PART III

Technology and Market Structure

Chapter 9

Research and Development

> Innovation is an activity in which "dry holes" and "blind alleys" are the rule, not the exception.
> —Jorde and Teece, "Innovation and Cooperation: Implications for Competition and Antitrust"

Innovation is the search for, and the discovery, development, improvement, adoption, and commercialization of new processes, new products, and new organizational structures and procedures. Firms spend substantial amounts on research and development (R&D). In the developed countries, industries can be characterized according to the ratio of their R&D expenditure to output sales. Industries that exhibit high ratios include aerospace (23%), office machines and computers (18%), electronics (10%), and drugs (9%). Industries with R&D expenditure to output ratios of less than 1% include food, oil refining, printing, furniture, and textiles (OECD, 1980 data).

So far, in our analysis we have assumed that a production process or know-how can be characterized by a well-defined *production function* or by its dual the *cost function* (see section 3.1). Moreover, we have assumed that the production function is exogenous to the firms and viewed as a "black box" by the producers. In this chapter we analyze how firms can influence what is going on inside these black boxes by investing resources in innovation activities. We then analyze the methods by which society protects the right of innovators in order to enhance innovation activities in the economy.

Research and development is generally classified into two types: (a) *process innovation,* the investment in labs searching for cost-reducing technologies for producing a certain product, and (b) *product innovation*, the search for technologies for producing new products. It is often argued that from a logical point of view there is no difference between the two types of innovation since product innovation can be viewed as a cost-reducing innovation where the production cost is reduced from infinity (when the product was not available) to a finite level. However, many intuitively believe that there is a difference.

The concept of R&D is very difficult to understand and therefore to model, since the act of doing R&D means the production of knowledge or know-how (see Mokyr 1990 and Rosenberg 1994 for a historical overview of innovation, and Dosi 1988 and Freeman 1982 a survey of the literature and empirical evidence of innovation). Although we have so far always succeeded in avoiding discussion of the foundation of production functions and what know-how is, in this chapter we discuss precisely that by defining R&D as the act of creating (or changing) the production functions.

Section 9.1 (Process Innovation) classifies two types of process innovation. Section 9.2 (Innovation Race) analyzes how firms compete for discovering new technologies, and evaluates whether the equilibrium R&D level is below or above the socially optimal R&D level. Section 9.3 (Cooperation in R&D) analyzes how R&D is affected when firms coordinate their R&D efforts. Section 9.4 (Patents) analyzes how society encourages R&D by granting patent rights to innovators and suggests a method for calculating the optimal duration of patents. Section 9.5 (Licensing an Innovation) explains why firms tend to license their patented technologies to competing firms. Section 9.6 (International R&D Races) analyzes why governments subsidize R&D for exporting firms. In the appendix, section 9.7 analyzes patent law from historical and legal perspectives. Section 9.8 discusses the legal approach to cooperative R&D.

9.1 Classifications of Process Innovation

This section classifies process (cost-reducing) innovation according to the magnitude of the cost reduction generated by the R&D process. Consider an industry producing a homogeneous product, and suppose that the firms compete in prices (i.e., Bertrand competition, described in section 6.3 on page 107). Assume that initially, all firms possess identical technologies, meaning that they all produce the product with a unit production cost $c_0 > 0$. Then initially, there is a unique Bertrand equilibrium where all firms sell at unit cost $p_0 = c_0$, make zero profits, and produce a total of Q_0 units of output. This equilibrium is illustrated in Figure 9.1.

Suppose now that one and only one firm has the following R&D technology: the firm can construct a research lab engaging in cost-reducing innovation that leads to a unit-cost technology of $c < c_0$. Now, recalling from chapter 5 that the pure monopoly's profit-maximizing output and price can be found by equating $MR(Q) = c$, we distinguish between a large and a small cost-reducing innovation in the following way:

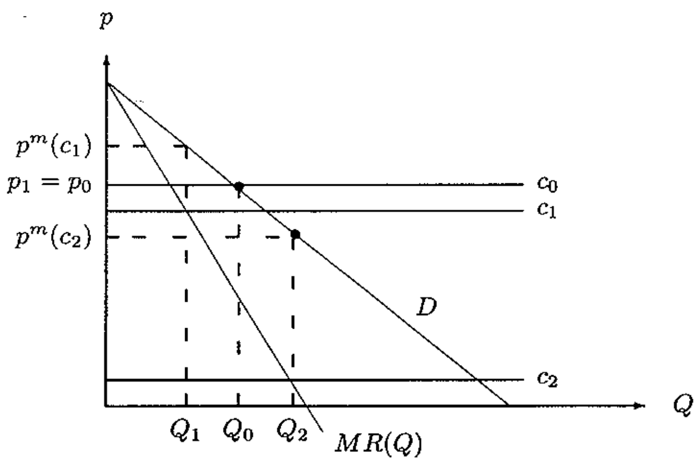

Figure 9.1: Classification of process innovation

DEFINITION 9.1 *Let $p^m(c)$ denote the price that would be charged by a monopoly firm whose unit production cost is given by c. Then,*

1. *Innovation is said to be* **large** *(or drastic, or major) if $p^m(c) < c_0$. That is, if innovation reduces the cost to a level where the associated pure monopoly price is lower than the unit production costs of the competing firms.*

2. *Innovation is said to be* **small** *(or nondrastic, or minor) if $p^m(c) > c_0$.*

Figure 9.1 illustrates the two types of process innovation. A cost reduction from c_0 to c_1 is what we call a small innovation. That is, the cost reduction is not large enough, implying that the innovating firm does not charge the pure monopoly price. In this case, the innovating firm will undercut all its rivals by charging a price of $p_1 = c_0 - \epsilon \approx c_0$, and will sell Q_0 units of output. In other words, a small innovation does not change the market price and the amount purchased by consumers. The only consequence of a small innovation is that the innovator sells to the entire market and makes strictly positive profit, equal to $(c_0 - c_1)Q_0$.

In contrast, a cost reduction from c_0 to c_2 in Figure 9.1 illustrates a large cost-reducing innovation, since the firm can undercut its rivals by simply charging the pure monopoly price associated with its new cost structure. That is, $p_2^m(c_2) < c_0$. Thus, a large innovation reduces the market price and increases quantity to Q_2.

Finally, note that Definition 9.1 connects the "physical" change of cost reduction with the market conditions (demand). That is, what we mean by small or large innovation depends on demand conditions and the market structure, in addition to the cost reduction itself.

9.2 Innovation Race

The timing of innovation plays a crucial role in the marketplace. There are two reasons why, in most cases, a firm that is first to discover a new technology or a new product gains an advantage over competing firms: First, the firm is eligible to obtain a patent protection that would result in earning monopoly profits for several years. Second, consumers associate the innovator with a higher-quality producer and will therefore be willing to pay a higher amount for the brand associated with the innovator.

Given the significance of becoming the first to discover, firms invest large sums in R&D, knowing that not discovering or discovering too late may result in a net loss from the innovation process. In this section we analyze the behavior of firms competing to discover a new product or a process, and we focus on the following questions: Do firms invest in R&D more or less than the socially optimal level? What is the impact of R&D competition on the expected date when the new product will be produced and marketed to consumers?

Assume that the discovery translates into a prize that can be viewed as the value of a patent associated with several years of earning monopoly profits.

Consider a two-firm industry searching for a new technology for producing a new product. The discovery of the product is uncertain. Each firm k, $k = 1, 2$, can engage itself in R&D by investing an amount of $\$I$ in a research lab. The payoff from R&D to a firm is as follows:

ASSUMPTION 9.1 *Once a firm invests $I in a lab, it has a probability α of discovering a technology that yields a profit of $V if the firm is the sole discoverer, $V/2 if both firms discover, and $0 if it does not discover.*

9.2.1 Equilibrium R&D in a race

We denote by $E\pi_k(n)$ the expected profit of firm k from investing in innovation when the total number of firms engaging in similar R&D is n,

$n = 1, 2$. Also, we denote by i_k ($i_k \in \{0, I\}$) the investment expenditure of firm k.

A single firm undertaking R&D

If only firm 1 invests in R&D, the firm discovers with probability α (therefore earning a profit of $V - I$) and does not discover with probability $1 - \alpha$ (earning a negative profit given by $-I$). Therefore, its expected profit is given by $E\pi_1(1) = \alpha V - I$. Hence, equating the expected profit to zero yields that the R&D investment decision of firm 1 is given by

$$i_1 = \begin{cases} I & \text{if } \alpha V \geq I \\ 0 & \text{otherwise.} \end{cases} \tag{9.1}$$

Two firms undertake R&D

The two-firm technology race highlights two important uncertainties facing firms engaging in R&D. First, there is *technological uncertainty*—whether or not the firm will discover the new product. Second, there is *market uncertainty*—whether or not the new product will be discovered by the rival firm.

When the two firms engage in R&D, the expected profit of each firm k is given by

$$E\pi_k(2) = \underbrace{\alpha(1-\alpha)V}_{\text{only } k \text{ discovers}} + \underbrace{\alpha^2 V/2}_{\text{both discover}} - I. \tag{9.2}$$

Equating (9.2) to zero implies that the following is a sufficient condition for having both firms profitably undertaking R&D:

$$i_1 = i_2 = I, \quad \text{if} \quad \frac{\alpha(2-\alpha)V}{2} \geq I. \tag{9.3}$$

Figure 9.2 illustrates the two conditions (9.1) and (9.3). When the combination of R&D cost and the success probability lies above the ray $E\pi_1(1) = 0$, no R&D is undertaken. That is, the combination of a low success probability or a high R&D cost yields the decision that innovation is not undertaken even under monopoly conditions.

Figure 9.2 also shows that when the R&D cost and probability combination lies between the curves $E\pi_k(2) = 0$ and $E\pi_1(1) = 0$, only one firm engages in R&D, whereas if this combination lies below $E\pi_k(2) = 0$, both firms undertake R&D.

9.2.2 Society's optimal R&D level

We now investigate what should be the number of firms that maximizes the society's welfare. In general, we should not expect that the equilib-

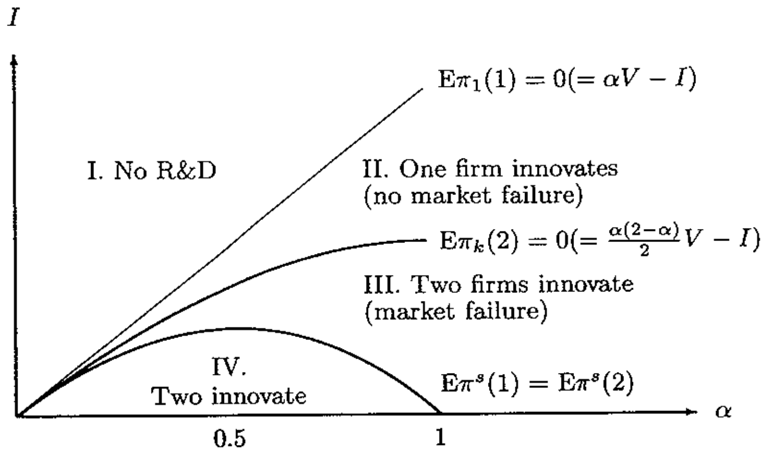

Figure 9.2: R&D race between two firms

rium number of firms calculated in the previous subsection is necessarily optimal since the action of undertaking R&D confers a negative externality on other firms engaging in the same R&D race. That is, from a social welfare point of view increasing the number of firms engaging in R&D will definitely increase the probability of discovery but will also increase the industry's aggregate R&D cost associated with R&D duplication. Therefore, without performing the actual calculation we find it hard to predict whether the equilibrium number of firms undertaking R&D is below or exceeds the optimal number.

We denote by $E\pi^S(n)$ the industry's expected profit when n firms undertake R&D, and associate the industry's expected profit with the welfare of the society.

When only one firm undertakes R&D ($n = 1$),

$$E\pi^S(1) = \alpha V - I = E\pi_1(1).$$

Thus, when there is only one firm, the social expected value of R&D coincides with the firm's expected profit from undertaking R&D.

When there are two firms undertaking R&D ($n = 2$),

$$E\pi^S(2) = \underbrace{2\alpha(1-\alpha)V}_{\text{only one firm discovers}} + \underbrace{\alpha^2 V}_{\text{both discover}} - 2I.$$

Comparing $E\pi^S(1)$ with $E\pi^S(2)$ yields that

$$E\pi^S(2) \geq E\pi^S(1) \quad \text{if and only if} \quad \alpha(1-\alpha)V \geq I.$$

Thus, in terms of Figure 9.2, any combination above the $E\pi^S(1) = E\pi^S(2)$ curve is associated with the situation where the socially optimal number of firms engaged in R&D is at most one.

Figure 9.2 is divided into four regions:

Region I: The combination of high innovation cost and a low probability of discovery makes it unprofitable for even a single firm to undertake innovation. It is obvious that if a single firm does not innovate, it is not beneficial for a society to engage in this R&D.

Region II: These combinations of cost and discovery probability leave room for only one firm to undertake R&D while still maintaining nonnegative expected profit. Since cost is relatively high (compared with the probability of discovery), there are no social benefits from having a second firm engaging in R&D.

Region III: A relatively low innovation cost makes it profitable for a second firm to engage in R&D. However, from the society's welfare point of view, the cost of duplicating the R&D effort ($2I$) is larger than the society's benefits from the increase in the likelihood of getting the discovery as a result of having a second firm engage in R&D. This a case of *market failure* which occurs because firms do not take into account how their R&D affect the profit of their rival firms.

Region IV: These combinations involve a low innovation cost, making it beneficial for both firms and the society to engage in the R&D race, despite the R&D cost duplication.

Proposition 9.1 *A market failure, a condition in which it is socially desirable to have at most one firm engaging in R&D but in equilibrium two firms engage in R&D, occurs only in Region III where the innovation cost, I, takes an intermediate value. Formally,*

$$E\pi^S(1) > E\pi^S(2) \text{ but } E\pi_k(2) > 0, \quad \text{when} \quad \alpha(1-\alpha)V < I < \frac{\alpha(2-\alpha)V}{2}.$$

In the literature, patent races are generally analyzed in continuous-time models, where the probability of discovery is a Poisson process that generates a constant probability of discovery at each point in time for a given R&D expenditure level (see Loury 1979; Lee and Wilde 1980; and Reinganum 1989 for such modeling). Fudenberg et al. (1983) analyze an industry where the probability of discovery increases with the length of time in which the R&D is conducted, and derive the conditions for having one firm preempt others from racing toward a discovery; see also Harris and Vickers 1985.

9.2.3 Expected date of discovery

Suppose that the race described in the previous subsection is repeated until one firm discovers the product. Then, what would be the expected date of discovery?

Before going to perform the calculations, we need the following lemma. The proof is given in an appendix (Section 9.9).

Lemma 9.1 *Let δ satisfy $0 < \delta < 1$. Then,*

$$\sum_{t=1}^{\infty} t\delta^{t-1} = \frac{1}{(1-\delta)^2}.$$

Let $T(n)$ denote the (uncertain) date when at least one firm discovers the product, given that n ($n \in \{1,2\}$) firms are engaged in R&D for discovering the same product. Also, let $ET(n)$ denote the expected date at which at least one firm discovers it.

A single firm

When only one firm engages in R&D ($n = 1$), the probability that $T(1) = 1$, (discovery occurs at the first date) is α. Next, the probability that $T(1) = 2$, (discovery occurs at the second date) is $(1-\alpha)\alpha$. That is, the probability that the firm does not discover at the first date times the probability that it discovers at the second date. Next, the probability that the firm discovers at the third date is $(1-\alpha)^2\alpha$. Hence, the expected date of discovery is given by

$$\begin{aligned} ET(1) &= \alpha 1 + (1-\alpha)\alpha 2 + (1-\alpha)^2\alpha 3 + (1-\alpha)^3\alpha 4 + \dots \\ &= \alpha \sum_{t=1}^{\infty} t(1-\alpha)^{t-1} \overset{\text{Lem 9.1}}{=} \frac{\alpha}{[1-(1-\alpha)]^2} = \frac{1}{\alpha}. \end{aligned} \tag{9.4}$$

Consequently, if the probability of discovery is $\alpha = 1/2$, then $ET(1) = 2$, and if $\alpha = 1/3$, then $ET(1) = 3$, and so on. Hence, as expected, an increase in the discovery probability α shortens the expected date of discovery.

Two firms

The probability that none of the firms discovers at a particular date is $(1-\alpha)^2$. Hence, the probability that at least one firm discovers at a particular date is $[1-(1-\alpha)^2] = \alpha(2-\alpha)$. Hence, $\text{prob}(T(2) = 1) = \alpha(2-\alpha)$. Next, $\text{prob}(T(2) = 2) = (1-\alpha)^2\alpha(2-\alpha)$ is the probability that none discovers at date 1 times the probability that at least one firm discovers at date 2. Therefore, the expected date of discovery when two firms engage in R&D is

$$\begin{aligned} ET(2) &= \alpha(2-\alpha)1 + (1-\alpha)^2\alpha(2-\alpha)2 + (1-\alpha)^4\alpha(2-\alpha)3 + \dots \\ &= \alpha(2-\alpha)\sum_{t=1}^{\infty} t(1-\alpha)^{2(t-1)} = \alpha(2-\alpha)\sum_{t=1}^{\infty} t\left[(1-\alpha)^2\right]^{t-1} \\ &= \frac{\alpha(2-\alpha)}{[1-(1-\alpha)^2]^2} = \frac{1}{\alpha(2-\alpha)}, \end{aligned} \tag{9.5}$$

where the fourth equality sign follows from Lemma 9.1. Comparing (9.4) with (9.5) yields $ET(2) < ET(1)$, meaning that opening more independent research labs shortens the expected date of discovery.

9.3 Cooperation in R&D

The antitrust legislation prohibits firms from engaging in activities that reduce competition and increase prices. Any attempt at collusion is sufficient to provoke lawsuit against the cooperating firms. However, the antitrust legislation is less clear about how to handle cases where firms establish research joint ventures (RJV) or just decide jointly how much to invest in their (separated) labs. The legal approach to RJV is addressed in the appendix (Section 9.8).

