-period profit by playing the monopoly's output in $t = 1$. That is, to set

$$q_1^1(4) = 3 \text{ and therefore earn } \pi_1(4) = \pi_1^m(4) + \pi_1^c(4) = 9 + 1 = 10. \tag{8.27}$$

Thus, if the incumbent is a high-cost firm, it would not attempt to limit its price and entry will occur.

Solving the game assuming a low-cost incumbent

Suppose that the incumbent (firm 1) is a low-cost firm ($c_1 = 0$). Then, if firm 2 were to know that firm 1 is a low-cost one, Table 8.2 shows that it would not enter since entry yields $\pi_2^c(0) < 0$. But since firm 2 does not know for sure that firm 1 is a low-cost one, the incumbent has the incentive to reveal it to firm 2. The purpose of this model is to demonstrate how limit pricing (or excess production) can serve as a means by which firm 1 can signal to firm 2 that it is a low-cost firm, thereby convincing firm 2 that entry is not profitable.

Proposition 8.11 *A low-cost incumbent would produce $q_1^1 = 5.83$, and entry will not occur in $t = 2$.*

Sketch of Proof. In order for the incumbent to convince firm 2 that it is indeed a low-cost firm, it has to do something "heroic." More precisely, in order to convince the potential entrant beyond all doubts that firm 1 is a low-cost one, it has to do something that a high-cost incumbent would never do—namely, it has to produce a first-period output level that is not profitable for a high-cost incumbent!

Now, a high-cost incumbent would not produce $q_1^1 = 5.83$ since

$$9.99 = (10-5.83)\times 5.83 - 4\times 5.83 + \pi_1^m(4) < \pi_1^m(4) + \pi_1^c(4) = 9+1 = 10. \tag{8.28}$$

That is, a high-cost incumbent is better off playing a monopoly in the first period and facing entry in the second period than playing $q_1^1 = 5.83$ in the first period and facing no entry in $t = 2$.

Finally, although we showed that $q_1^1 = 5.83$ indeed transmits the signal that the incumbent is a low-cost firm, why is $q_1^1 = 5.83$ the incumbent's profit-maximizing output level, given that the monopoly's output level is much lower, $q_1^m(0) = 5$. Clearly, the incumbent won't produce more than 5.83 since the profit is reduced (gets higher above the monopoly output level). Also (8.28) shows that any output level lower than 5.83 would induce entry, and given that entry occurs, the incumbent is best off playing monopoly in $t = 1$. Hence, we have to show that deterring entry by producing $q_1^1 = 5.83$ yields a higher profit than accommodating entry and producing the monopoly output level $q_1^1 = 5$ in $t = 1$. That is,

$$\pi_1(0)|_{q_1^1=5} = 25+13 = 38 < 49.31 = (10-5.83)\times 5.83+25 = \pi_1(0)|_{q_1^1=5.83},$$

hence, a low-cost incumbent will not allow entry and will not produce $q_1^1 < 5.83$. ■

8.4.7 Other entry-deterrence methods

The literature on entry deterrence explores various entry-deterring actions taken by incumbent firms (see survey articles by Neven [1989] and Wilson [1992]). One possible action referred to as *raising a rival's cost* is analyzed by Salop and Scheffman (1983). They suggest that incumbent firms may possess a variety of methods for raising the cost of entering firms. For example, one way of doing that is for the incumbent firm to sign high wage contracts, thereby raising the industry's labor cost. Another, is for the incumbent to lobby for higher tax rates. As noted earlier, potential entrants may be immune from these entry-deterring strategies since they may not be subjected to binding (wage and other cost) contracts. Note that in order for these actions to constitute entry-deterring methods, one needs to show that the these methods do not result in having the incumbent going bankrupt.

Another possible action analyzed in Aghion and Bolton 1987 suggests that incumbent firms rush to sign contracts with buyers in order to preempt entry. Gallini (1984) suggests that an incumbent can minimize its loss to firms producing potentially more advanced brands by simply licensing their own older technologies to potential entrants. The idea is that without licensing, potential entrants would develop superior technologies that would wipe out producers of older technologies. Finally, Spiegel (1993) demonstrates that incumbent firms can deter entry by subcontracting with other incumbent firms producing competing brands. Intuitively, if those firms have different cost structure, horizontal subcontracting reduces average costs of the incumbent firms, thereby reducing the likelihood that entry will occur.

Another way in which entry can be deterred is for the incumbent to deny access to a new technology by acquiring a patent right for its technology (see Gilbert and Newbery 1982). Finally, Scherer (1979) and Schmalensee (1978) analyze the FTC complaint that the four major cereal producers managed to deter entry by *proliferating* product varieties, thereby leaving insufficient room for the entry of new brands. Their result stems from the assumption of that the incumbent's decision to produce a brand is irreversible; however, subsection 8.4.5 demonstrates that incumbents may be better off to withdraw from the production of some brands in the presence of entry rather than fighting it.

8.5 Contestable Markets

Baumol, Panzar, and Willig (1982) proposed a market structure that describes the behavior of incumbent firms constantly faced by threats of entry. The main assumption underlying this market structure is that

entry does not require any sunk cost. Note that with the absence of sunk cost incumbent firms are subject to a hit-and-run entry, meaning that potential entrants can costlessly enter and exit the industry without having to wait until they generate a sufficient amount of revenue to recover the sunk cost of entry. Therefore, if incumbent firms do not have any cost advantage over potential entrants, a contestable market equilibrium will result in having an incumbent firm making only normal (zero) profit.

Assume that in a homogeneous product industry there is one incumbent firm facing entry by potential competitors. Let all firms have identical and increasing returns-to-scale technologies summarized by the cost function $TC(q_i) = F + cq_i$, and assume that the inverse aggregate demand facing the industry is given by $p = a - Q^d$.

DEFINITION 8.1

1. *An* **industry configuration** *is the incumbent's pair* (p^I, q^I) *of price charged and quantity produced.*
2. *An industry configuration is said to be* **feasible** *if*
   (a) *At the incumbent's price* p^I*, the quantity demanded equals the incumbent's quantity supplied. That is, if* $p^I = a - q^I$.
   (b) *The incumbent makes a nonnegative profit. That is,* $p^I q^I \geq F + cq^I$.
3. *An industry configuration is said to be* **sustainable** *if no potential entrant can make a profit by undercutting the incumbent's price. That is, there does not exist a price* p^e *satisfying* $p^e \leq p^I$ *and a corresponding entrant's output level* q^e *satisfying* $q^e \leq a - p^e$*, such that* $p^e q^e > F + cq^e$.
4. *A feasible industry configuration is said to be a* **contestable-markets equilibrium** *if it is sustainable.*

Thus, an industry configuration is sustainable if no other firm could make a strictly positive profit by setting a lower or equal price while producing no more than the quantity demanded by the consumers.

A contestable-market equilibrium is illustrated in Figure 8.11, where the price p^I and quantity produced q^I satisfy the consumers' aggregate demand curve and, in addition, lie on the incumbent's average total-cost function thereby ensuring that the incumbent does not incur a loss. Hence, this configuration is feasible.

Now, given that all firms share the same cost structure, it is clear that under the industry configuration illustrated in Figure 8.11 no other

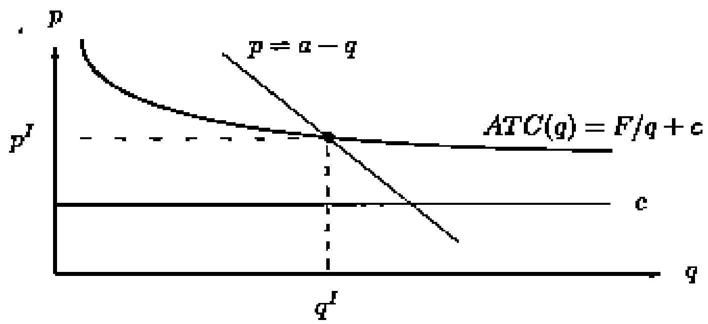

Figure 8.11: Contestable-markets equilibrium

firm could lower the price and make a strictly positive profit. Hence, this configuration is sustainable.