In this section, we do not address problems such as how firms manage to implicitly or explicitly coordinate their research efforts and how the research information is shared by the participating firms (see Combs 1993 and Gandal and Scotchmer 1993). Instead, we analyze how firms determine their research efforts, taking into consideration that they compete in the final good's market after the research is completed. This problem has been the subject of many papers (see Choi 1993; d'Aspremont and Jacquemin 1988; Kamien, Muller, and Zang 1992; Katz 1986, and Katz and Ordover 1990).

In this section we analyze a two-stage game in which at $t = 1$, firms determine (first noncooperatively and then cooperatively) how much to invest in cost-reducing R&D and, at $t = 2$, the firms are engaged in a Cournot quantity game in a market for a homogeneous product, where the demand function is given by $p = 100 - Q$.

The process-innovation R&D technology

We denote by x_i the amount of R&D undertaken by firm i, $i = 1, 2$, and by $c_i(x_1, x_2)$ the unit production cost of firm i, which is assumed to be a function of the R&D investment levels of both firms. Formally, let

$$c_i(x_1, x_2) \equiv 50 - x_i - \beta x_j \quad i \neq j,\ i = 1, 2, \quad \frac{3-\sqrt{7}}{2} < \beta < 1. \tag{9.6}$$

That is, the unit production cost of each firm declines with the R&D of both firms, where the parameter β measures the effect of firm j's R&D level on the unit production cost of firm i. Formally,

DEFINITION 9.2 *We say that R&D technologies exhibit (positive)* **spillover effects** *if* $\beta > 0$.

That is, if $\beta > 0$, the R&D of each firm reduces the unit cost of both firms. For example, spillover effects occur when some discoveries are made public during the innovation process (some secrets are not kept). Also, this positive externality can emerge from the labs investing in infrastructure or from research institutes and universities that benefit all other firms as well (see Jaffe 1986 for empirical evidence). Assuming $\beta > 0$ implies that R&D exhibits only positive spillover effects. However, note that in some cases β can be negative if the R&D of a firm also involves vandalism activities against competing firms, such as radar jamming or spreading false information and computer viruses.

Finally, to close the model we need to assume that R&D is costly to firms. Formally, denote by $TC_i(x_i)$ the cost (for firm i) of operating an R&D lab at a research level of x_i.

ASSUMPTION 9.2 *Research labs operate under decreasing returns to scale. Formally,*

$$TC_i(x_i) = \frac{(x_i)^2}{2}.$$

Assumption 9.2 implies that the cost per unit of R&D increases with the size of the lab. That is, higher R&D levels require proportionally higher costs of lab operation. Note that this assumption heavily affects the results because if labs were to operate under increasing returns (say, by having to pay a high fixed cost for the construction of the lab), firms would always benefit from operating only a single lab (that serves both firms) when firms are allowed to cooperate in R&D.

Subsection 9.3.1 calculates the firms' profit maximizing R&D levels when firms do not cooperate. Subsection 9.3.2 calculates the R&D levels that maximizes the firms' joint profit when firms are allowed to coordinate their R&D levels while still maintaining two separate labs.

9.3.1 Noncooperative R&D

We look for a subgame perfect equilibrium (Definition 2.10) where firms choose their R&D expenditure levels in the first period and their output levels in the second periods. We find this equilibrium by first solving for the Nash equilibrium in the second period and then working backwards, we solve for the first-period R&D levels.

The second period

The second-period Cournot competition takes place after the cost reduction innovation is completed. Hence, the postinnovation c_1 and c_2 are treated as given. Thus, our Cournot analysis of section 6.1 on page 98 applies; so, if we recall (6.7), the Cournot profit levels are given by

$$\pi_i(c_1, c_2)|_{t=2} = \frac{(100 - 2c_i + c_j)^2}{9} \quad \text{for } i = 1, 2,\ i \neq j. \tag{9.7}$$

The first period

In the first period, each firm noncooperatively chooses its level of R&D given the R&D level of the rival firm. That is, we look for a Nash equilibrium (Definition 2.4 on page 18) in R&D levels. Formally, substituting (9.6) into (9.7), for a given level of x_j, firm i chooses x_i to

$$\begin{aligned}\max_{x_i} \pi_i &= \frac{1}{9}[100 - 2(50 - x_i - \beta x_j) + 50 - x_j - \beta x_i]^2 - \frac{(x_i)^2}{2} \\ &= \frac{1}{9}[50 + (2-\beta)x_i + (2\beta - 1)x_j]^2 - \frac{(x_i)^2}{2}. \end{aligned}\tag{9.8}$$

The first-order condition yields

$$0 = \frac{\partial \pi_i}{\partial x_i} = \frac{2}{9}[50 + (2-\beta)x_i + (2\beta-1)x_j](2-\beta) - x_i.$$

Given that the payoff functions are symmetric between the two firms, we look for a symmetric Nash equilibrium where $x_1 = x_2 \equiv x^{nc}$, where x^{nc} is the common noncooperative equilibrium R&D level invested by each firm when the firms do not cooperate. Thus,

$$x^{nc} = \frac{50(2-\beta)}{4.5 - (2-\beta)(1+\beta)}. \tag{9.9}$$

9.3.2 Cooperative R&D

Under cooperative R&D, firms jointly choose R&D levels that will maximize their joint profits, knowing that in the second period they will compete in quantities.

The firms seek to jointly choose x_1 and x_2 to

$$\max_{x_1,x_2}(\pi_1 + \pi_2),$$

where π_i, $i = 1, 2$ are given in (9.7). The first-order conditions are given by

$$0 = \frac{\partial(\pi_1 + \pi_2)}{\partial x_i} = \frac{\partial \pi_i}{\partial x_i} + \frac{\partial \pi_j}{\partial x_i}.$$

The first term measures the marginal profitability of firm i from a small increase in its R&D (x_i), whereas the second term measures the marginal increase in firm j's profit due to the spillover effect from an increase in i's R&D effort. Hence,

$$\begin{aligned} 0 &= \frac{2}{9}[50 + (2-\beta)x_i + (2\beta - 1)x_j](2-\beta) - x_i \\ &+ \frac{2}{9}[50 + (2-\beta)x_j + (2\beta - 1)x_i](2\beta - 1). \end{aligned}$$

Assuming that second order conditions for a maximum are satisfied, the first order conditions yield the cooperative R&D level

$$x_1^c = x_2^c = x^c = \frac{50(\beta + 1)}{4.5 - (\beta + 1)^2}. \tag{9.10}$$

We now compare the industry's R&D and production levels under noncooperative R&D and cooperative R&D.

Proposition 9.2

1. *Cooperation in R&D increases firms' profits.*
2. *If the R&D spillover effect is large, then the cooperative R&D levels are higher than the noncooperative R&D levels. Formally, if $\beta > \frac{1}{2}$, then $x^c > x^{nc}$. In this case, $Q^c > Q^{nc}$.*
3. *If the R&D spillover effect is small, then the cooperative R&D levels are lower than the noncooperative R&D levels. Formally, if $\beta < \frac{1}{2}$, then $x^c < x^{nc}$. In this case, $Q^c < Q^{nc}$.*

Proof. For part 1, clearly, the firms could decide to set the R&D at the noncooperative levels. However, if they set $x^c \neq x^{nc}$, it means that their joint profit must increase. Parts 2 and 3 follow from comparing (9.9) with (9.10). The quantity comparisons follow from the simple fact that in a Cournot market structure, the aggregate quantity increases with a decline in unit production costs. ■

The intuition behind parts 2 and 3 of Proposition 9.2 is as follows. First note that under noncooperation, each firm sets its R&D level to reduce its own cost, ignoring that fact that it reduces the cost of the other firm as well. Now, if β is high, (the spillover effect is intense), then under cooperation the firms set R&D levels higher than the noncooperative levels since under cooperation firms take into account the effect of their R&D on their joint profits. When the spillover effect is small, the effect of each firm on the cost reduction of the other firm is small, hence, when firms do not cooperate each firm has a lot to gain from R&D since under small spillover effects, the R&D intensifies the cost advantage of the firm that undertakes a higher level of R&D.

Shaffer and Salant (1998) have pointed out some problems associated with the commonly used assumption that the two labs are engaged in an equal amount of R&D. They have shown that even though the aggregate R&D cost of identical firms in a research joint venture would be the lowest if they invested equally to reduce subsequent production costs, nonetheless members may enlarge their overall joint profit by instead signing agreements which mandate unequal investments. If we apply their analysis to our simple example, it turns out that unequal R&D levels maximize joint profit if the spillover parameter, β, is sufficiently low, $1 < 2(1-\beta)^2$ or $\beta < 0.3$; implying that we need to assume that $\beta > 0.3$ in order to make the analysis of this section valid.

Finally, in the present analysis the profit of firms must be higher under cooperation than under noncooperation since under cooperation in the first stage the firms can always invest at the noncooperative R&D level and earn the same profit as under noncooperation. However, Fershtman and Gandal (1994) show that the profit of the firms may be lower under cooperation in a (different) game where firms compete in R&D in the first period but collude in the second period. This happens since, depending on the second-period profit-sharing rule, each firm may over-invest in R&D in order to negotiate a larger fraction of the (cooperative) profit in the second period.

9.4 Patents

A *patent* is a legal document, granted by a government to an inventor, giving the inventor the sole right to exploit the particular invention for a given number of years (see an appendix (Section 9.7) for a detailed analysis of patent law). It is widely accepted that the patent system is useful for encouraging new product development and process innovation despite the market distortion it creates by granting temporary monopoly rights to new firms. Thus, the patent system is essential to growing economies. Empirically, it is very hard to measure the social value of a patent since patented invention tend to be rapidly imitated (or be "patented around" the patented innovation), so the knowledge is diffused into many firms, into other industries (see Mansfield 1965), and into other countries. One way to solve the problem how to measure the social value of a patented innovation is to count the number of times the innovation is cited in other patented innovations (see Trajtenberg 1990).

Formally, the patent system has two social goals: To provide firms with the incentives for producing know-how, and to make the new information concerning the new discoveries available to the public as fast as possible. In other words, society recognizes that information dispersion is a key factor in achieving progress and that public information reduces duplication of R&D. Note that the information-dissemination goal of the patent may somewhat contradict the pure interpretation of the patent law stating that a future innovation is patentable only if it does not infringe on earlier patented inventions. That is, on the one hand, society desires to disclose the information behind the invention in order to enhance the research by other firms; on the other hand, other firms would not be able to patent a technology that infringes on older patents. However, providing the public with the information on patented technologies definitely reduces extra cost resulting from R&D duplication in the sense that it prevents the wheel from being reinvented.

The reason why innovators need extra protection lies in the fact that know-how is a very special entity, compared with other products such as chairs, cars, and cheese: know-how is easy to duplicate and steal. Once a firm makes its invention known to others, other firms would immediately start with imitation followed by intense competition, thereby reducing the price to unit cost. With zero profits, no firm would ever engage in R&D, and the economy would stagnate forever.

The goal of the patent system is to reward innovators. The drawback of the system is that it creates a price distortion in the economy since those goods produced under patent protection will be priced differently from goods under no patent protection.

There are different kinds of patents, such as patents given for a new product, a new process, or a substance and a design patent. In order for an invention to be classified as worthy of a patent it has to satisfy three criteria: it has to be novel, nontrivial, and useful. In practice, it is hard to measure whether an invention satisfies these criteria, and therefore, patents tend to be approved as long as they do not infringe on earlier patented innovations. For a discussion of the legal side of the patent system and intellectual property see the appendix (Section 9.7). This appendix discusses many important legal and economic aspects of patent protection.

In this section, we discuss only one important and difficult aspect

of the patent system, the duration of patent protection. For example, in the United States, inventors are generally rewarded with seventeen years of patent protection, and in Europe with around twenty years of protection. Here, we wish to investigate what factors affect a society's optimal duration of patents.

We now provide a simple method for calculating the optimal duration of a patent that was proposed in Nordhaus 1969 and Scherer 1972. As in section 9.3, consider a firm that is capable of undertaking a process innovation R&D. An investment of x in R&D reduces the firm's unit cost from $c > 0$ to $c - x$. The cost of undertaking R&D at level x is the same as in Assumption 9.2. We assume that the innovation is minor (see Definition 9.1), so the innovating firm profit-maximizing price (assuming that the unit cost of all competing firms remain c) is $p = c$. Hence, there will be no change in output as a result of the innovation.

Figure 9.3 illustrates the market before and after the process innovation reduces the unit cost of the innovating firm by x, assuming a market demand given by $p = a - Q$, where $a > c$. Since there is no change in

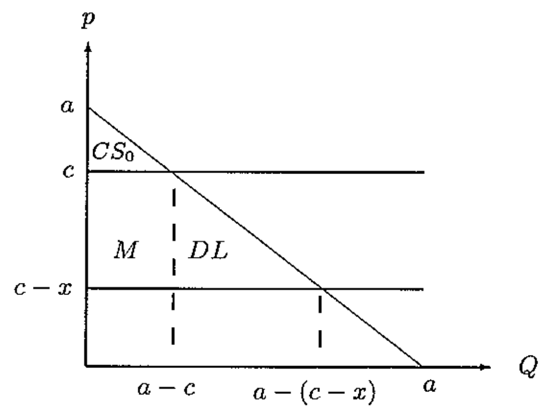

Figure 9.3: Gains and losses due to patent protection

price charged to the consumers, the area M in Figure 9.3 measures the innovator's gain in profit due to the innovation. Assuming that the government sets the patent life for $T \geq 0$ periods ($T = 17$ in the United States), we see that the innovator enjoys a profit of M for T periods, and zero profit from period $T + 1$ and on.

The area DL in Figure 9.3 is the *society's deadweight loss* resulting from the monopoly power held by the patent holder for T periods. That is, in periods $t = 1, 2, \ldots T$ the society's benefit from the innovation is only the monopoly's profits M (assuming that the profits are distributed to consumers, say, via dividends). In periods $t = T+1, T+2, \ldots$, after the patent expires, all firms have access to the new technology, and the equilibrium price falls to $c - x$. Hence, after the patent expires, the gain to the society is the sum of the areas, $M + DL$, since the removal of the monopoly rights expands output and increases consumer surplus by DL. It is clear from Figure 9.3 that

$$M(x) = (a-c)x \quad \text{and} \quad DL(x) = \frac{x^2}{2}. \tag{9.11}$$

Since the patent means monopoly rights for several periods, we need to develop a dynamic model in order to determine the optimal patent duration. Therefore, let ρ, $0 < \rho < 1$, denote the discount factor. Recall that the discount factor is how much a dollar next year is worth today. In perfect markets, the discount factor is also inversely related to the interest rate. That is, $\rho = 1/(1+r)$, where r is the market real-interest rate.

In what follows, we consider a two-stage game. In the first stage, the government sets the duration of the patent life T knowing how a firm would react and invest in cost-reducing R&D. In the second stage, at $t = 1$, the innovator takes the patent life as given, and chooses his or her R&D level. Then, during the periods $t = 1, \ldots T$, the innovator is protected by the patent rights and collects a monopoly profit for T periods.

9.4.1 Innovator's choice of R&D level for a given duration of patents

Denote by $\pi(x; T)$ the innovator's present value of profits when the innovator chooses an R&D level of x. Then, in the second stage the innovator takes the duration of patents T as given and chooses in period $t = 1$ R&D level x to

$$\max_x \pi(x; T) = \sum_{t=1}^{T} \rho^{t-1} M(x) - TC(x). \tag{9.12}$$

That is, the innovator chooses R&D level x to maximize the present value of T years of earning monopoly profits minus the cost of R&D. We need the following Lemma. The proof is given in an appendix, (Section 9.9).

Lemma 9.2

$$\sum_{t=1}^{T} \rho^{t-1} = \frac{1-\rho^T}{1-\rho}.$$

Hence, by Lemma 9.2 and (9.11), (9.12) can be written as

$$\max_x \frac{1-\rho^T}{1-\rho}(a-c)x - \frac{x^2}{2},$$

implying that the innovator's optimal R&D level is

$$x^I = \frac{1-\rho^T}{1-\rho}(a-c). \tag{9.13}$$

Hence,

Proposition 9.3

1. *The R&D level increases with the duration of the patent. Formally, x^I increases with T.*
2. *The R&D level increases with an increase in the demand, and decreases with an increase in the unit cost. Formally, x^I increases with an increase in a and decreases with an increase in c.*
3. *The R&D level increases with an increase in the discount factor ρ (or a decrease in the interest rate).*

The intuition behind Proposition 9.3 is as follows. When the duration of patents increases, the firm will be protected for a longer period and therefore will be selling more units over time. Thus, a higher R&D level would correspond to a unit-cost reduction for for a higher volume of production, which would make the process innovation even more profitable. The prediction of part 3 of Proposition 9.3 should remind you of your macroeconomics classes, where the Keynesian and ISLM approaches assumed that investment increases when interest rates fall. Here, we obtain this result: when the discount factor increases, say due to a drop in the real interest rate, the firm's present value of discounted profits increases, thereby making innovation more profitable.

9.4.2 Society's optimal duration of patents

We now turn to the first stage of the game, where the government legislates the duration of the patent to maximize social welfare, taking into account how the duration of patents affects the innovators' R&D level. As represented in Figure 9.3, the society's welfare is $CS_0 + M$ from the date the invention occurs, and $CS_0 + M + DL$ from the date when the monopoly's patent right expires.

Formally, the social planner calculates profit-maximizing R&D (9.13) for the innovator, and in period $t = 1$ chooses an optimal patent duration T to

$$\max_T W(T) \equiv \sum_{t=1}^{\infty} \rho^{t-1}\left[CS_0 + M(x^I)\right] + \sum_{t=T+1}^{\infty} \rho^{t-1} DL(x^I) - \frac{(x^I)^2}{2}$$
$$\text{s.t. } x^I = \frac{1-\rho^T}{1-\rho}(a-c). \tag{9.14}$$

Since

$$\sum_{t=T+1}^{\infty} \rho^{t-1} = \rho^T \sum_{t=0}^{\infty} \rho^t = \frac{\rho^T}{1-\rho}$$

and using (9.11), (9.14) can be written as choosing T^* to maximize

$$W(T) = \frac{CS_0 + (a-c)x^I}{1-\rho} - \frac{(x^I)^2}{2}\frac{1-\rho-\rho^T}{1-\rho} \quad \text{s.t.} \quad x^I = \frac{1-\rho^T}{1-\rho}(a-c). \tag{9.15}$$

Thus, the government acts as a leader since the innovator moves after the government sets the patent length T, and the government moves first and chooses T knowing how the innovator is going to respond.

We denote by T^* the society's optimal duration of patents. We are not going to actually perform this maximization problem in order to find T^*. In general, computer simulations can be used to find the welfare-maximizing T in case differentiation does not lead to an explicit solution, or when the discrete nature of the problem (i.e., T is a natural number) does not allow us to differentiate at all. However, one conclusion is easy to find:

Proposition 9.4 *The optimal patent life is finite. Formally, $T^* < \infty$.*

Proof. It is sufficient to show that the welfare level under a one-period patent protection ($T = 1$) exceeds the welfare level under the infinite patent life ($T = \infty$). The proof is divided into two parts for the cases where $\rho < 0.5$ and $\rho \geq 0.5$.

First, for $\rho < 0.5$ when $T = 1$, $x^I(1) = a - c$. Hence, by (9.15),

$$W(1) = \frac{CS_0 + (a-c)^2}{1-\rho} - \frac{(a-c)^2}{2}\frac{1-2\rho}{1-\rho} = \frac{CS_0}{1-\rho} + \frac{(a-c)^2}{1-\rho}\frac{1+2\rho}{2}. \tag{9.16}$$

When $T = +\infty$, $x^I(+\infty) = \frac{a-c}{1-\rho}$. Hence, by (9.15),

$$W(+\infty) = \frac{CS_0}{1-\rho} + \frac{(a-c)^2}{(1-\rho)^2} - \frac{(a-c)^2}{2(1-\rho)^2} = \frac{CS_0}{1-\rho} + \frac{(a-c)^2}{2(1-\rho)^2}. \tag{9.17}$$

A comparison of (9.16) with (9.17) yields that

$$W(1) > W(\infty) \iff \frac{(a-c)^2}{1-\rho}\frac{1+2\rho}{2} > \frac{(a-c)^2}{2(1-\rho)^2} \iff \rho < 0.5. \tag{9.18}$$

Second, for $\rho \geq 0.5$ we approximate T as a continuous variable. Differentiating (9.15) with respect to T and equating to zero yields

$$T^* = \frac{\ln[3 + \sqrt{6 + \rho^2 - 6\rho} - \rho] - \ln(3)}{\ln(\rho)} < \infty.$$

Now, instead of verifying the second-order condition, observe that for $T = 1$, $\mathrm{d}W(1)/\mathrm{d}T = [(a-c)^2\rho(1-5\rho)\ln(\rho)]/[2(1-\rho)^2] > 0$ for $\rho > 0.2$. ■

The result obtained in Proposition 9.4 is important because it is often argued in the literature that innovators should be granted an infinite patent life. The logic behind the infinite-patent-life argument is that in order to induce an innovator to undertake the optimal R&D level, the innovator should be rewarded with the entire profit stream from the innovation, which could last forever. That is, without the infinite patent protection, the innovator cannot capture all the rents from future sales associated with the innovation, and hence will not innovate at the optimal level. However, Proposition 9.4 shows that the monopoly distortion associated with an infinitely lived monopoly is larger than the innovation distortion associated with an insufficient reward to the innovator. Chou and Shy (1991, 1993) have found that this result also holds for patents given for product innovation (rather than for a process innovation in the present case). Also, Stigler (1968) provides an interesting calculation leading to an optimal patent life of seventeen years.

9.5 Licensing an Innovation

Licensing of technologies is common on both the national and the international scales. Over 80 percent of the inventions granted patents are licensed to other firms, where some are exclusively licensed and others are licensed to several manufacturers. Given this observation, we ask in this section why a firm that invested a substantial amount of resources in R&D would find it profitable to license its technology to a competing firm that has not invested in R&D. Several answers to this questions are given in the literature on patent licensing and surveyed in Kamien 1992. We answer this question by considering the following example:

Consider the simple two-firm Cournot example illustrated in Figure 9.3 and suppose that firm 1 has invented a (minor) cost-reducing process indicated by a lower unit cost $c_1 = c - x$, where c is the unit cost of the noninnovating firm 2 ($c_2 = c$).

No licensing

If firm 1 does not license its technology, the firms play Cournot, where in section 6.1 on page 98 we calculated that $\pi_1^c(c_1, c_2) > \pi_2^c(c_1, c_2)$ and $q_1^c(c_1, c_2) > q_2^c(c_1, c_2)$. That is, firm 1, with the lower unit cost, produces a higher amount and earns a higher profit than firm 2.

Licensing

Suppose that firm 1 negotiates with firm 2 for granting permission to firm 2 to use the less costly technology. There can be several types of licensing. For example, there can be a *fixed-fee* license (a fee that is independent of the output produced by firm 2), or firm 1 can charge firm 2 with a *per-unit fee* for every unit sold by firm 2.

Consider a per-unit fee case (that is very common in the electronics and entertainment industries, for example) in which firm 2 buys the technology for producing at unit cost of $c_1 < c_2$, and has to pay firm 1 the sum of $\$\phi$ for every unit it sells.

Although it is clear that the two firms have some surplus to divide between themselves, when firm 2 buys the cost-saving technology from firm 1, we take the simplest approach by assuming that firm 1 is a leader which offers firm 2 a take-it-or-leave-it contract to pay a per-unit fee of $\$\phi$. In other words, in the first stage firm 1 offers the technology to firm 2 for a per-unit fee. In the second stage, firm 2 can either reject the offer, or accept the offer and then choose how much to produce.

We now seek to find the profit-maximizing per-unit of output fee, ϕ, that firm 1 charges firm 2 for its cost-reducing technology. Clearly, firm 1 sets $\phi = (c_2 - c_1) - \epsilon \approx c_2 - c_1 = x$. That is, firm 1 charges a per-unit fee that is almost the size of the unit cost reduction associated with the licensed technology. Therefore, under this licensing contract, the (fee inclusive) per-unit cost facing firm 2 is now given by $c_{2'} = c_1 + \phi = c_2 - \epsilon \approx c_2$. Hence, in a Cournot equilibrium firm 2 would not change its quantity produced, and therefore, its profit level does not change.

In contrast, the profit of firm 1 is now given by $\pi_1 = \pi_1^c(c_1, c_2) + \phi q_2^c(c_1, c_2)$. That is, firm 1 gains all the surplus generated by the cost reduction in the production of firm 2. Therefore, we can state the following proposition.

Proposition 9.5

1. *In a Cournot environment, licensing a cost-reducing innovation can increase the profit of all firms.*
2. *In a Cournot environment, welfare increases when firms license cost-reducing innovations.*

The last part of the proposition follows from the fact that in our example, firms do not change their output levels and therefore the market price does not change. Hence, consumers' welfare remains unchanged. The profit of firm 1 increases, however, implying an aggregate welfare increase.

9.6 Governments and International R&D Races

We observe that governments never completely leave R&D to be performed by the free markets. Governments' intervention in R&D starts with the establishment of mandatory school systems and universities and ends with direct subsidies to firms or industries. In the developing countries, the gross estimation of the domestic R&D expenditure is around 3 to 3.5 percent of the GDP. Out of that, 30 to 60 percent is government financed.

In this section we analyze two examples in which international competition between firms located in different countries generates an incentive for each government to subsidize the R&D for the firm located in its country. Subsection 9.6.1 analyzes how a governmental subsidy to a domestic firm can secure the international dominance of the domestic firm in an international market for a new product. Subsection 9.6.2 analyzes governmental subsidies to process-innovation R&D.

9.6.1 Subsidizing new product development

Consider Krugman's (1986) illustration of how governments can enhance the international strategic position of the firms located in their countries. Suppose that there are only two civilian aircraft manufacturers in the entire world, and that the world consists of two countries, the United States and the European Community. Suppose that the U.S. manufacturer is called Boeing and the European firm is called Airbus. Each firm is considering developing the future super-large passenger plane, the "megacarrier," intended to transport six hundred passengers and having a flight range exceeding eighteen hours. Suppose further that each firm has a binary choice: develop (and produce) or don't develop (and don't produce). Table 9.1 demonstrates the profit levels of each firm under the four possible market outcomes.

Table 9.1 demonstrates what several civil aviation specialists frequently argue, that given the high development costs, a two-firm market is inconsistent with having positive profit levels. That is, in this kind of market, there can be at most one firm earning strictly positive profit. The Nash equilibrium (see Definition 2.4 on page 18) for this game is given in the following proposition.

		AIRBUS			
		Produce		Don't Produce	
BOEING	Produce	-10	-10	50	0
	Don't Produce	0	50	0	0

Table 9.1: Profits of Boeing and Airbus under no gov't intervention

Proposition 9.6 *In the Boeing-Airbus game, there exist exactly two Nash equilibria: (Produce, Don't Produce) and (Don't Produce, Produce).*

Now suppose that the EC subsidizes Airbus by providing fifteen units of money for the development of a European megacarrier. Table 9.2 illustrates the profit levels of the two aircraft manufacturers under the four possible outcomes.

		AIRBUS			
		Produce		Don't Produce	
BOEING	Produce	-10	5	50	0
	Don't Produce	0	65	0	0

Table 9.2: Profits of Boeing and Airbus under the EC subsidy

In this case, we can assert the following:

Proposition 9.7 *Under the EC subsidy, a unique Nash equilibrium is given by having Airbus play Produce and having Boeing play Don't Produce.*

Thus, by subsidizing product development, a government can secure the world dominance of the domestic firm in a product having large development costs relative to the potential market size. Although we have shown that the EC can guarantee its dominance in the megacarriers market by providing a subsidy to Airbus, it is not clear that the welfare of the EC residents increases with such a policy, since the EC residents will have to pay for this subsidy in one form or another!

9.6.2 Subsidizing process innovation

Following Brander and Spencer 1983 and 1985, consider two countries denoted by $i = 1, 2$, each of which has one firm producing a homogeneous product only for export, to be sold in the world market. The world's

demand for the product is $p = a - Q$; assume that the preinnovation unit cost of each firm is c where $0 < c < a$.

Let x_i denote the amount of R&D sponsored by the government in country i. We assume that when government i undertakes R&D at level x_i, the unit production cost for the firm producing in country i is reduced to $c - x_i$, $i = 1, 2$. As in Assumption 9.2 on page 230 we assume that the total cost to government i of engaging in R&D at level x_i is $TC_i(x_i) = (x_i)^2/2$.

Since we assumed that the two firms play a Cournot quantity game in the world market, for given R&D levels x_1 and x_2, (6.5) and (6.7) (see section 6.1 on page 98) imply that the profit level of the firm located in country i is

$$\pi_i = \frac{[a - 2(c - x_i) + c - x_j]^2}{9} = \frac{(a - c + 2x_i - x_j)^2}{9}.$$

We denote by W_i the welfare of country i, which is defined as the sum of the profit earned by firm i minus the R&D cost. Altogether, each government i takes x_j as given and chooses an R&D level x_i, to maximize the welfare of its country. That is, government i solves

$$\max_{x_i} W_i \equiv \pi_i - TC_i(x_i) = \frac{(a - c + 2x_i - x_j)^2}{9} - \frac{(x_i)^2}{2}.$$

The first-order condition yields how the government of country i sets its R&D level in response to the R&D set is country j. Thus,

$$x_i \equiv R_i(x_j) = 4(a - c) - 4x_j \quad i, j = 1, 2; \quad i \neq j. \tag{9.19}$$

Note that the countries' best-response functions are strategic substitutes (see Definition 7.2 on page 140), reflecting the fact that if one country increases its R&D level, the other reduces it. Equation (9.19) shows that if country j does not subsidize its R&D ($x_j = 0$), then the government of country i sets a strictly positive R&D level, $x_i = 4(a - c) > 0$. Hence,

Proposition 9.8 *If initially the world is characterized by no government intervention, it is always beneficial for at least one country to subsidize R&D. That is, the increase in profit from export sales associated with the cost-reducing R&D dominates the cost of R&D.*

Solving (9.19) yields the unique symmetric Nash equilibrium R&D levels given by

$$x_1^n = x_2^n = \frac{4(a - c)}{5}.$$

Proposition 9.9 *In a Nash equilibrium of an R&D game between two governments, each government subsidizes the R&D for the firm located in its country. Also, the equilibrium levels of the R&D subsidies increase with a shift in the world demand (a) and decrease with the initial unit-production cost (c).*

Thus, when demand rises, governments increase their R&D subsidies since cost reduction is magnified by larger sales.

Finally, the reader should not interpret this model as the ultimate argument for having governments subsidize R&D of the exporting firms, because this model does not explain why the government itself should perform the R&D. In other words, why does the private sector not invest in R&D, given that the firms' increase in profit can more than cover the R&D cost? Why cannot banks finance this innovation? Also, it is unlikely that governments possess all the information needed to decide which R&D is profitable and which is not. For arguments against protection, see Baldwin 1967. For a comprehensive survey of strategic trade policy, see Krugman 1986.

The result obtained in this subsection has been mitigated in several papers. First, Dixit and Grossman (1986) have shown that in a general equilibrium model (as compared with our partial equilibrium framework) the incentive for protection becomes weaker. Second, Eaton and Grossman (1986) have shown that the choice of policy instrument for helping the domestic industry depends heavily on the assumed market structure. Hence, since governments never know exactly whether the market structure is Cournot or a different one, the optimal policy may simply be not to intervene. Third, Gaudet and Salant (1991) show that the Brander and Spencer result is a special case because if one country has a large number of exporting firms and one has a small number of exporting firm, the optimal policy for the government in the country with the large number of firms may be a tax (instead of a subsidy) that will induce some firms to exit.

9.7 Appendix: Patent Law

A patent application is submitted to the Patent Office. Then the Patent Office examines the application and does research to determine whether the claims made by the petitioner fulfill the criteria for granting a patent. In many cases patents are denied by the Patent Office, and the innovator resubmits the application. During this time, it often happens that other innovators apply for similar patents, and in this case, the question of who invented first has to be answered by the Patent Office.

After the patent is granted, the patentee is given exclusive rights to make, use, or sell the invention, to the absolute exclusion of others. In the United States, the patent is granted for seventeen years and cannot be renewed.

9.7.1 History of Patent Law

The history of the patent system can be traced to medieval times in Europe when commerce became controlled by various groups and guilds. The reader interested in more details is referred to Kaufer 1989 and Miller and Davis 1990. The earlier patents issued by the Crown in England were a method used by the monarch to control various sectors in return for some benefits. That is, early patent rights were not as concerned with inventions as with the protection of the monarchy itself. In 1623 the Statute of Monopoly ended the period of unrestricted granting of monopolies by the Crown. In fact, the development of patent law was needed to secure monopoly rights for special reasons, such as to reward the innovators, rather than for the unrestricted granting of monopoly rights. In 1624 England passed a statute to regularize previously arbitrary "letters of patents" issued by the Crown.

The life of a patent was set at fourteen years because fourteen is two times seven, and seven years was the normal length of an apprenticeship (the time needed to train a professional, say a doctor). Then, the patent could be extended for seven additional years, reaching a maximum number of twenty-one years of patent protection. It is possible that the current U.S. system of seventeen years represents a compromise between fourteen and twenty-one years.

In the New World the colonies began granting patents; the colonists recognized that society could benefit from rewarding the innovators. All this led to the statement in the U.S. Constitution that

> The Congress shall have the power...To promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries.

Then, in 1836 the U.S. Patent Office was given the authority to examine proposed inventions and to determine whether they meet the criteria of the Patent Statute. In what follows, we will refer to the Patent Act of 1952 as the Patent Law.

9.7.2 Types of Patents

A patent can be granted for products, processes, plants, and design. However, any invention related to abstract ideas is not patentable. For example, the first person to prove Lemma 9.1 on page 228 (or any other lemma in this book) was not entitled to a patent right, since this "invention" is classified as an abstract idea, or a mathematical formula. However, note that applications for abstract ideas of theories may be patentable.

9.7.3 Criteria for granting a patent

In order for an invention to be entitled to a patent, it has to satisfy three requirements: novelty, nonobviousness, and usefulness. According to the patent law, novelty refers to the lack of prior domestic or foreign patenting, publication, use, or sale. Nonobviousness refers to the requirement that the invention must demonstrate some advance over "prior art" so that the ordinary mechanic skilled in prior art would not have been capable of making this advance. The purpose of the usefulness (or utility) requirement is to prevent patenting inventions that are based only on ingenuity and novelty but do not serve any purpose. This requirement also intends to steer the R&D towards inventing welfare-increasing inventions rather than useless ones.

9.7.4 First to invent versus first to file

The U.S. patent law differs from those of other countries in one major respect—the priority assignment given to one of several agents filing for the same patent. The general rule in the United States is that the innovator is the one who conceived first. However, one exception prevails, the case in which a second innovator reduces the invention into practice and the first innovator did not exercise continuous diligence. Thus, an innovator who is the first to conceive the innovation and the first to reduce it to practice has a definite priority in getting the patent.

The U.S. system is referred to as the *first-to-invent* system, which is not exercised by other countries. The EC and Japan use a different priority system, referred to as the *first-to-file* system. Obviously, the first-to-file system is easier to enforce. Problems arise nowadays when claiming a priority over international patents, since an invention could be recognized by one patent system but not the other.

9.7.5 Copyrights

Copyright gives an exclusive right to the copyright owner to reproduce the work and its derivatives in the form of copying or recording, and are given on the basis of pure originality, which refers to the act of authorship or artistic creativity, and not necessarily on novelty. The duration of the copyright ownership extends to the author's lifetime plus fifty years.