It should be noted that the contestable market structure can be used to describe an industry comprised of multiproduct firms, that is, firms producing a variety of different products (see Baumol, Panzar, and Willig 1982). Indeed, the advantage of using the contestable-markets market structure is that it can be applied to more realistic industries where firms produce more than one product, especially since all other market structures are defined for single-product firms that are rarely observed.

The contestable-market equilibrium defined in Definition 8.1 relies on the assumption that firms do not incur any sunk cost upon entry, and therefore can costlessly enter and exit the industry. This assumption is problematic since it is hard to imagine an industry where firms do not have to sink any irrevocable investment prior to entry. That is, firms generally conduct a market survey, place advertisements, and pay fees prior to entry, and these costs are definitely sunk and cannot be recovered. Moreover, Stiglitz (1987) pointed out the significance of this assumption by showing that if entrants face even tiny sunk costs prior to entry, then the only subgame perfect equilibrium is an incumbent charging a monopoly price and making a monopoly profit. In other words, although the contestable-market equilibrium yields a result that the incumbent makes zero profit, Proposition 8.5 showed that introducing even tiny sunk cost would imply that in a subgame perfect equilibrium of an entry-deterrence game, the incumbent makes pure monopoly profit. Therefore, the sensitivity of the market outcome to the existence of even small sunk cost is highly problematic because these two models have contradictory policy recommendations. On the one hand, contestable-market equilibrium implies that no intervention is needed by the antitrust authorities since even a single firm would charge the socially efficient price. On the other hand, the introduction of even a small sunk costs turns our model into a sustained monopoly, one which the antitrust authority would like to challenge. Finally, Schwartz (1986) has shown that, despite the assumed easy exit of potential entrants, hit and run entry is unprofitable if incumbent's price responses are sufficiently rapid.

8.6 Appendix: Merger and Antitrust Law

Section 7 of the Clayton Act (1914) states that

> No person engaged in commerce or in any activity affecting commerce shall acquire, directly or indirectly, the whole or any part of the stock or other share capital...shall acquire the whole or any part of the assets of another person engaged also in commerce or in any activity affecting commerce where in any line of commerce in any section of the country, the effect of such acquisition may be substantially to lessen competition, or to tend to create a monopoly.

Section 7 of the Clayton Act (amended in 1950) was needed because Sections 1 and 2 of the Sherman Act (1890) were not sufficient to halt mergers that would increase concentration and would reduce competition. Of course a question remains about why an increase in concentration would reduce competition and raise prices. This idea is built on two premises: First, that collusion or tacit coordination is less likely to succeed in less concentrated markets, where price cuts are less likely to be noticed by rival firms; and second, that antitrust should be viewed as consumer protection and that consumers tend to lose when faced by monopoly sellers. The discussion in this section is divided into two parts. We first discuss the procedure by which the FTC (Federal Trade Commission) and the DOJ (Department of Justice) can intervene in order to challenge a merger. Then, we proceed to the details, that the two agencies use to measure the effect of a merger. The interested reader is referred to Asch 1983; Fisher 1987; Gellhorn 1986; Salop 1987; and White 1987 for further reading and more references.

8.6.1 Challenging a merger

The monitoring of merger activities is in the hands of the FTC and the DOJ. The FTC issues guidelines to the DOJ recommending what types of mergers should be challenged. It is important to note that these guidelines do not constitute a law, but rather recommendations to the DOJ for starting to take actions against undesired mergers. In practice, firms with assets or sales in excess of \$100 million must report acquisitions of assets valued in excess of \$15 million. A merger does not take place until the FTC or the DOJ determines the competition effects of such an acquisition. With this procedure, very few cases are brought to courts since in most cases the FTC evaluation is sufficient for providing the signals to the acquiring firm about whether it should proceed with the acquisition or call it off.