To obtain a copyright ownership, the author or the artist must demonstrate that he or she has contributed something to the final production or a reproduction. Thus, a reproduction of a book in modern style or with new decorations may be eligible for copyright protection because the author or artist has contributed something that did not exist in the earlier version. The Copyright Act also allows computer programs and sound recordings to receive copyright protection. Finally, the law permits reproduction of various works mainly for noncommercial purposes, such as education.

9.8 Appendix: The Legal Approach to R&D Joint Ventures

Two major questions are faced by the regulators regarding cooperative R&D: First, whether the act of joining together itself reduces competition, thereby violating antitrust laws. More precisely, should R&D joint ventures be considered as procompetitive or anticompetitive in the product's market? Second, even if R&D joint ventures are anticompetitive, are there efficiency gains associated with joint R&D that dominate the welfare loss resulting from anticompetitive behavior in the final-good market?

Clearly, unless the R&D joint ventures offer gains in efficiency associated with more productive and less costly R&D, there is no reason to permit it. For this reason, antitrust cases brought against firms cooperating in R&D are judged by the rule of reason rather than by the per se rule. The following discussion of the legal approach to cooperative R&D is based on Brodley 1990 and Jorde and Teece 1990.

The U.S. legal system seems to be less supportive of R&D joint ventures than the EC and Japan. According to the Clayton Act, allegations that firms use price fixing permit suing for treble damages. Therefore, there is a question of whether cooperation in R&D can open a channel of communication among firms to explicitly or implicitly collude on prices. Despite these suspicions, Congress has recognized the potential benefits associated with cooperative R&D and in 1984 enacted the National Cooperative Research Act (NCRA), which states that joint R&D ventures must not be held illegal per se. The NCRA established a registration procedure for joint R&D ventures. The firms that do follow the registration procedure are immune from paying treble damages on any antitrust violation. Instead, the maximum penalty for registered firms is limited to damages, interest, and costs.

In sum, the U.S. law attempts to distinguish between joint R&D and joint commercialization decisions by cooperating firms. The former is legal, and the latter is illegal. The reader should note that sometimes it is hard to distinguish between the two processes since the decision to commercialize an invention can be viewed as the last step of the R&D process. That is, it is possible that one firm has a comparative advantage in theoretical product development, while the other has one in making an innovation marketable. In this case, society may benefit from the formation of a joint venture despite the fact that joint commercialization may result in higher prices than those that obtain under pure competition, since otherwise, there might be no product at all. This approach is more common in Japan, where commercialization is an integral part of the R&D process.

9.9 Mathematical Appendix

Proof of Lemma 9.1 First, recall the (high school) identity given by

$$\sum_{t=0}^{\infty}\delta^t \equiv 1+\delta+\delta^2+\delta^3+\ldots=\frac{1}{1-\delta}.$$

Next,

$$\begin{aligned}\sum_{t=1}^{\infty}t\delta^{t-1} &= 1+2\delta+3\delta^2+4\delta^3+\ldots\\ &= (1+\delta+\delta^2+\delta^3+\ldots)+(\delta+\delta^2+\delta^3+\ldots)+(\delta^2+\delta^3+\ldots)\\ &= \frac{1}{1-\delta}+\frac{\delta}{1-\delta}+\frac{\delta^2}{1-\delta}+\ldots=\left(\frac{1}{1-\delta}\right)\left(\frac{1}{1-\delta}\right)=\frac{1}{(1-\delta)^2}.\end{aligned}$$

■

Proof of Lemma 9.2 Using the high school identity given at the beginning of this appendix section, we have it that

$$\begin{aligned}\sum_{t=1}^{T}\rho^{t-1}=\sum_{t=0}^{T-1}\rho^t &= \frac{1}{1-\rho}-\rho^T-\rho^{T+1}-\rho^{T+2}-\ldots\\ &= \frac{1}{1-\rho}-\rho^T(1+\rho+\rho^2+\rho^3+\ldots)\\ &= \frac{1}{1-\rho}-\frac{\rho^T}{1-\rho}=\frac{1-\rho^T}{1-\rho}.\end{aligned}$$

■

9.10 Exercises

1. Consider the classification of process R&D given in section 9.1. Suppose that the aggregate inverse-demand function is given by $p = a - Q$, and

that initially all the firms have identical unit costs measured by c_0, where $c_0 < a < 2c_0$. Suppose that one and only one of the firms is able to reduce its unit cost to $c_1 = 2c_0 - a$. Using Definition 9.1 infer whether this process innovation is considered to be minor or major.

2. Consider a three-firm version of the patent-race model studied in section 9.2. Suppose that each one of the three firms is capable of developing a product. Let V denote the monetary value of the patent associated with the new product. Each firm can construct a research lab provided that it invests $\$I$ in the lab. Assume that if a firm constructs a lab, it has a probability of $\alpha = 1/2$ of discovering the product.
   If only one firm discovers the product, it will earn a profit equal to the full value of the patent (i.e., $\$V$). If only two firms discover, then each will earn $\$V/2$, and if all three discover, then each will earn $\$V/3$. Answer the following questions.
   (a) Assuming that $I = 1$, calculate the minimal value of V that ensures that each firm will invest in constructing a lab.
   (b) Suppose now that firm 3 went out of business, and that a foreign firm purchased the two remaining firms. Calculate the minimal value of V that would induce the foreign owner of the two firms to run the two separate research labs instead of operating only one lab.
3. Consider the calculations of the expected time of discovery described in subsection 9.2.3. Suppose that n ($n \geq 2$) firms are engaged in R&D, where the probability of discovery by each firm at each date is α, $0 < \alpha < 1$. Answer the following questions.
   (a) What is the probability that none of the firms discovers at a particular date?
   (b) What is the probability that at least one firm discovers at a particular date?
   (c) Calculate the expected date of discovery.
4. Consider the Boeing-Airbus game described in Table 9.1 on page 242.
   (a) Calculate the minimal subsidy to Airbus that will ensure that Airbus will develop the megacarrier. Explain!
   (b) Suppose that the EC provides Airbus with fifteen units of money as a subsidy. Which subsidy by the U.S. government to Boeing would guarantee that Boeing will develop this megacarrier?
   (c) Suppose that the EC provides Airbus with fifteen units of money as a subsidy. Is there any level of subsidy given by the U.S. government that would deter Airbus from developing this airplane?
   (d) From your answer to the previous question, conclude whether the world benefits by having both governments subsidizing their own aircraft manufacturing firms. Explain!

9.11 References

Baldwin, R. 1967. "The Case Against Infant-Industry Tariff Protection." *Journal of Political Economy* 77: 295–305.

Brander, J., and B. Spencer. 1983. "International R&D Rivalry and Industrial Strategy." *Review of Economic Studies* 50: 707–722.

Brander, J., and B. Spencer. 1985. "Export Subsidies and International Market Share Rivalry." *Journal of International Economics* 18: 83–100.

Brodley J. 1990. "Antitrust Law and Innovation Cooperation." *Journal of Economic Perspectives* 4: 97–112.

Choi, J. 1993. "Cooperative R&D with Product Market Competition." *International Journal of Industrial Organization* 11: 553–571.

Chou, C., and O. Shy. 1991. "New Product Development and the Optimal Duration of Patents." *Southern Economic Journal* 57: 811–821.

Chou, C., and O. Shy. 1993. "The Crowding-Out Effects of Long Duration of Patents." *RAND Journal of Economics* 24: 304–312.

Combs, K. 1993. "The Role of Information Sharing in Cooperative Research and Development." *International Journal of Industrial Organization* 11: 535–551.

d'Aspremont, C., and A. Jacquemin. 1988. "Cooperative and Noncooperative R&D in Duopoly with Spillovers." *American Economic Review* 78: 1133–1137.

Dosi, G. 1988. "Sources, Procedures, and Microeconomic Effects of Innovation." *Journal of Economic Literature* 26: 1120–1171.

Dixit, A., and G. Grossman. 1986. "Targeted Export Promotion With Several Oligopolistic Industries." *Journal of International Economics* 21: 233–249.

Eaton, J., and G. Grossman. 1986. "Optimal Trade and Industrial Policy under Oligopoly." *Quarterly Journal of Economics* 2: 383–406.

Fershtman, C., and N. Gandal. 1994. "Disadvantageous Semicollusion." *International Journal of Industrial Organization* 12: 141–154.

Freeman, C. 1982. "The Economics of Industrial Innovation." 2nd ed. Cambridge, Mass.: MIT Press.

Fudenberg, D., R. Gilbert, J. Stiglitz, and J. Tirole. 1983. "Preemption, Leapfrogging, and Competition in Patent Races." *European Economic Review* 22: 3–31.

Gandal, N., and S. Scotchmer. 1993. "Coordinating Research Through Research Joint Ventures." *Journal of Public Economics* 51: 173–193.

Gaudet, G., and S. Salant. 1991. "Increasing the Profits of a Subset of Firms in Oligopoly Models with Strategic Substitutes." *American Economic Review* 81: 658–665.

Harris, C., and J. Vickers. 1985. "Perfect Equilibrium in a Model of Race." *Review of Economic Studies* 52: 193–209.

Jaffe, A. 1986. "Technological Opportunity and Spillovers of R&D: Evidence from Firm's Patents, Profits, and Market Value." *American Economic Review* 76: 984–1001.

Jorde, M., and D. Teece. 1990. "Innovation and Cooperation: Implications for Competition and Antitrust." *Journal of Economic Perspectives* 4: 75–96.

Kamien M. 1992. "Patent Licensing." In *Handbook of Game Theory,* edited by R. Aumann, and S. Hart. Amsterdam: North-Holland.

Kamien, M., E. Muller, and I. Zang. 1992. "Research Joint Ventures and R&D Cartel." *American Economic Review* 82: 1293–1306.

Katz, M. 1986. "An Analysis of Cooperative Research and Development." *Rand Journal of Economics* 17: 527–543.

Katz, M., and J. Ordover. 1990. "R&D Cooperation and Competition." *Brookings Papers on Economic Activity: Microeconomics,* 137–203.

Kaufer, E. 1989. *The Economics of the Patent System.* New York: Hardwood Academic Publishers.

Krugman, P. 1986. *Strategic Trade Policy and the New International Economics.* Cambridge, Mass.: MIT Press.

Lee, T., and L. Wilde. 1980. "Market Structure and Innovation: A Reformulation." *Quarterly Journal of Economics* 94: 429–436.

Loury, G. 1979. "Market Structure and Innovation." *Quarterly Journal of Economics* 93: 395–410.

Mansfield, E. 1965. "Rates of Return from Industrial R&D." *American Economic Review,* Papers and Proceedings 55: 741–766.

Mokyr, J. 1990. *The Lever of Riches: Technological Creativity and Economic Progress.* Oxford: Oxford University Press.

Miller, A., and M. Davis. 1990. *Intellectual Property, Patents, Trademarks, and Copyright in a Nutshell.* 2nd ed. St. Paul, Minn.: West Publishing.

Nordhaus, W. 1969. *Invention Growth, and Welfare: A Theoretical Treatment of Technological Change.* Cambridge, Mass.: MIT Press.

Reinganum, J. 1989. "The Timing of Innovation: Research, Development, and Diffusion." In *Handbook of Industrial Organization,* edited by R. Schmalensee and R. Willig. Amsterdam: North-Holland.

Rosenberg, N. 1994. *Exploring the Black Box.* Cambridge: Cambridge University Press.

Scherer, F. M. 1972. "Nordhaus' Theory of Optimal Patent Life: A Geometric Reinterpretation." *American Economic Review* 62: 422–427

Shaffer, G., and S. Salant. 1998. "Optimal Asymmetric Strategies in Research Joint Ventures." *International Journal of Industrial Organization,* 16: 195–208.

Stigler, G. 1968. *The Organization of the Industry.* Homewood, Ill.: Richard D. Irwin.

Trajtenberg, M. 1990. "A Penny for Your Quotes: Patent Citations and the Value of Innovation." *Rand Journal of Economics* 21: 172–187.

Chapter 10

The Economics of Compatibility and Standards

> Standards are always out of date. That is what makes them standards.
> —Alan Bennet, *Forty Years On* (1969)

Perhaps the most easily observed phenomenon is that people do not live alone. People (and all other animals) tend to live in groups (called villages, towns, cities, or countries) since they benefit from interacting with other people. In addition to the pure social observation that people just enjoy being around other people, the benefits of being and working together can be explained as follows:

Production: Most production processes involve teams or groups of people using other (complementary) intermediate products, such as machinery and computers. Therefore, for the production to be efficient, machinery, computers, and all other equipment supporting workers must be designed in a way that (a) different workers would be able to use the same equipment, and (b) the output generated by a certain machine would be able to be used by another worker operating a different machine.

Consumption: People "enjoy" consuming goods that are also used by other people. They like to watch the same movies, to exchange books, and to listen to music of the same composers. People observe what others buy and try to match their consumption with that of their neighbors.

Thus, we can conclude that product or brand *compatibility* affects both the productivity of workers and the welfare of consumers. In what follows, we start with some descriptive definitions. Later on in the chapter, we shall give more precise definitions.

DEFINITION 10.1

1. *Brands of products are said to be* **compatible**, *if they can work together, in the sense that the output of one brand can be operated or used by other brands. In this case, we say that the brands operate on the same* **standard**.

2. *Brands are said to be* **downward compatible** *if a newer model is compatible with an older model, but not necessarily the other way around.*

3. *Consumers' preferences are said to exhibit* **network externalities** *if the utility of each consumer increases with the number of other consumers purchasing the same brand.*

Examples for compatibility include products such as video and audio equipment (records and tapes), languages, railroad gauges, power supply, computer operating systems, computer software, communication equipment (the phone system, fax and telex machines, cellular and radio phones), keyboards (QWERTY versus DVORAK), and banks and automatic teller machines (ATMs).

More precisely, video tapes operate on various different standards such as VHS, Beta, and different sizes such as 8mm and VHS size. Music is recorded on LP's (long-play records), compact cassettes, and Compact Disks (CD). Cellular phones, which use airwaves instead of cables, are used in two different standards: analog or digital. The commonly used QWERTY (the first six letters on the upper row of the keyboard) English keyboard was designed so that it slows the typist, since fast typing is technically impossible on mechanical typewriters. The newer DVORAK system allows faster typing; however, people were reluctant to switch to it (see David 1985). Compatibility of automatic teller machines comes into effect when a customer carrying a bank card issued by one bank can withdraw cash from a machine servicing the clients holding a card issued by another bank. In fact, in Israel all the banks collude in the sense that any bank's card can be used on all teller machines. We will show later in this chapter that this behavior is indeed profitable to banks. Finally, extensions to the seven-bit ASCII code (the most widely used as a standard for saving and transmitting computer files) to eight-bit for the purpose of increasing the number of characters from 2^7 to 2^8

yielded several incompatible standards offered by MS-DOS, Macintosh, and other computers.

Downward compatibility is commonly observed in the software industry, where a newer version can read output files generated by the old version, but in many cases the older version cannot input files generated by the newer version.

An example for preferences exhibiting network externalities include all communication equipment. That is, it is unlikely that a person would purchase a phone knowing that nobody else uses it.

To illustrate the significance of the choice of standards on the profits of firms in a certain industry, Table 10.1 demonstrates a two-firm industry producing a product that can operate on two standards: standard α and standard β. Table 10.1 demonstrates a normal form game

		FIRM B			
		Standard α		Standard β	
FIRM A	Standard α	a	b	c	d
	Standard β	d	c	b	a

Table 10.1: Standardization game

where each firm can choose to construct its product to operate on standard α or standard β. The profits levels of the two firms are given by the nonnegative parameters a, b, c, and d, where the profit of each firm is affected by the standard choices of the two firms. We look for the Nash equilibria for this game (Definition 2.4 on page 18).

Proposition 10.1

1. *If* $a, b > \max\{c, d\}$, *then the industry produces on a single standard, that is,* (α, α) *and* (β, β) *are Nash equilibria.*

2. *If* $c, d > \max\{a, b\}$, *then the industry produces on two different standards, that is,* (α, β) *and* (β, α) *are Nash equilibria.*

Part 1 of Proposition 10.1 resembles the Battle of the Sexes game (see Table 2.2 on page 17), where the profit levels are high when the firms produce compatible brands (on the same standard). Industrywide compatibility is observed in the banking industry (ATM machines) and in many electronic appliances industries. Part 2 of the proposition demonstrates a polar case, where the firms can increase their profit by differentiating their brands and hence by constructing them to operate on different standards. Examples for this behavior include the computer industry's producing computer brands operating on different operating systems, and automobiles that are produced with model-specific parts. Thus, in this chapter we investigate firms' incentives to standardize and the effects of their choices on consumers' welfare.

There is a substantial amount of literature on compatibility issues. For a comprehensive discussion on the nature of standards see Kindleberger 1993. For literature surveys, see Farrell and Saloner 1987; David and Greenstein 1990; and Gabel 1991. Gandal 1994 provides some empirical evidence for the existence of network externalities in the computer software industry.

Our discussion of the economics of standardization is divided into three approaches: Section 10.1 (Network Externalities) analyzes an industry where consumer preferences exhibit network externalities. Section 10.2 (Supporting Services) shows that people's tendency to use products that are identical or compatible to the products purchased by others need not be explained by assuming that consumers' preferences exhibit network externalities. That is, it is possible that people will end up using compatible products even if their welfare is not directly affected by the consumption choice of other people. Section 10.3 (Components) analyzes interface compatibility of components that are to be combined into a single, usable system. Two applications of these theories are not discussed in this chapter. First, Conner and Rumelt 1991 provides an application of network externalities to explain why software firms do not always protect the software against copying. Second, an application is discussed in section 17.1, where we show that when the choice of restaurants depends on the choice of other consumers, a restaurant may refrain from raising its prices even when it faces a demand that exceeds its seating capacity.

10.1 The Network Externalities Approach

In this section we present the basic network-externality model, where consumers' valuation of a brand increases with the number of other consumers using the same brand.

10.1.1 The interdependent demand for communication services

One of the first attempts to model the aggregate demand for communication services is given in Rohlfs 1974.

The demand for phone services

Our point of departure is that the utility that a subscriber derives from a communication service increases as others join the system. Consider

a group of a continuum of potential phone users indexed by x on the unit interval $[0, 1]$. Unlike the study of the Hotelling location model of subsection 7.3.1 in which we interpreted consumers indexed by a high x as consumers oriented toward brand B, and consumers indexed by a low x as consumers oriented toward brand A, here, since we have only one type of service, we interpret consumers indexed by a low x as those who love to subscribe to a phone system (high willingness to pay), and consumers indexed by a high x as those who have less desire for subscribing to a phone system (low willingness to pay).

We denote by n, $0 \leq n \leq 1$ the total number of consumers who actually subscribe to the phone system, and by p the price of subscribing to the phone system. Altogether, we define the utility of a consumer indexed by x, $0 \leq x \leq 1$, as

$$U^x = \begin{cases} n(1-x) - p & \text{if he or she subscribes to the phone system} \\ 0 & \text{if he or she does not subscribe.} \end{cases} \tag{10.1}$$

Thus, the utility of each subscriber exhibits network externalities since it increases with n (the number of consumers subscribing to the phone system).

We now derive the consumers' aggregate demand for phone services. We first look at a particular consumer denoted by $\hat{x}$ who is at a given price p indifferent to the alternatives of subscribing to the phone system and not subscribing. In view of (10.1), the indifferent consumer is found by

$$0 = n(1 - \hat{x}) - p.$$

Since the number of consumers is given by $n = \hat{x}$, we have it that

$$0 = \hat{x}(1 - \hat{x}) - p \quad \text{or} \quad p = \hat{x}(1 - \hat{x}), \tag{10.2}$$

which is drawn in Figure 10.1. The price p_0 in Figure 10.1 intersects twice the 'flipped' U-shaped curve (at points $\hat{x}_0^L$ and $\hat{x}_0^H$). The interpretation for the two intersection points is that for a given price p_0 there can be two levels of demand: a low level, measured by $n = \hat{x}_0^L$, that is associated with a small number of subscribers, hence, by (10.1) with a low valuation by each subscriber, and therefore with a small number of users, and so forth. In contrast, at the given price p_0 there can be a high demand measured by $n = \hat{x}_0^H$, hence a high valuation by each subscriber, and therefore, a large number of subscribers, and so forth. However, only point $\hat{x}_0^H$ is a stable demand equilibrium, since at the intersection point $\hat{x}_0^L$ a small increase in the number of subscribers would make the phone subscription more desirable, thereby causing all the consumers indexed on $[\hat{x}_0^L, \hat{x}_0^H]$ to subscribe.

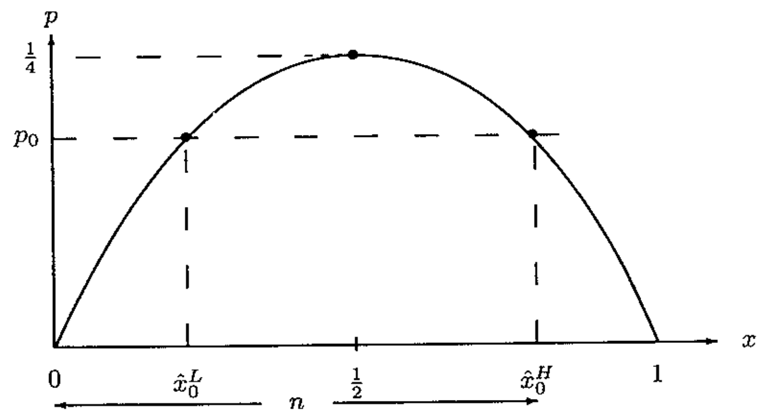

Figure 10.1: Deriving the demand for telecommunication services

The point $\hat{x}_0^L$ is defined in the literature as the *critical mass* for a given price p_0 to indicate that at a given price, any increase in the number of subscribers would shift the demand (number of subscribers) to the point $\hat{x}_0^H$.

The problem of the monopoly phone company

Now suppose that there is only one monopoly firm providing phone services, and suppose that the marginal cost of adding a subscriber is negligible, after the PTT (Public Telephone and Telegraph) company has already wired all the houses. We now ask what price maximizes the PTT's profit (equals revenue in our case)? To solve this problem, we formulate the PTT's profit-maximization problem, which is to choose $\hat{x}$ that solves

$$\max_{\hat{x}} \pi(\hat{x}) \equiv p(\hat{x})\hat{x} = \hat{x}(1-\hat{x})\hat{x} = (\hat{x})^2(1-\hat{x}). \tag{10.3}$$

The profit function (10.3) is drawn in Figure 10.2. The first- and second-order conditions for (10.3) are given by

$$0 = \frac{\partial \pi}{\partial x} = 2x - 3x^2 \quad \text{and} \quad \frac{\partial^2 \pi}{\partial x^2} = 2 - 6x. \tag{10.4}$$

Now, equation (10.4) and Figure 10.2 completely describe how the profit level is affected by changing the number of subscribers. Clearly, the profit is zero when there are no subscribers ($\hat{x} = 0$). The profit is

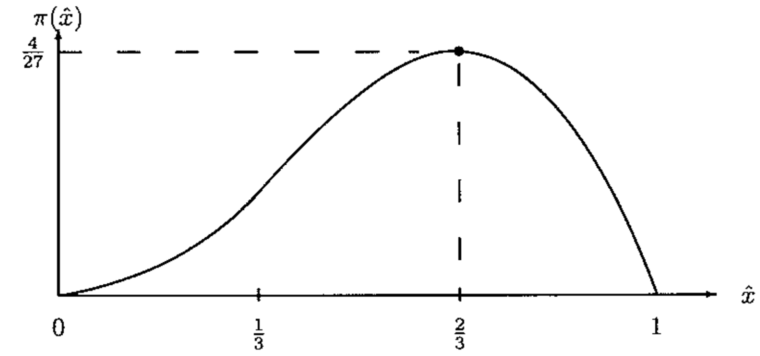

Figure 10.2: The PTT profit function in the presence of network externalities

also zero when the entire population subscribes, since in order to have the entire population subscribing, the PTT should set the price to zero.

The first-order condition shows that $\hat{x} = 0$ and $\hat{x} = 2/3$ are extremum points. In addition, the second-order condition shows that the second derivative is negative for $\hat{x} > 1/3$, implying that $\hat{x} = 2/3$ is a local maximum point. Since the first-order condition is positive for all $0 < \hat{x} < 2/3$, it must be that $\hat{x} = 2/3$ is a global maximum point. Hence,

Proposition 10.2 *A monopoly phone company's profit-maximizing subscription price is set such that the number of subscribers exceeds half of the consumer population but is less than the entire population.*

10.1.2 The standardization-variety tradeoff

In the previous subsection we confined the analysis to a single service. In this subsection we develop a different model in which we assume that there are two brands of the product and heterogeneous consumers, in the sense that each consumer prefers one brand over the other. There are two firms, each producing a different brand, brand A and brand B.

We assume a continuum of consumers, normalize the population size to 1, and assume that a $(0 < a < 1)$ consumers prefer brand A over brand B, whereas b $(0 < b < 1)$ consumers prefer brand B over brand A, where $a + b = 1$.

The Farrell and Saloner (1986) model assumes that the utility of each consumer type increases with the number of consumers buying the same brand. However, if a consumer purchases the less desired brand, his utility falls by $\delta > 0$. Formally, the utility functions of types A and

B consumers are given by

$$U^A = \begin{cases} x_A & \text{buys brand } A \\ x_B - \delta & \text{buys brand } B \end{cases} \qquad U^B = \begin{cases} x_A - \delta & \text{buys brand } A \\ x_B & \text{buys brand } B \end{cases} \tag{10.5}$$

where x_A denotes the number of consumers purchasing brand A and x_B denotes the number of consumers purchasing brand B, $x_A + x_B = 1$. The parameter δ also reflects the extra amount of money that a consumer is willing to pay to get his or her ideal brand.

DEFINITION 10.2

1. *If $x_A = 1$ and $x_B = 0$, we say that the product is standardized on A.*
2. *If $x_A = 0$ and $x_B = 1$, we say that the product is standardized on B.*
3. *If $x_A > 0$ and $x_B > 0$, we say that the product is produced with incompatible standards.*
4. *An allocation of buyers between brands x_A and x_B is called an equilibrium, if no single buyer would benefit from switching to the competing brand, given that all other consumers do not switch from their adopted brand.*

Equilibrium adoption of brands

We first seek necessary conditions for a single standard to be an equilibrium. Observe that in the following analysis, since we assume a continuum of consumers, each consumer is negligible in the sense that if a single consumer switches from buying brand A to buying brand B, then it will not affect the aggregate number of A and B users measured by x_A and x_B. Now, if the industry is standardized on A ($x_A = 1$), then it must be that type B consumers would not benefit from switching from A to B, implying that $1 - \delta > 0$. That is, a consumer prefers to consume the same brand as the others rather than consuming alone his or her most preferred brand (i.e., if the network effect dominates the ideal good effect). Therefore,

Proposition 10.3

1. *If $\delta < 1$, then two equilibria exist: one in which A is the standard ($x_A = 1$) and one in which B is the standard ($x_B = 1$).*
2. *If $\delta > 1$, no single-standard equilibrium exists.*

We now investigate under what conditions the industry will produce two incompatible brands, that is, under what conditions $x_A = a$ and $x_B = b$ is an equilibrium. In this equilibrium, a type A consumer would not switch to B if $a > b - \delta$. Since $b = 1 - a$, we have it that $a > \frac{1-\delta}{2}$. Similarly, type B would not switch if $b > \frac{1-\delta}{2}$. Hence,

Proposition 10.4 *If the number of each type of consumers is sufficiently large, then there exists a two-standard equilibrium. Formally, if $a, b > \frac{1-\delta}{2}$, then $x_A = a$, $x_B = b$ is an equilibrium.*

Figure 10.3 illustrates the parameter range for which the two-standard equilibrium exists. As the utility loss from consuming the less preferred

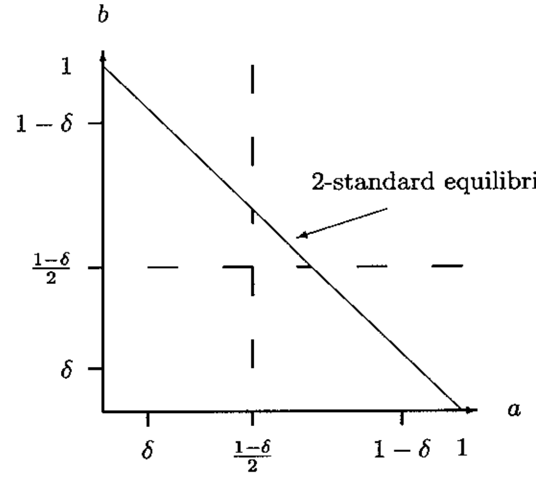

Figure 10.3: Two-standard (incompatibility) equilibrium

brand parameter δ increases, the parameter range for which incompatibility is an equilibrium increases. That is, if $\delta \geq 1$, a two-standard equilibrium always exists.

Efficiency of brand adoption

We define the economy's social welfare function as the sum of consumers' utilities. Formally, let $W \equiv aU^A + bU^B$. In view of the three possible outcomes described above, we have it that

$$W = \begin{cases} a + b(1-\delta) & \text{if } A \text{ is the standard} \\ a^2 + b^2 & \text{if there are incompatible standards} \\ a(1-\delta) + b & \text{if } B \text{ is the standard.} \end{cases} \tag{10.6}$$

Comparing these social welfare levels yields

Proposition 10.5 *If there are more consumers oriented toward brand A than there are consumers oriented toward brand B ($a > b$), then standardization on A is socially preferred to standardization on B.*

We now ask under what condition the incompatibility equilibrium outcome is socially preferred to a single-brand standardization. It follows from (10.6) that incompatibility is preferred over standardization on A, if $a^2 + b^2 > a + b - b\delta = 1 - b\delta$, or $\delta > \frac{1-a^2-b^2}{b}$. Using the fact that $b = 1-a$, we see that this last condition is equivalent to $\delta > 2a$, or $a < \frac{\delta}{2}$. Similarly, incompatibility is socially preferred over standardization on B if $b < \frac{\delta}{2}$. However, these conditions cannot both hold if $\delta < 1$ since in this case $a + b < \frac{\delta}{2} + \frac{\delta}{2} < 1$. Hence,

Proposition 10.6

1. *If the network preference effect is strong relative to the disutility from consuming the less preferred brand ($\delta < 1$), then the incompatibility equilibrium is socially inefficient.*
2. *If $\delta > 1$, incompatibility is socially optimal if $a < \frac{\delta}{2}$ and $b < \frac{\delta}{2}$.*

Is there a market failure?

We first ask whether standardization on a single-brand equilibrium may not be socially desirable. Proposition 10.3 shows that as long as $\delta < 1$, there are two equilibria in which the industry produces on a single standard. However, (10.6) implies that if there are more consumers oriented toward A, standardization on A socially dominates standardization on B. Hence,

Proposition 10.7 *An equilibrium in which the industry standardizes on the less socially preferred brand exists.*

However, note that in this case there is also a good equilibrium where the industry's standard is the more popular brand, so one can assume that with a minor coordination, consumers can choose the socially preferred standard.

How can it happen that an industry specializes on the wrong brand? Consider a dynamic scenario (which is not analyzed in this section) such that $a > b$ and brand B exists in the market before brand A attempts to enter the market. In this case, the firm producing brand A will not be able to enter the market. In the literature, this situation is generally described as a case where the existence of an *installed base* (brand B) has prevented the emergence of the more popular brand A.

We now seek to investigate whether a market failure can occur under the incompatibility equilibrium. Let us take an example: $a = b = 0.5$ and $\delta = 0.6$. Proposition 10.4 implies that incompatibility is an equilibrium since $1/2 > (1 - 0.6)/2 = 0.2$. However, since $\delta = 0.6 < 1$, Proposition 10.6 implies that incompatibility is inefficient. Hence,

Proposition 10.8 *An equilibrium in which the industry produces two incompatible brands need not be socially efficient.*

Finally, the opposite of Proposition 10.8 holds:

Proposition 10.9 *If incompatibility ($x_A = a$ and $x_B = b$) is efficient, then the incompatibility equilibrium exists and is unique.*

Proof. If incompatibility is efficient, then part 1 of Proposition 10.6 implies that $\delta > 1$. Since $a > 0$ and $b > 0$, Proposition 10.4 implies that incompatibility is an equilibrium. Also, Proposition 10.3 implies that an equilibrium where an industry is standardized on a single standard does not exist. ■

10.2 The Supporting Services Approach

The analysis of the previous section was based on the assumption that consumers' value for a product increases when other consumers purchase a compatible or an identical brand. However, despite the fact that the network-externalities assumption is intuitive and appealing for modeling products such as telecommunication systems, where the utility of each consumer is directly related to the network size, the models themselves do not explain why people behave this way. So the remaining question is whether "network effects" can prevail even without assuming that consumers' preferences exhibit network externalities.

We therefore turn now to models describing consumers who do not derive satisfaction from the consumption of other consumers. Instead, consumers gain satisfaction from the product itself and the variety of (brand-specific) complementary products that we call *supporting services*. The literature utilizing this approach includes Chou and Shy 1990, 1993, and 1996, and Church and Gandal 1992a,b, 1993. In many instances, supporting services are incompatible across brands. For examples, most software packages are designed to operate on one operating system (such as UNIX, DOS, Macintosh, OS, etc.) and do not operate on the other operating systems. Videotapes recorded on the NTSC television system (used in North America and Japan) cannot be played in Europe or in the Middle East, where the dominant television standard is PAL. For a discussion of the newly emerging high-definition television standards see Farrell and Shapiro 1992 and the references therein.

10.2.1 Network effects without network externalities

Consider consumers who can freely choose between two computer brands named brand A (short for Artichoke computers) and brand B (short for Banana computers). Each consumer is endowed with Y dollars to be spent on one unit of hardware and the variety of software written for the specific hardware purchased. We denote by p_i the price of computer brand i, $i = A, B$. Hence, given a total budget of Y, a consumer purchasing brand i spends $E_i \equiv Y - p_i$ on i's specific software.

We denote by N_i the total number of software packages that can be run on an i machine. The utility of a consumer purchasing system i is defined as an increasing function of the number of software packages compatible with machine i, $i = A, B$. Consumers are uniformly indexed by δ on the interval $[0, 1]$ according to their relative preference towards computer brand B. We define the utility of a consumer type δ as

$$U^\delta \equiv \begin{cases} (1-\delta)\sqrt{N_A} & \text{if she is an } A\text{-user} \\ \delta\sqrt{N_B} & \text{if she is a } B\text{-user.} \end{cases} \tag{10.7}$$

Thus, the utility function (10.7) describes preferences exhibiting "love for variety" of software. That is, a consumer's preferences toward a specific brand are affected by a fixed parameter, (δ or $(1-\delta)$), and by the number of software packages available for each brand, (N_A and N_B). Figure 10.4 illustrates how consumers are distributed according to their preferences toward the two brands.

Figure 10.4: Consumers' distribution of tastes

The consumer who is indifferent to the choice between system A and system B is denoted by $\hat{\delta}$, which is found from (10.7) by solving

$$(1-\hat{\delta})\sqrt{N_A} = \hat{\delta}\sqrt{N_B}. \tag{10.8}$$

Thus, in equilibrium, a consumer indexed by $\delta < \hat{\delta}$ is an A-user whereas a consumer indexed by $\delta > \hat{\delta}$ is a B-user. The total number of A-users is denoted by $\delta_A \equiv \hat{\delta}$, and the total number of B-users is given by $\delta_B \equiv (1-\hat{\delta})$. Altogether,

$$\frac{\delta_B}{\delta_A} = \frac{1-\hat{\delta}}{\hat{\delta}} = \sqrt{\frac{N_B}{N_A}}. \tag{10.9}$$

Hence,

Proposition 10.10 *The brand with the higher market share is supported by a larger variety of software. Formally, $\delta_A \ge \delta_B$ if and only if $N_A \ge N_B$.*

Proposition 10.10 confirms widely observed phenomena, for example, the Intel-based machines (PCs) have the largest market share and are supported by the largest variety of software compared to machines based on other chips.