8.6.2 Merger guidelines

The purpose of horizontal merger guidelines is to describe the analytical process that the agencies will employ to decide whether to challenge a merger; the guidelines are issued by the FTC and are suggestive rather than definitive. Salop (1987) summarizes five criteria that characterize those used by the FTC and the DOJ for evaluating a proposed merger: (1) the scope of the market upon which the merger may have anticompetitive effects; (2) the effect on concentration; (3) the ease of entry into the market; (4) other factors related to the case of collusion in the market; and (5) efficiency gains (such as cost reduction) associated with the merger.

In 1982, the Reagan administration came up with new merger guidelines (released in 1984 and modified in 1992). The scope of the relevant market was defined in price terms. That is, the relevant antitrust market is defined as a set of products and a geographical area where firms could profitably raise prices by at least 5% above the premerger price for at least one year. These guidelines suggest that a merger should not be challenged if the postmerger Herfindahl-Hirshman concentration index I_{HH}, defined by (8.2), satisfies

1. $I_{HH} < 1000$;
2. $1000 < I_{HH} < 1800$, and $\Delta I_{HH} < 100$;
3. $I_{HH} > 1800$, and $\Delta I_{HH} < 50$.

Thus, a merger is more likely to be challenged when it results in a higher concentration ratio and when it results in a more significant change in concentration. More precisely, at low postmerger concentration levels, a merger resulting in a change in the I_{HH} of a less than 100 would not be challenged. However, at a high postmerger I_{HH}, a merger leading to a change of less than 100 but greater than 50 is likely to be challenged.

In the above, ΔI_{HH} measures the difference in the I_{HH} measure before and after the proposed merger. For example, if firm 1, maintaining a market share s_1, and firm 2, maintaining a market share of s_2, merge, then the market share of the newly merged firm is expected to be s_1+s_2. In this case,

$$\Delta I_{HH} = (s_1+s_2)^2 - [(s_1)^2+(s_2)^2] = 2s_1s_2.$$

The higher the concentration is, the more likely merger is to be challenged even if the merger causes only a small increase in the degree of concentration.

Several authors, for example Farrell and Shapiro (1990) and those found in their references, have criticized the use of the I_{HH} as a reliable measure of a merger-induced change in concentration because it assumes that the merged firms maintain the exact sum of the market shares the merged firms had prior to the merger. However, it is likely that the sum of the market shares of the merged firm would fall after the merger in the case where entry barriers do not prevail.

Finally, in 1992 the DOJ and the FTC released modified horizontal merger guidelines (see, Department of Justice and Federal Trade Commission Horizontal Merger Guidelines, April 2, 1992). The release marks the first time that the two federal agencies that share antitrust enforcement jurisdiction have issued joint guidelines. The new guidelines reflect the experience of the DOJ and the FTC in applying the 1984 merger guidelines. The 1992 guidelines modify the test for identifying the relevant market. The 1984 guidelines hypothesized a uniform price increase to identify the market. Under the 1992 guidelines the price increase is not necessarily uniform. Instead, the new guidelines assume that a hypothetical monopolist may increase prices for some localities more than for others.

Similar to the 1984 guidelines, a post merger concentration level of $I_{HH} < 1000$ classifies the market in the region as unconcentrated. A post merger concentration of $1000 \leq I_{HH} \leq 1800$ is regarded as moderately concentrated. Mergers producing $\Delta I_{HH} > 100$ raise significant competitive concerns depending on the factors set forth in Sections 2-5 of the 1992 guidelines. Post merger concentration level $I_{HH} > 1800$ is regarded as highly concentrated. Mergers yielding a change in concentration $50 < \Delta I_{HH} \leq 100$ raise significant competitive concerns depending on the factors set forth in Sections 2-5 of the 1992 guidelines. Mergers yielding $\Delta I_{HH} > 100$ are regarded as likely to create or enhance market power or facilitate its exercise. This presumption may be overcome by showing that the factors set forth in Sections 2-5 of the 1992 guidelines make it unlikely that the merger will enhance market power.