The software industry

We have not yet discussed how the variety (number) of each brand-specific software is being determined in each software industry. Instead of fully modeling the software industry, we conjecture that the number of software packages supporting each machine should be proportional to the aggregate amount of money spent on each type of software. We therefore make the following assumption:

ASSUMPTION 10.1 *The number of software packages (variety) supporting each brand is proportional to the aggregate expenditure of the consumers purchasing the brand-specific software. Formally,*

$$N_A = \hat{\delta}E_A = \hat{\delta}(Y - p_A) \quad \textit{and} \quad N_B = (1-\hat{\delta})E_B = (1-\hat{\delta})(Y - p_B).$$

Substituting into (10.9) yields

$$\hat{\delta} = \frac{E_A}{E_A + E_B} = \frac{Y - p_A}{2Y - p_A - p_B}. \tag{10.10}$$

Network effects

The following proposition (part 4 in particular) demonstrates how network effects can prevail without assuming network externalities.

Proposition 10.11 *An increase in the price of hardware A (p_A) will*

1. *decrease the number of A-users (δ_A decreases);*
2. *increase the number of B-users (δ_B increases);*
3. *decrease the variety of software written for the A machine (N_A decreases) and increase the variety of B-software (N_B increases); and*

4. decrease the welfare of A-users and increase the welfare of B-users.

Proof. Part 1 follows from (10.10) since $\partial\hat{\delta}/\partial p_A < 0$. Part 2 immediately follows since $\delta_B = 1 - \delta_A$. Part 3 follows from Assumption 10.1 since as $\hat{\delta}$ decreases and p_A increases, it is implied that N_A must decrease while N_B must increase. Part 4 follows from (10.7), since a decrease in N_A decreases the utility of an A-user, whereas an increase in N_B increases the utility of a B-user. ■

When p_A increases, Assumption 10.1 implies that two factors exist that cause a reduction in the variety of A-software: First, the direct effect ($Y - p_A$ decreases), that is, A-users spend more on hardware and therefore less on software; and second, the indirect effect via a reduction in the number of A-users ($\hat{\delta}$ decreases). Assumption 10.1 also implies that N_B increases since there are more B-users.

Part 3 of Proposition 10.11 demonstrates the network effect generated by an increase in hardware price p_A on the welfare of B-users as follows:

$$p_A \uparrow \Longrightarrow \delta_A \downarrow \Longrightarrow \delta_B \uparrow \Longrightarrow N_B \uparrow \Longrightarrow U^{B-user} \uparrow .$$

That is, a decrease in the number of A-users causes an increase in the number of B-users, which in turn increases the variety of B-software, which increases the welfare and number of B-users, and so on.

10.2.2 Partial compatibility

Note that 100 percent compatibility is never observed. For example, you have probably noticed that sometimes you fail to transmit a fax to a remote fax machine because the other machine does not fully respond to all standards. You have probably also noticed that some record and tape players are not rotating at the same speed. Also, even when the manufacturer asserts that his computer (say) is DOS compatible, there are always some packages of software that can operate on one machine, but "refuse" to operate on another. In that sense, 100 percent compatibility is actually never observed.

Perhaps the main advantage of using the supporting-services approach to model network behavior is that it allows an easy interpretation for modeling the concept of partial compatibility.

DEFINITION 10.3 *A computer brand i is said to be* **partially compatible** *with a ρ_i ($0 \le \rho_i \le 1$) degree of compatibility with computer brand j if a fraction ρ_i of the total software written specifically for brand j can also be run on computer brand i.*

It should be pointed out that Definition 10.3 does not imply that compatibility is a symmetric relation. In other words, it is possible that

a computer of a certain brand is designed to be able to read software developed for rival machines, but the rival machines are not designed to read software not specifically designed for them. In the extreme case, in which $\rho_i = 1$ but $\rho_j = 0$ (machine i can read j software, but machine j cannot read i software), we say that the machines are *one-way compatible.*

The number of software packages written specifically for machine i is denoted by n_i, $i = A, B$. The main feature of this model is that the machines can be partially compatible in the sense that in addition to its own software, each machine can also run a selected number of software packages written for its rival machine. That is, ρ_i measures the proportion of machine j software that can be run on an i machine, $i, j = A, B$ and $i \neq j$. Therefore, the total number of software packages available to an i-machine user is equal to

$$N_A = n_A + \rho_A n_B \quad \text{and} \quad N_B = n_B + \rho_B n_A. \tag{10.11}$$

We will not develop the complete model. The complete computer and software industry equilibrium is developed in Chou and Shy 1993. Instead, in what follows we merely illustrate the main insights of this model.

Suppose that the software industry produces a positive variety of both types of software. That is, $n_A > 0$ and $n_B > 0$. Now, for the sake of illustration, let N_A and N_B be kept constants. Figure 10.5 shows the equilibrium n_A and n_B levels associated with the given N_A and N_B.

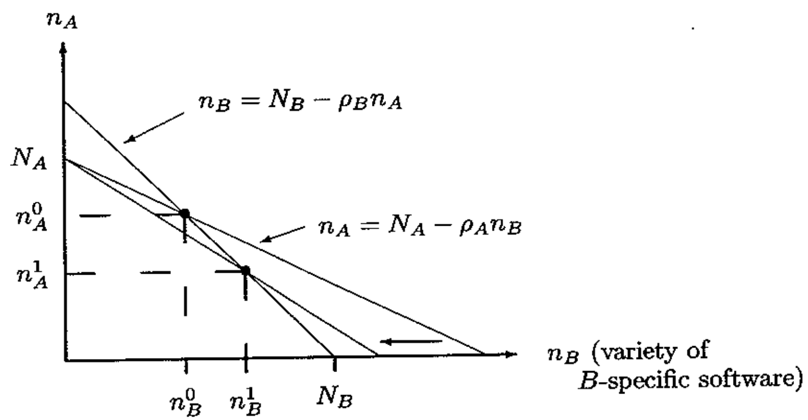

Figure 10.5: Equilibrium variety of brand-specific software

The line N_A shows the combinations of brand-specific software n_A and the rival brand-specific software n_B associated with a constant level N_A of A-usable software available to A-users, for a given level of compatibility ρ_A. Similarly, the line N_B shows all the n_A and n_B combinations associated with a constant level of B-usable software N_B. The point $\langle n_A^0, n_B^0 \rangle$ is the equilibrium variety of software written specifically for A and B machines.

Now, suppose that the producer of computer A makes its machine more compatible with B software (i.e., ρ_A increases). Hence, the line N_A tilts to the left because in order to keep the number of A-usable software at a constant level there is less need for A-specific software (since A-users can use more of B-software). Therefore, the new software-variety equilibrium is now given at the point $\langle n_A^1, n_B^1 \rangle$ in Figure 10.5. Consequently,

Proposition 10.12 *When there are two software industries, each producing brand-specific software, an increase in the degree of compatibility of the A-machine with the software written for the B-machine,*

1. *will reduce the variety of software specifically written for the A-machine (n_A decreases);*
2. *will increase the variety of software specifically written for the B-machine (n_B increases); and*
3. *will reduce the total variety of software available to A-users and will increase the total variety of software available to B-users (N_A decreases and N_B increases).*

The last part of the Proposition is proved in Chou and Shy 1993. The significance of the proposition (which was actually known to many computer makers a long time before it was known to economists) is that it shows that a computer manufacturer may refrain from making its machine more compatible with the software supporting the rival machine because compatibility with the rival machine's software will induce software writers to write more software for the rival machine (since part of it is usable for both machines), thereby making the rival machine more attractive to consumers. This result explains why computer manufacturers may choose different operating systems for their machines.

It should be pointed out that there could be reasons other than the one in Proposition 10.12 for why firms make their brand less compatible with other brands. For example, in subsection 12.2.2 we show other cases in which firms choose to differentiate themselves from other firms by producing products of different quality.

10.3 The Components Approach

In the previous sections we introduced two approaches to the economics of networks: (a) the network-externality approach, where a consumer's valuation of a certain brand is affected by the number of consumers purchasing a similar or an identical brand, and (b) the supporting-services approach, where a consumer's valuation of a brand is affected by the number of supporting services (supporting software) supporting the specific brand.

The components approach discussed in this section is similar to the supporting services approach in two aspects: First, it does not assume that consumers' preferences exhibit a consumption externality; second, it assumes complementarity, in the sense that just as computers yield no utility without the supporting software, the basic computer component does not yield utility without a complementary monitor component.

10.3.1 The basic model

The components models were first introduced in Matutes and Regibeau 1988 and Economides 1989.

The product

Consider a product that can be decomposed into two (perfect complements) components. For example, a computer system can be decomposed into a basic unit and a monitor. The basic unit and the monitor are perfect complements since a consumer cannot use one component without using the other. Another example is a stereo system, which is generally decomposed into an amplifier and speakers.

We denote the first component (the basic unit) by X and the second component (the monitor) by Y.

Firms and Compatibility

There are two firms capable of producing both components, which can be assembled into systems. We denote by X_A the first component produced by firm A, and by Y_A the second component produced by firm A. Similarly, firm B produces components X_B and Y_B. With no loss of generality, we simplify by assuming that production is costless.

Turning to compatibility, we can readily see that since the components are perfect complements, each consumer must purchase one unit of X with one unit of Y. The question of compatibility here is whether a consumer can combine components from different manufacturers when he or she purchases and assembles the system. Formally,

DEFINITION 10.4

1. *The components are said to be* **incompatible** *if the components produced by different manufacturers cannot be assembled into systems. That is, systems X_AY_B and X_BY_A do not exist in the market.*

2. *The components are said to be* **compatible** *if components produced by different manufacturers can be assembled into systems. That is, X_AY_B and X_BY_A are available in the market.*

Consumers

There are three consumers, denoted by AA, AB, and BB, with heterogeneous preferences toward systems. We denote by p_i^x and p_i^y the price of component X and component Y produced by firm i, respectively, $i = A, B$.

Each consumer has an ideal combination of components. That is, if $p_A^x = p_B^x$ and $p_A^y = p_B^y$, then consumer AA would always choose system X_AY_A over X_BY_B, consumer BB would choose system X_BY_B over X_AY_A, and if the systems are compatible (see Definition 10.4), then consumer AB would choose system X_AY_B.

A consumer who purchases system X_iY_j would pay a total price of $p_i^x + p_j^y$ for this system, $i, j = A, B$. We denote by U_{ij} the utility level of consumer ij, whose ideal system is X_iY_j, $ij \in \{AA, AB, BB\}$, and assume that for $\lambda > 0$

$$U_{ij} \equiv \begin{cases} 2\lambda - (p_i^x + p_j^y) & \text{if purchasing system } X_iY_j \\ \lambda - (p_j^x + p_j^y) & \text{if purchasing system } X_jY_j \\ \lambda - (p_i^x + p_i^y) & \text{if purchasing system } X_iY_i \\ -(p_j^x + p_i^y) & \text{if purchasing system } X_jY_i \\ 0 & \text{if does not purchase any system.} \end{cases} \tag{10.12}$$

Thus, in this simple model each consumer has a different ideal system (under equal prices). The utility function (10.12) shows that a consumer purchasing his ideal system gains a (net of prices) utility level of 2λ. If the system he buys has one component from his ideal system and one component from his less preferred system, his (net of prices) utility level is reduced by λ. Finally, a consumer who purchases a system in which both components are produced by his less preferred manufacturer has a (net of prices) utility level of 0. Clearly, given the threshold utility level of 0, no system will be purchased unless its total cost is lower than 2λ.

10.3.2 Incompatible systems

Suppose that the components produced by different manufacturers are incompatible (see Definition 10.4), so that only two systems are produced: system X_AY_A and system X_BY_B. We denote by q_i the number of systems sold by firm i, and by p_i the price of system i (both components), $i = A, B$. That is, the price of system X_AY_A is $p_A \equiv p_A^x + p_A^y$ and the price of system X_BY_B is $p_B \equiv p_B^x + p_B^y$. Thus, the profit function of firm i is $\pi_i = p_iq_i$, $i = A, B$. We look for a Nash-Bertrand equilibrium in prices. Formally,

DEFINITION 10.5 *An incompatible-components equilibrium is a pair of price p_A^I and p_B^I, a pair of quantities q_A^I and q_B^I such that for a given p_j^I, firm i chooses p_i^I to $\max_{p_i} \pi_i(p_i, p_j^I)$* s.t. *$q_i =$ number of consumers maximizing (10.12) by choosing system i ($i, j = A, B$, $i \neq j$).*

Before characterizing the equilibria we can show that

Lemma 10.1 *There does not exist an equilibrium where one firm sells to all consumers.*

Proof. If firm A sells to all customers, then it must set $p_A = 0$. But even at this price, if for $\epsilon > 0$ sufficiently small, firm B sets $p_B = \epsilon$, consumer BB would purchase system X_BY_B. ■

What Lemma 10.1 tells us is that if an equilibrium exists, then consumer AA buys a system from firm A and consumer BB buys from firm B. Therefore,

Proposition 10.13 *There exist three equilibria:*
In one equilibrium firm A sells system X_AY_A to consumers AA and AB while firm B sells system X_BY_B to consumer BB. In this equilibrium, $p_A^I = \lambda$, $q_A^I = 2$, $p_B^I = 2\lambda$, $q_B^I = 1$.

In the second equilibrium, firm B sells system X_BY_B to consumers BB and AB while firm A sells system X_AY_A to consumer AA. In the second equilibrium, $p_A^I = 2\lambda$, $q_A^I = 1$, $p_B^I = \lambda$, $q_B^I = 2$.

In the third equilibrium, firm A sells system X_AY_A to consumer AA, firm B sells system X_BY_B to consumer BB, and consumer AB is not served. In this equilibrium, $p_A^I = p_B^I = 2\lambda$ and $q_A^I = q_B^I = 1$.

In any equilibrium the firms' profit levels are given by $\pi_A^I = \pi_B^I = 2\lambda$.

Proof. Since the first two equilibria are symmetric, it is sufficient to look at the first equilibrium. We have to show that firm A cannot increase its profit by reducing its price to a level at which it would sell to all three consumers (undercutting firm B). That is,

$$\pi_A^I = 2p_A^I \geq 3(p_B^I - 2\lambda). \tag{10.13}$$

Similarly, we have to show that firm B cannot increase its profit by reducing its price p_B to p_A where it would sell to two consumers BB and AB.

$$\pi_B^I = p_B^I \geq 2p_A^I. \tag{10.14}$$

In fact, one should also check a third possibility in which firm B deviates by reducing the price to a level where all the three consumers purchase system X_BY_B. However, such a deviation is not profitable since firm B has to set $p_B = p_A^I - 2\lambda = \lambda - 2\lambda < 0$.

First, note that our candidate equilibrium prices satisfy equations (10.13) and (10.14), so no firm would find it profitable to reduce its price.

Second, no firm could profitably deviate by raising its price since if firm B raises its price above 2λ, consumer BB will not purchase system X_BY_B. Similarly, if firm A raises its price above λ, consumer AB will not purchase any system.

We still have to show that consumers AA, AB, and BA maximize their utility(10.12) by choosing system AA and that consumer BB maximizes utility by choosing system X_BY_B. To do that, we need to calculate the equilibrium utility levels of all customers. Thus, in equilibrium we have it that

$$U_{AA}^I = 2\lambda - p_A^I = \lambda; \quad U_{BB}^I = 2\lambda - p_B^I = 0; \quad U_{AB}^I = \lambda - \lambda = 0. \tag{10.15}$$

It is easy to verify that consumer AA would not purchase system BB since system BB would yield a utility level of $-p_B = -2\lambda < U_{AA}^I$. Similarly, consumer BB would not purchase system AA since system AA would yield a utility level of $-p_A = -\lambda < U_{BB}^I$. Also, consumer AB would not purchase system BB since $p_B^I > p_A^I$, and both yield a (net of prices) utility level of λ.

Finally, to show that $p_A^I = p_B^I = 2\lambda$ constitute (the third) equilibrium, note that if, say, firm A reduces its price to $p_A = \lambda$, consumer AB buys system AA, and we have the first equilibrium. Since in all equilibria, firm A's profit level is $\pi_A^I = 2\lambda$, a deviation will not occur. ■

We define the consumer surplus as the sum of consumers' utilities. Hence,

$$CS^I \equiv U_{AA}^I + U_{BB}^I + U_{AB}^I = \lambda. \tag{10.16}$$

We define the economy's welfare as the sum of firms' profit levels and consumer surplus. Thus,

$$W^I \equiv \pi_A^I + \pi_B^I + CS^I = 2p_A^I + p_B^I + CS^I = 2\lambda + 2\lambda + \lambda = 5\lambda. \tag{10.17}$$

The equilibrium social-welfare level given in (10.17) is simply the sum of the (net of prices) utility levels of all the consumers, which equals twice 2λ for consumers AA and BB, who consume their ideal systems, and λ for consumer AB, who purchases the system X_AY_A but whose ideal Y component is Y_B.

10.3.3 Compatible systems

When firms design their components to be compatible with components produced by the rival firm, two more systems become available to consumers: system X_AY_B and system X_BY_A. We look for an equilibrium where each consumer buys (assembles) his ideal system. In this equilibrium, each firm i sells two units of component X_i and two units of component Y_i, $i = A, B$.

DEFINITION 10.6 *A compatible components equilibrium is the set of component prices* $p_A^x, p_A^y, p_B^x, p_B^y$ *and quantities of components sold by each firm* $q_A^x, q_A^y, q_B^x, q_B^y$ *such that for given* p_j^x *and* p_j^y, *firm* i *chooses* p_i^x *and* p_i^y *to* $\max \pi_i(p_i^x, p_i^y, p_j^x, p_j^y)$ s.t. q_i^x *and* q_i^y *are the number of consumers maximizing (10.12) by choosing components* X_i, Y_i, *respectively.*

Proposition 10.14 *There exists an equilibrium where each consumer purchases his ideal system. In this equilibrium all components are equally priced at* $p_A^x = p_A^y = p_B^x = p_B^y = \lambda$, *and a firm's profit levels are* $\pi_A^c = \pi_B^c = 3\lambda$.