Sections 2-5 consider potential adverse competitive effects of mergers, in addition to market concentration measured by the I_{HH}. These effects include (i) the likelihood of coordination among firms; (ii) conditions revealing implicit or explicit coordination such as common price, fixed price differentials, stable market shares, or consumer or territorial restrictions; (iii) detection of conditions making punishments on deviations from collusion more effective, thereby increasing the likelihood of collusion; (iv) the likelihood that a merger between firms distinguished by differentiated products to cause a price increase for all differentiated brands; (v) ability of rival sellers to replace lost competition.

8.7 Appendix: Entry Deterrence and Antitrust Law

Single-firm conduct is covered by Section 2 of the Sherman Act (1890), under which it would be a violation of the antitrust law for an incumbent firm to engage in actions that would limit competition, as stated in Section 2 of the Sherman Act (1890):

> Every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States, or with foreign nations, shall be deemed guilty of a felony.

Thus, Section 2 focuses on the unilateral conduct of a firm, whereas Section 1 focuses on the duality of actions among firms. More precisely, the essence of an offense under Section 1 is the act of joining together to conspire to limit competition, and therefore, the main concern is to find an agreement among firms. In contrast, Section 2 is concerned with the act of a monopoly that misuses its power by taking exclusionary actions.

Predatory prices are condemned, but there is little agreement on what defines predatory prices. A proof of pricing below average cost constitutes predatory pricing, and in this case the burden of proof is on the defendant to show that either the prices are not below average cost, or that the low prices are temporary, for promotional reasons only. However, prices that exceed average cost can still be considered as predatory if they are set in order to eliminate competition with other firms since any attempt to monopolize is a felony under Section 2.

Another violation of Section 2 is a *refusal to deal*, which refers to cases where a monopoly at one level of a chain of distribution refuses to deal with the next level in order to gain a monopoly position at both levels. Finally, product innovation is not considered to be a violation of Section 2 even if the introduction of the new product into the market makes it difficult for other firms to compete or even survive.

We conclude by discussing how the FTC handles anticompetitive behavior of incumbent firms. Section 5(a)(1) of the Federal Trade Commission Act (1914) states:

> Unfair methods of competition in or affecting commerce, and unfair or deceptive acts or practices in or affecting commerce, are declared unlawful.

In earlier periods after the FTC was established, the FTC concentrated on promoting "fair trade practices" among trade associations. Over the years, the FTC extended its role in enforcing these laws by conducting repeated investigations for the purpose of finding violations of firms that use a variety of anticompetitive methods, described earlier in the chapter, in order to maintain their dominance in the market. When the FTC suspects a violation, it opens an investigation against the suspected firms and looks at the product's design and its distribution channels in order to find a clue about whether these activities deter potential entrants from entering into the market. Investigations of these kinds are generally made public and by themselves encourage more firms to enter the market with competing brands, knowing that predatory activities will not be sustained.

8.8 Exercises

1. The bicycle industry consists of seven firms. Firms 1, 2, 3, 4 each has 10% market share, and firms 5, 6, 7 each has 20% market share. Using the concentration measures defined in Section 8.1, answer the following questions:
   (a) Calculate I_4 for this industry.
   (b) Calculate the I_{HH} for this industry.
   (c) Now, suppose that firms 1 and 2 merge, so that the new firm will have a market share of 20%.
      i. Calculate the post merger I_{HH}.
      ii. Calculate the change in the I_{HH} caused by the merger. That is, calculate ΔI_{HH}.
      iii. Using the merger guidelines described in subsection 8.6.2, evaluate the proposed merger and predict whether this merger will be challenged or not. Explain!
2. In an industry there are three firms producing a homogeneous product. Let q_i denote the output level of firm i, $i = 1, 2, 3$, and let Q denote the aggregate industry-production level. That is, $Q = q_1 + q_2 + q_3$. Assume that the demand curve facing the industry is $p = 100 - Q$. Solve the following problems:
   (a) Find the Cournot equilibrium output and profit level of each firm.
   (b) Now suppose that firms 2 and 3 merge into a single firm that we call firm 4. Calculate the profit level of firm 4 under a Cournot market structure.

(c) Do firms 2 and 3 benefit from this merger?

(d) Now suppose that firm 1 merges with firm 4. Does firm 4 benefit from the merger with firm 1?

(e) Explain why the first and the second mergers yield different results regarding the profitability of mergers.

3. Consider the merger among firms producing complementary components studied in subsection 8.2.3. Suppose that consumers desire computer systems composed of one computer (denoted as product X), and two diskettes (denoted as product Y). Thus, our consumers treat computers and diskettes as perfect complements where, for each computer, the consumers need two diskettes. Let p_X denote the price of a computer, and p_Y denote the price of a single diskette. Thus, the price of a computer system is $p_S = p_X + 2p_Y$. Formally, let the demand function for computer systems be given by

$$Q = \alpha - p_S = \alpha - (p_X + 2p_Y), \quad \text{where } Q = x = y/2, \ \alpha > 0.$$

Answer the following questions assuming that production is costless.

(a) Suppose that the X producer and the Y producer are independent. Solve for the Nash-Bertrand equilibrium in prices. Calculate the equilibrium prices, the quantity produced of each product, and firms' profit levels.

(b) Now suppose that firms X and Y merge under a single ownership. Calculate the monopoly equilibrium prices, the quantity produced of each product, and the monopoly's profit.

(c) Is this merger welfare-improving? Compare system prices and profits of the firms before and after the merger.

4. Consider the contestable-markets market structure defined in section 8.5. Suppose that in the industry there is one incumbent firm and several potential competitors all having identical technologies summarized by the cost function $TC(q_i) = 100 + (q_i)^2$, where q_i is the output of firm i. Solve for a contestable-markets equilibrium assuming that the (inverse) aggregate demand facing the industry is given by $p = 60 - 4Q^d$.

8.9 References

Aghion, P., and P. Bolton. 1987. "Contracts as a Barrier to Entry." *American Economic Review* 77: 388–401.

Asch, P. 1983. *Industrial Organization and Antitrust Policy*. New York: John Wiley & Sons.

Bain, J. 1956. *Barriers to New Competition*. Cambridge, Mass.: Harvard University Press.

Bain, J. 1972. *Essays on Price Theory and Industrial Organization*. Boston: Little, Brown.

Baumol, W., J. Panzar, and R. Willig. 1982. *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace Jovanovich.

Davidson, C., and R. Deneckere. 1984. "Horizontal Mergers and Collusive Behavior." *International Journal of Industrial Organization* 2: 117–132.

Dixit, A. 1979. "A Model of Duopoly Suggesting a Theory of Entry Barriers." *Bell Journal of Economics* 10: 20–32.

Dixit, A. 1980. "The Role of Investment in Entry-Deterrence." *Economic Journal* 90: 95–106.

Eaton, B. C., and R. Lipsey. 1980. "Exit Barriers are Entry Barriers: The Durability of Capital as a Barrier to Entry." *Bell Journal of Economics* 11: 721–729.

Economides, N., and S. Salop. 1992. "Competition and Integration Among Complements and Network Market Structure." *Journal of Industrial Economics* 40: 105–123.

Farrell, J., and C. Shapiro. 1990. "Horizontal Mergers: An Equilibrium Analysis." *American Economic Review* 80: 107–126.

Fisher, F. 1987. "Horizontal Mergers: Triage and Treatment." *Journal of Economic Perspectives* 1: 23–40.

Friedman, J. 1979. "On Entry Preventing Behavior and Limit Price Models of Entry." In *Applied Game Theory*, edited by S. Brams, A. Schotter, and G. Schwödiauer. Würzburg: Physica-Verlag.

Gallini, N. 1984. "Deterrence through Market Sharing: A Strategic Incentive for Licensing." *American Economic Review* 74: 931–941.

Gaudet, G., and S. Salant. 1992. "Mergers of Producers of Perfect Complements Competing in Price." *Economics Letters* 39: 359–364.

Gellhorn, E. 1986. *Antitrust Law and Economics*. St. Paul, Minn.: West Publishing Co.