Proof. Since firm A sells two components of X and one component of Y, while firm B sells two components of Y and one component of X, equilibrium prices should be at levels so that firms could not profitably reduce the price of one component in order to sell this component to additional customers. For example, in equilibrium, firm A sells component X_A to consumers AA, and AB. Reducing the price of Y_A to $p_B^y - \lambda$ would induce consumer BB to buy component Y from firm A (note that in order to attract consumers from the competing firms, the price reduction should be at least λ). However, reducing a component price to zero cannot constitute a profit-maximizing deviation. By symmetry, firm B will not find it profitable to reduce its price to $p_A^x - \lambda = 0$.

Finally, since all prices are equal each consumer purchases his ideal brand, yielding equilibrium utility levels of

$$u_{AA}^c = u_{AB}^c = u_{BB}^c = 2\lambda - \lambda - \lambda = 0. \tag{10.18}$$

For this reason, no firm would find it profitable to increase a component's price since each consumer would not pay more than 2λ for a system. ■

Hence, when all components are compatible, the (aggregate) consumer surplus, firms' profit levels, and the social welfare level are given by

$$CS^c = 0; \quad \pi_A^c = \pi_B^c = 3\lambda; \quad W^c \equiv \pi_A^c + \pi_B^c + CS^c = 6\lambda. \qquad (10.19)$$

Like equation (10.17), equation (10.19) demonstrates that the social welfare is the sum of the (net of prices) utility levels.

10.3.4 Compatibility versus incompatibility

We now wish to examine the effects of components compatibility on firms' profit and consumers' utility levels, aggregate consumers' surplus, and the social welfare. Comparing (10.15) with (10.18) yields

Proposition 10.15 *Consumers are never better off when the firms produce compatible components than when firms produce incompatible components.*

However, comparing Propositions 10.13 with 10.14 yields

Proposition 10.16 *All firms make higher profits when they produce compatible components than when they produce incompatible components.*

Also, comparing (10.17) with (10.19) yields

Proposition 10.17 *Social welfare is higher when firms produce compatible components.*

In order to explain Proposition 10.15 we need to compare the systems' prices under the compatibility and incompatibility regimes (given in Propositions 10.13 and 10.14). Under the incompatibility regime, two consumers pay each λ for the system they buy. Under compatibility, each consumer pays 2λ for each system. Hence, total consumer expenditure under compatibility exceeds the expenditure under incompatibility by 2λ, but the (net of prices) utility level of consumer AB (the "mixing" consumer) rises by only λ. Thus, firms extract a surplus that exceeds the aggregate utility gains from compatibility, thereby reducing aggregate consumer surplus under the compatibility regime.

Proposition 10.16 can be explained by the following: First, under compatibility the mixing consumer is willing to pay more because he can now buy his ideal system. Second, compatibility reduces price competition between the component-producing firms since under incompatibility both firms are forced to lower the price of their system in order to attract the mixing consumer to choose their systems, given that the systems are not ideal for this consumer. This competition is relaxed when the components are compatible.

Finally, Proposition 10.17 shows that the welfare gains derived from having firms increase their profits by making their components compatible exceeds the welfare loss to consumers from the high component prices under compatibility.

10.3.5 How firms design their components

Proposition 10.16 shows that firms collect higher profits when all components are compatible with the components produced by the rival firms than they collect when firms produce incompatible components. We now ask whether an outcome where both firms choose to produce compatible components can be realized as an equilibrium for game in which firms choose both prices and the design of the components.

Consider a two-stage game where in period 1 firms choose whether to design their components to be compatible with the components produced by the rival firm. In period 2, given the design of the components, firms compete in prices, as described in subsections 10.3.2 and 10.3.3.

The subgame perfect equilibrium for this game turns out to be very simple because the compatibility decision by one firm forces an externality on the rival firm, in the sense that the compatibility of components is a symmetric relation, meaning that if component X_A is compatible with component Y_B, then component Y_B is compatible with component X_A. In other words, the market effect of having firm A make its X_A component compatible with component Y_B is equivalent to having firm B make its Y_B component compatible with X_A. Similarly, the outcome in which firm B makes its X_B component compatible with firm A's Y_A component is equivalent to firm A's making its Y_A component compatible with B's X_B component. It is important to note that this externality is a feature of the component approach discussed here, but it does not occur in the supporting-services approach (see Definition 10.3). That is, in the supporting-services approach we can have it that machine A reads B's software, but not the other way around.

Thus, given this externality feature of the components approach, it is sufficient for one firm to decide on compatibility to produce a market outcome identical to that which would result from both firms deciding on compatibility. Therefore, Proposition 10.16 implies that

Proposition 10.18 *In the two-stage game, a subgame perfect equilibrium yields compatible components.*

10.4 Exercises

1. Consider the supporting-services approach model developed in subsection 10.2.
   (a) For a given hardware price of brand A, p_A, what is the price of computer B beyond which firm B would have a zero market share?
   (b) Suppose that $p_A > p_B$, and suppose that the income of each consumer doubles to $2Y$, while hardware prices remain unchanged. Calculate the effect this increase in incomes on (i) the market shares (δ_A and δ_B), and on (ii) the ratio of the number of software packages written for computer A to the number of software packages written for computer B.
2. Consider the component approach analyzed in subsection 10.3, but assume that there are four consumers: consumer AA, consumer BB, consumer AB, and consumer BA.
   (a) If the components are incompatible, prove that no Nash-Bertrand equilibrium in system prices p_A and p_B (as defined in Definition 10.5) exists.
   (b) If the components are compatible, calculate the symmetric equilibrium prices of all components, firms' profit levels, and consumers' surplus.

10.5 References

Chou, C., and O. Shy. 1990. "Network Effects without Network Externalities." *International Journal of Industrial Organization* 8: 259–270.

Chou, C., and O. Shy. 1993. "Partial Compatibility and Supporting Services." *Economics Letters* 41: 193–197.

Chou, C., and O. Shy. 1996. "Do Consumers Gain or Lose When More People Buy the Same Brand?" *European Journal of Political Economy* 12: 309–330.

Church, J., and N. Gandal. 1992a. "Integration, Complementary Products and Variety." *Journal of Economics and management Strategy* 1: 651–676.

Church, J., and N. Gandal. 1992b. "Network Effects, Software Provision, and Standardization." *Journal of Industrial Economics* 40: 85–104.

Church, J., and N. Gandal. 1993. "Complementary Network Externalities and Technological Adoption." *International Journal of Industrial Organization* 11: 239–260.

Conner, K., and R. Rumelt. 1991. "Software Piracy: An Analysis of Protection Strategies." *Management Science* 37: 125–139.

David, P. 1985. "Clio and the Economics of QWERTY." *American Economic Review* 75: 332–336.

David, P., and S. Greenstein. 1990. "The Economics of Compatibility Standards: An Introduction to Recent Research." *Economics of Innovation and New Technology* 1: 3–42.

Economides, N. 1989. "Desirability of Compatibility in the Absence of Network Externalities." *American Economic Review* 79: 1165–1181.

Farrell, J., and G. Saloner. 1986. "Standardization and Variety." *Economics Letters* 20: 71–74.

Farrell, J., and G. Saloner. 1987. "The Economics of Horses, Penguins, and Lemmings." In *Production Standardization and Competitive Strategies,* edited by L. G. Gable. Amsterdam: North-Holland.

Farrell, J., and C. Shapiro. 1992. "Standard Setting in High-Definition Television." *Brookings Papers on Economic Activity: Microeconomics,* 1–93.

Gabel, L. 1991. *Competitive Strategies for Product Standards.* London: McGraw Hill.

Gandal, N. 1994. "Hedonic Price Indexes for Spreadsheets and an Empirical Test of Network Externalities." *RAND Journal of Economics* 25: 160–170.

Katz, M., and C. Shapiro. 1985. "Network Externalities, Competition, and Compatibility." *American Economic Review* 75: 424–440.

Katz, M., and C. Shapiro. 1986. "Technology Adoption in the Presence of Network Externalities." *Journal of Political Economy* 94: 822–841.

Kindleberger, C. 1983. "Standards as Public, Collective and Private Goods." *KYKLOS* 36: 377–396.

Matutes, C., and P. Regibeau. 1988. "Mix and Match: Product Compatibility Without Network Externalities." *RAND Journal of Economics* 19: 221–234.

Rohlfs, J. 1974. "A Theory of Interdependent Demand for a Communication Service." *Bell Journal of Economics* 5: 16–37.

PART IV

Marketing

Chapter 11

Advertising

> Hardly any business practice causes economists greater uneasiness than advertising.
> —L. Telser, "Advertising and Competition"

Advertising is an integral part of our life. Each one of us is constantly bombarded by advertising for products and services in a wide variety of forms. We watch advertising on TV, listen to advertising on the radio, read ads in newspapers, in magazines, on outdoor billboards, on buses and trains, receive a large amount of so-called junk mail, and we transmit advertising via word-of-mouth and by wearing brand-name labels on our clothes.

Despite this basic observation, very little is understood about the effects of advertising. Advertising is generally defined as a form of providing information about prices, quality, and location of goods and services. Advertising differs from other forms of information transmissions (like stock-exchange data and guidebooks) in two respects: First, the information is transmitted by the body who sells the product, and second, the buyer does not always have to pay to receive the information (or pays a little with his or her value of time of watch a TV ad or to sort out the relevant ads in the Sunday newspaper).

What is the purpose of advertising? We first need to acknowledge that advertising must serve a purpose for some agents in the economy since—as a matter of fact—firms, governments, and individuals spend large sums of money on advertising. It is generally estimated that developed economies spend more than 2 percent of their GNPs on advertising (see Schmalensee 1972, 1986). The expenditure of firms on advertising is generally measured in terms of advertising expenditure divided by the value of sales. These ratios vary drastically across products and industries. The ratio of advertising expenditure to sales of vegetables may be as low as 0.1 percent, whereas for cosmetics or detergents, this ratio may be as high as 20 to 60 percent.

There have been many attempts to correlate industry types, product characteristics, geographical locations, and other characteristics with advertising-to-sales ratios. However, in most cases advertising still remains a mystery since neither empirically nor theoretically can we explain why different firms spend different amounts on advertising. For example, Adams and Brock (1990) report that the Big Three car producers in the United States, which are ranked among the largest advertisers in the country, happen to have different advertising-to-sales ratios. In 1986 the largest producer, GM (which spent \$285 million on advertising), spent \$63 per car, whereas Ford spent \$130 and Chrysler spent \$113 per car (though they spent less overall than GM). This may hint of economies of scale in car advertising.

Earlier modern authors, e.g. Kaldor (1950), held the idea that advertising is "manipulative" and reduces competition and therefore reduces welfare for two reasons: First, advertising would persuade consumers to believe wrongly that identical products are differentiated because the decision of which brand to purchase depends on consumers' perception of what the brand is rather than on the actual physical characteristics of the product. Therefore, prices of heavily advertised products would rise far beyond their cost of production. Second, advertising serves as an entry-deterring mechanism since any newly entering firm must extensively advertise in order to surpass the reputation of the existing firms. Thus, existing firms use advertising as an entry-deterrence strategy and can maintain their dominance while keeping above-normal profit levels.

More recent authors, Telser (1964), Nelson (1970, 1974), and Demsetz (1979), proposed that advertising serves as a tool for transmitting information from producers to consumers about differentiated brands, thereby reducing consumers' cost of obtaining information about where to purchase their most preferred brand.

Nelson (1970) distinguishes between two types of goods: *search goods* and *experience goods*. Consumers can identify the quality and other characteristics of the product before the actual purchase of search goods. Examples include tomatoes or shirts. Consumers cannot learn the quality and other characteristics of experience goods before the actual purchase. Examples include new models of cars and many electrical appliances with unknown durability and failure rates. Note that this distinction is not really clear-cut, since we cannot fully judge the quality of a tomato until we eat it, and we cannot fully judge the quality of a shirt until after the first wash!

What Nelson claims is that the effects of advertising may vary between these two groups of products, because consumers do not depend

on information obtained from the manufacturers concerning search products (since consumers find it by themselves). However, consumers do rely on advertisements when they purchase experience goods. Several tests have also confirmed that advertising of experienced products is more intensive (in terms of the ratio of advertising expenditure to sales) than advertising of search goods.

The economics literature distinguishes between two types of advertising: *persuasive advertising* and *informative advertising*. Persuasive advertising intends to enhance consumer tastes for a certain product, whereas informative advertising carries basic product information such as characteristics, prices, and where to buy it. In the following two subsections we analyze these two types of advertising and ask whether from a social welfare point of view, firms engage in too little or too much advertising.

11.1 Persuasive Advertising

In this subsection we analyze persuasive advertising. That is, advertising that boosts the industry demand for the advertised product(s). We first investigate what the optimal advertising level is, assuming that the demand for the good is monotonically increasing with the firm's advertising level. Then, we ask whether from a social welfare point of view there is too much or too little advertising.

11.1.1 The monopoly's profit-maximizing level of advertising

Consider a monopoly firm selling a single product in a market where the demand curve is given by

$$Q(A,p) = \beta A^{\epsilon_A} p^{\epsilon_p}, \quad \text{where } \beta > 0,\ 0 < \epsilon_A < 1, \quad \text{and} \quad \epsilon_p < -1. \tag{11.1}$$

The parameter A denotes the firm's expenditure on advertising, Q and p denote the quantity demanded and the price for this product. Thus, the quantity demanded is monotonically increasing with the level of advertising (A) but at a decreasing rate (since $\epsilon_A < 1$).

Denoting by $\eta_A(A,p)$ and $\eta_p(A,p)$ the demand advertising elasticity and price elasticity respectively, and recalling subsection 3.2.1, where we showed the exponents of the variables in an exponential demand function (illustrated in Figure 3.4) are the elasticities of the corresponding variables, the reader can verify that

$$\eta_A \equiv \frac{\partial Q(A,p)}{\partial A}\frac{A}{Q} = \epsilon_A \quad \text{and} \quad \eta_p \equiv \frac{\partial Q(A,p)}{\partial p}\frac{p}{Q} = \epsilon_p. \tag{11.2}$$

Let c denote the unit cost of the product. The monopoly has two choice variables: the price (p) and the advertising expenditure (A). Thus, the monopoly solves

$$\max_{A,p} \pi(A,p) \equiv pQ - cQ - A = \beta A^{\epsilon_A} p^{\epsilon_p+1} - c\beta A^{\epsilon_A} p^{\epsilon_p} - A. \quad (11.3)$$

The first-order condition with respect to price is given by

$$0 = \frac{\partial \pi(A,p)}{\partial p} = \beta A^{\epsilon_A}(\epsilon_p + 1)p^{\epsilon_p} - c\beta A^{\epsilon_A}\epsilon_p p^{\epsilon_p - 1}, \quad (11.4)$$

implying that

$$p^M = \frac{\epsilon_p}{\epsilon_p + 1}c \text{ and hence } \frac{p^M - c}{p^M} = \frac{-1}{\epsilon_p}. \quad (11.5)$$

The first-order condition with respect to advertising level is given by

$$0 = \frac{\partial \pi(A,p)}{\partial A} = \beta \epsilon_A A^{\epsilon_A - 1} p^{\epsilon_p}(p - c) - 1, \quad (11.6)$$

implying that

$$\frac{p^M - c}{p^M} = \frac{1}{\beta \epsilon_A A^{\epsilon_A - 1} p^{\epsilon_p + 1}}. \quad (11.7)$$

Equating equations (11.5) with (11.7) yields

$$\frac{\epsilon_A}{-\epsilon_p} = \frac{1}{\beta A^{\epsilon_A - 1} p^{\epsilon_p + 1}} = \frac{A^M}{p^M Q^M}, \text{ where } Q^M \equiv Q(p^M). \quad (11.8)$$

Equation (11.8) is known as the Dorfman-Steiner (1954) condition. Therefore,

Proposition 11.1 *A monopoly's profit-maximizing advertising and price levels should be set so that the ratio of advertising expenditure to revenue equals the (absolute value of the) ratio of the advertising elasticity to price elasticity. Formally,*

$$\frac{A^M}{p^M Q^M} = \frac{\epsilon_A}{-\epsilon_p}.$$

Thus, a monopoly would increase its advertising-to-sales ratio as the demand becomes more elastic with respect to the advertising (ϵ_A is close to 1), or less elastic with respect to price (ϵ_p is close to zero).

11.1.2 Too much or too little persuasive advertising?

Persuasive advertising was defined as a method of information transmission that boosts the demand for the advertised product. Thus, persuasive advertising makes the good attractive to consumers and therefore has the potential to increase welfare. This does not imply that persuasive advertising must be truthful. All that persuasive advertising does is to provide an image for the product that would induce the consumer to purchase the product in order to identify with the message or people portrayed in the ads.

Dixit and Norman (1978) have proposed an extremely simple method for evaluating the welfare effect of persuasive advertising. Consider a simplified version of the demand function (11.1) where $\beta = 64$, $\epsilon_A = 0.5$, and $\epsilon_p = -2$. For this case, we assume that

$$Q = 64\sqrt{A}p^{-2} \quad \text{or} \quad p = \frac{8A^{1/4}}{Q^{1/2}}. \quad (11.9)$$

Taking the unit production cost to equal $c = 1$, the monopolist chooses p^M and A^M to maximize

$$\max_{A,p} \pi(A,p) = pQ - 1Q - A = 64A^{1/2}p^{-1} - 64A^{1/2}p^{-2} - A. \quad (11.10)$$

The first-order condition with respect to p is given by

$$0 = \frac{\partial \pi(A,p)}{\partial p} = \frac{-64\sqrt{A}}{p^2} + \frac{128\sqrt{A}}{p^3}, \quad (11.11)$$

implying that $p^M = 2$ and hence, $Q^M = 16\sqrt{A}$. Since the demand function has a constant elasticity, the monopoly price is independent of the level of advertising. The first-order condition with respect to A is given by

$$0 = \frac{\partial \pi(A,p)}{\partial A} = \frac{64}{2\sqrt{A}p} - \frac{64}{2\sqrt{A}p^2} - 1, \quad (11.12)$$

implying that $A^M = 64$ and hence, $Q^M = 16\sqrt{64} = 128$.