Gelman, J., and S. Salop. 1983. "Judo Economics: Capacity Limitation and Coupon Competition." *Bell Journal of Economics* 14: 315–325.

Geroski, P., R. Gilbert, and A. Jacquemin. 1990. *Barriers to Entry and Strategic Competition*. Chur, Switzerland: Hardwood Academic Publishers.

Gilbert, R., and D. Newbery. 1982. "Pre-emptive Patenting and the Persistence of Monopoly." *American Economic Review* 72: 514–526.

Jacquemin, A. 1987. *The New Industrial Organization*. Cambridge, Mass.: MIT Press.

Judd, K. 1985. "Credible Spatial Preemption." *Rand Journal of Economics* 16: 153–166.

Milgrom, P., and J. Roberts. 1982. "Limit Pricing and Entry Under Incomplete Information: An Equilibrium Analysis." *Econometrica* 50: 443–459.

Neven, D. 1989. "Strategic Entry Deterrence: Recent Developments in the Economics of Industry." *Journal of Economic Surveys* 3: 213–233.

Ordover, J., G. Saloner, and S. Salop. 1990. "Equilibrium Vertical Foreclosure." *American Economic Review* 80: 127–141.

Perry, M. 1989. "Vertical Integration: Determinants and Effects." In *Handbook of Industrial Organization,* edited by R. Schmalensee, and R. Willing. Amsterdam: North-Holland.

Ravenscraft, D., and F. M. Scherer. 1987. *Mergers, Sell-offs, and Economic Efficiency.* Washington, D.C.: The Brookings Institution.

Salant, S., S. Switzer, and R. Reynolds. 1983. "Losses from Horizontal Merger: The Effects of an Exogenous Change in Industry Structure on Cournot-Nash Equilibrium." *Quarterly Journal of Economics* 98: 185–199.

Salinger, M. 1988. "Vertical Mergers and Market Foreclosure." *Quarterly Journal of Economics* 77: 345–356.

Salop, S. 1987. "Symposium on Mergers and Antitrust." *Journal of Economic Perspectives* 1: 3–12,

Salop, S., and D. Scheffman. 1983. "Raising Rival's Cost." *American Economic Review* 73, 267–271.

Schelling, T. 1960. *The Strategy of Conflict.* Cambridge, Mass.: Harvard University Press.

Scherer, F. M. 1979. "The Welfare Economics of Product Variety: An Application to the Ready-To-Eat Cereals Industry." *Journal of Industrial Economics*, 28: 113–133.

Schmalensee, R. 1978. "Entry Deterrence in the Ready-To-Eat Breakfast Cereal Industry." *Bell Journal of Economics* 9: 305–327.

Schwartz, M. 1986. "The Nature and Scope of Contestability Theory. *Oxford Economic Papers* 38(Suppl.): 37–57.

Sonnenschein, H. 1968. "The Dual of Duopoly in Complementary Monopoly: Or, Two of Cournot's Theories are One." *Journal Political Economy* 76: 316–318.

Spence, M. 1977. "Entry, Capacity, Investment and Oligopolistic Pricing." *Bell Journal of Economics* 8: 534–544.

Spiegel, Y. 1993. "Horizontal Subcontracting." *Rand Journal of Economics* 24: 570–590.

Stiglitz, J. 1987. "Technological Change, Sunk Costs, and Competition." *Brookings Papers on Economic Activity* 3 (Special Issue on Microeconomics: 883–937.

Tirole, J. 1988. *The Theory of Industrial Organization.* Cambridge, Mass.: MIT Press.

von Weizsäcker. 1980. "A Welfare Analysis of Barriers to Entry." *RAND Journal of Economics* 11: 399–420.

White, L. 1987. "Antitrust and Merger Policy: A Review and Critique." *Journal of Economic Perspectives* 1: 13–22.

Wilson, R. 1992. "Strategic Models of Entry Deterrence." In *Handbook of Game Theory,* edited by R. Aumann, and S. Hart. Amsterdam: North-Holland.