In order to check whether the monopoly advertises at the socially optimal level we first need to calculate the consumer surplus associated with each advertising level. The shaded area in Figure 11.1 shows the consumer surplus associated with a given advertising level A and the monopoly price $p^M = 2$. Hence, for a given advertising level A, the consumer surplus is given by

$$\begin{aligned} CS(A) &= \int_0^{16\sqrt{A}} 8\frac{\sqrt[4]{A}}{\sqrt{Q}}dQ - 2 \times 16\sqrt{A} \quad (11.13) \\ &= 2 \times 8\sqrt[4]{A}\left[Q^{1/2}\right]_0^{16\sqrt{A}} - 32\sqrt{A} = 32\sqrt{A}. \end{aligned}$$

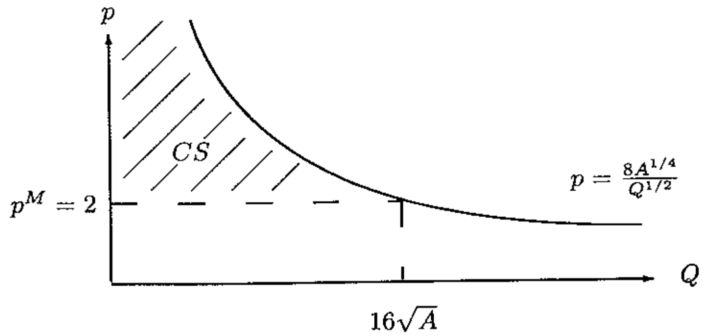

Figure 11.1: Consumer surplus for a given persuasive-advertising level

Assuming a monopoly price of $p^M = 2$, the firm's profit level as a function of the level of advertising is given by

$$\pi(A,2) = 2Q(A) - Q(A) - A = 32\sqrt{A} - 16\sqrt{A} - A = 16\sqrt{A} - A. \quad (11.14)$$

The social planner takes the monopoly price $p^M = 2$ as given and chooses an advertising level A^* to

$$\max_A W(A) \equiv CS(A) + \pi(A,2) = 48\sqrt{A} - A. \quad (11.15)$$

The first-order condition is given by

$$0 = \frac{\partial W(A)}{\partial A} = \frac{24}{\sqrt{A}} - 1. \quad (11.16)$$

Hence, the socially optimal advertising level is $A^* = 24^2 > 64 = A^M$. Notice that this social optimum is not a "first-best" optimum, since a first-best optimum requires marginal cost pricing. Hence,

Proposition 11.2 *Given a monopoly market structure, the equilibrium level of persuasive advertising is below the socially optimal level.*

Finally, the model presented in this section is very special and is given for the purpose of introducing one method for evaluating the welfare effects of persuasive advertising. We note here several problems concerning the robustness of Proposition 11.2. First, is it appropriate to use the consumer surplus as a welfare measure when the demand (utility) is affected by the advertising level? Second, even if this measure is appropriate, since the model is a partial equilibrium one, the measure

does not capture the entire welfare effect associated with an increase in the demand for the advertised product. That is, an increase in the demand for one product would decrease the demand for other products (say, for substitute products). Hence, the change in consumers' surplus in other markets should be taken into account.

11.2 Informative Advertising

Consumers often rely on information for their purchases. Without advertising, few consumers would be exposed to the variety of existing products, the price distribution, and the location of specific products. As Nelson points out, advertising can serve as a tool for transmitting this information to consumers and therefore should not be considered as an unnecessary activity. In fact, Benham (1972) has shown that prices are lower in markets where prices of eyeglasses are advertised than in markets where prices are not advertised.

The literature investigating the welfare effects of informative advertising concentrates on the conventional question of whether there is too little or too much informative advertising. Butters (1977) develops a model in which firms advertise the price of a homogeneous product and finds that the aggregate advertising level determined in a monopolistic competition equilibrium is socially optimal. Thus, Butters shows that informative advertising need not always be detrimental. Grossman and Shapiro (1984) consider a world of product differentiation where consumers who are located on the circumference of a circle (see subsection 7.3.2) are able to recognize a brand only if the producer advertises. This model provides ambiguous results about the excessiveness of informative advertising. Thus, the literature demonstrates that whether informative advertising is excessive or not depends on the specific functional form used for describing the industry. Recently, Meurer and Stahl (1994) developed a model in which some consumers are informed about two differentiated products and some are not, and in which both advertising and prices are choice variables. They show that social welfare may increase or decrease, depending on the level of advertising.

We proceed by developing a very simple model to analyze this question. Obviously, the answer that will be given here is not robust. However, the purpose of developing this model is to present one approach for how to model this type of question.

Consider a single-consumer, single-product market. Let p be the price of the product and assume that p is exogenously given (e.g., p is regulated). Let m denote the consumer's benefit from purchasing one unit of the product. Altogether, we assume that the utility function of the consumer is given by

$$u = \begin{cases} m - p & \text{if he purchases the product} \\ 0 & \text{if does not purchase.} \end{cases} \quad (11.17)$$

There are two firms producing the same product and offering it for sale at a price of p. With no loss of generality, assume that production is costless so that the only cost firms have to bear is the cost of sending an advertisement to the consumer. Formally, assume that each firm has a single decision variable, which is whether or not to advertise. The cost of advertising is given by a constant denoted by A.

The consumer may receive a total of 0, 1, or 2 ads from the firms. If the consumer receives one ad, he buys the product from the firm that sent it. If he receives no ads, he buys none, and if he receives two ads, he splits the transaction equally between the firms, that is, he pays $p/2$ to each firm. Note that this assumption is similar to the assumption that the consumer flips a coin when he receives two ads, thereby yielding an expected revenue of $p/2$ to each firm. Therefore, the profit of firm i, $i = 1, 2$ is given by

$$\pi_i = \begin{cases} p - A & \text{if only firm } i\text{'s ad is received} \\ \frac{p}{2} - A & \text{if both firms' ads are received} \\ -A & \text{if firm } i \text{ sends an ad, but the ad is not received} \\ 0 & \text{if firm } i \text{ does not advertise (and hence does not sell).} \end{cases} \quad (11.18)$$

The fact that a firm sends an ad does not imply that the consumer will indeed receive it. For instance, even if the firm invests \$$A$ in a TV ad, it is possible that the consumer will not be watching TV at the time that the ad runs on the air. Formally, let δ, $0 < \delta < 1$, be the probability that a message sent by a certain firm would be received by the consumer. Therefore, the expected profit of firm i, $i = 1, 2$, is given by

$$E\pi_i = \begin{cases} \delta(1-\delta)(p - A) + \delta^2(\frac{p}{2} - A) - (1-\delta)A & \text{both advertise} \\ \delta(p - A) - (1-\delta)A & \text{only } i \text{ advertises} \\ 0 & i \text{ does not.} \end{cases} \quad (11.19)$$

Comparing the expected profits in the first and second rows in (11.19) to the reservation profit of 0 yields:

Proposition 11.3 *For a given value of p, $p \le m$,*

1. *at least one firm will engage in advertising if and only if*

$$\frac{p}{A} \ge \frac{1}{\delta};$$

2. *two firms will engage in advertising if*

$$\frac{p}{A} \ge \frac{2}{\delta(2-\delta)}.$$

Figure 11.2 illustrates the combinations of the receiving probability parameter(δ) and the ratio of price to advertising cost (p/A) associated with having no firm, one firm, or two firms placing ads. Clearly,

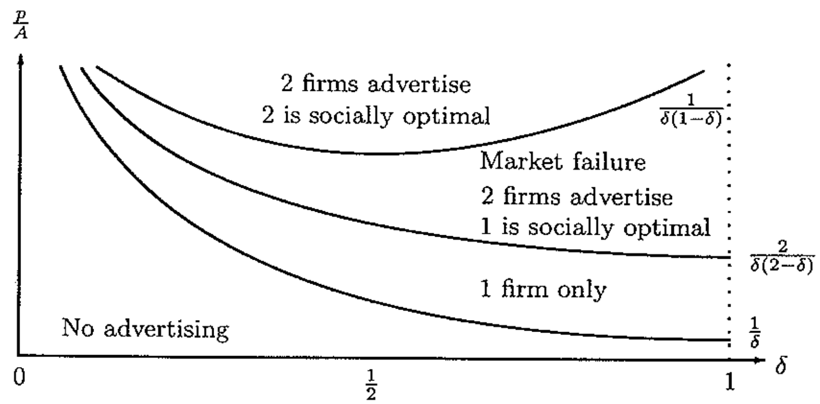

Figure 11.2: Equilibrium and socially-optimal amount of advertising

for a low receiving probability δ, or for a high advertising cost relative to the price (low p/A), no firm would place an ad. As either δ or p/A increase, the number of firms placing ads also increases.

We now turn to the welfare analysis. The problem solved by the social planner is to choose the number of firms that advertise in order to maximize the expected sum of consumer surplus and firms' profits. First, observe that if both firms advertise, the probability that at least one firm would sell is $2\delta(1-\delta) + \delta^2 = \delta(2-\delta)$ (which is twice the probability that one ad will be received while the other will not, plus the probability that both ads are received). Formally, the expected social welfare as a function of the number of ads is given by

$$EW = \begin{cases} \delta(2-\delta)m - 2A & \text{if both firms advertise} \\ \delta m - A & \text{if only one firm advertise} \\ 0 & \text{both do not advertise.} \end{cases} \quad (11.20)$$

Observe that the price p does not appear in (11.20) since it is merely a transfer from the consumer to the firms. Hence, we can infer that as long as $p < m$, a market failure is likely to occur because firms do not capture the entire consumer surplus and therefore will underadvertise compared with what a social planner would choose. Therefore, in order to check whether too many firms engage in advertising from a social viewpoint, we set $p = m$, implying that all consumer surplus is absorbed in the firms' profits. In this case, (11.20) implies that it is socially optimal to have two firms sending ads (rather than a single firm) if and only if $m/A = p/A > 1/[\delta(1-\delta]$. However, Proposition 11.3 implies that a weaker parameter restriction is needed for having an equilibrium where two firms send ads. That is, $m/A = p/A > 2/[\delta(2-\delta)]$. Hence, in Figure 11.2 the area between the curves given by $2/[\delta(2-\delta)] < p/A < 1/[\delta(1-\delta)]$ represents the parameter range where both firms advertise in equilibrium, but it is socially optimal to have only one.

Proposition 11.4 *In a model where some placed ads do not reach the consumer, there exists a parameter range where too many firms engage in advertising from a social welfare point of view. Formally, if* $p = m$ *this range is given by* $2/[\delta(2-\delta)] < p/A < 1/[\delta(1-\delta)]$.

Finally, what happens when the advertising technology improves, in the sense that there is a higher probability that ads sent to consumers actually arrive? Formally, when $\delta \to 1$, the upper curve marked by $1/[\delta(1-\delta)] \to +\infty$ which means for given values of $m = p$ and A, there always exist values of δ sufficiently close to 1 such that it is not socially optimal to have two firms engage in advertising. The intuition is as follows. Since sending ads is costly, and since $\delta \to 1$ implies that ads are always received, then it is sufficient to have only one firm sending an ad in order for the consumer to receive the information about the product.

11.3 Targeted Advertising

The literature on advertising assumes that advertising is either persuasive or informative. That is, the nature of advertising is always treated as exogenously given, thereby ignoring the question of how firms choose the content for their advertising.

The underlying observation is that societies are composed of heterogeneous consumers with different rankings (preferences) over products. Thus, firms are unable to advertise and sell their brands to all types of consumers and therefore must limit the scope of their advertising by choosing a narrow group of consumers to which their advertising appeals. There may be three reasons for that: First, it is impossible to classify products' attributes that are (highly) valued by all consumers. Second, given the high cost of advertising, firms and advertising agencies may find it profitable to narrow the scope of advertising to a limited group

of consumers. Third, ignoring advertising costs, since product differentiation may facilitate price competition, firms may intentionally choose to target a limited consumer group.

The purpose of this section is to propose a framework for modeling firms' choice of advertising methods and the resulting targeted consumer group, where firms' advertising must be confined to choosing a single advertising method and therefore a single consumer target group. For example, a firm may choose to advertise its brand by emphasizing one attribute of the product that is preferred by at least one consumer group but is not found in a competing brand. Alternatively, instead of advertising the product's attributes, a firm may target its advertising to a certain age group (young or old) or to inexperienced consumers and ignore the (attributes) quality differences among the competing brands.

11.3.1 Firms and consumers

There are two firms denoted by i, $i = 1, 2$, producing differentiated brands, which we will refer to as brand 1 and brand 2, respectively.

There are two types of buyers: There are N consumers, who are first-time buyers that we call the inexperienced consumers. In addition, there are E consumers, who have purchased the product before and whom we call *experienced* consumers. Figure 11.3 illustrates how the consumer population is divided between consumer types. We assume that the

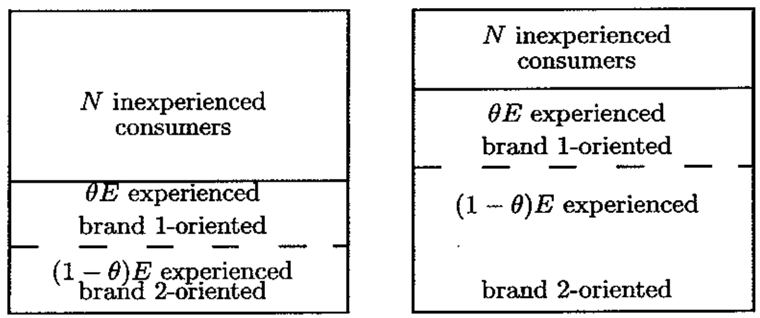

Figure 11.3: Targeted advertising: Experienced versus inexperienced consumers out of total population: *Left:* $E < N$; *Right:* $E > N$

group of experienced consumers is divided into two subgroups: those who prefer to purchase brand 1 over brand 2, and those who prefer brand 2 over brand 1. Let θ, $0 < \theta < 1$, be the fraction of brand 1-oriented consumers (among experienced consumers). Therefore, $(1-\theta)$ is the fraction of brand 2-oriented consumers (among experienced consumers). Thus, out of a total of E experienced consumers, there are θE brand 1-oriented and $(1-\theta)E$ brand 2-oriented consumers.

11.3.2 Advertising methods

There are two advertising methods: A firm can use persuasive advertising, a strategy denoted by P. Alternatively, a firm can use informative advertising, a strategy denoted by I. Thus, each firm i chooses s^i from an action set given by $S \equiv \{P, I\}$. For our purposes, we assume that no firm can employ more than one advertising method, that is, a firm can choose P or I but not both! One justification for such a strong assumption would be that advertising agencies tend to specialize in a single advertising method (or philosophy). Therefore, if a firm would like to use both advertising methods, it has to employ two advertising agencies, which may increase cost more than profit.

To simplify our model, we assume that choosing advertising methods is the only strategic variable available to firms. Thus, in this model, we ignore prices and assume that firms seek to maximize the number of consumers buying their brand. We denote by $\Pi \equiv \langle \pi^1, \pi^2 \rangle$ the vector of profit levels (which equals the number of customers buying from each firm.) We make the following assumption:

ASSUMPTION 11.1

1. *Persuasive advertising attracts only inexperienced consumers. Formally, if firm i chooses $s^i = P$, then*
   (a) *if firm j does not use persuasive advertising, then all inexperienced consumers purchase brand i, that is, $\pi^i = N$ if $s^j \neq P$;*
   (b) *if both firms use persuasive advertising, then all inexperienced consumers are equally divided between the two firms, that is, $\pi^i = N/2$ if $s^j = p$.*
2. *Informative advertising attracts only the experienced consumers who are oriented toward the advertised brand. Formally, if firm 1 chooses $s^1 = I$, then $\pi^1 = \theta E$, and if firm 2 chooses $s^2 = I$, $\pi^2 = (1-\theta)E$.*

Table 11.1 demonstrates the profit level of each firm and the industry aggregate profit under all four possible outcomes $\langle s^1, s^2 \rangle$. We look for a Nash equilibrium (see Definition 2.4) in the above strategies.

Profit \ Outcome	$\langle P,P \rangle$	$\langle P,I \rangle$	$\langle I,P \rangle$	$\langle I,I \rangle$
π^1	$N/2$	N	θE	θE
π^2	$N/2$	$(1-\theta)E$	N	$(1-\theta)E$
$\pi^1+\pi^2$	N	$N+(1-\theta)E$	$\theta E + N$	E

Table 11.1: Profits for firms under different advertising methods

Proposition 11.5

1. *A necessary condition for having both firms using persuasive advertising is that the number of inexperienced consumers exceeds the number of experienced consumers ($N > E$). In this case, $\langle P,P \rangle$ is a unique equilibrium if $1 - \frac{N}{2E} < \theta < \frac{N}{2E}$.*
2. *A necessary condition for having both firms using informative advertising is that the number of experienced consumers is more than twice the number of inexperienced consumers ($E > 2N$). In this case, $\langle I,I \rangle$ is a unique equilibrium if $\frac{N}{E} < \theta < 1 - \frac{N}{E}$.*
3. *If brand 1 is unpopular among experienced users, then firm 1 uses persuasive advertising and firm 2 uses informative advertising. Formally, $\langle P,I \rangle$ is an equilibrium if $\theta < \min\{\frac{N}{E}; 1 - \frac{N}{2E}\}$.*
4. *If brand 1 is sufficiently popular among experienced users, then firm 1 uses informative advertising and firm 2 uses persuasive advertising. Formally, $\langle I,P \rangle$ is an equilibrium if $\theta > \max\{\frac{N}{2E}; 1 - \frac{N}{E}\}$.*

Proposition 11.5 is illustrated in Figure 11.4. The upper part of Figure 11.4 corresponds to part 1 of Proposition 11.5, where the number of experienced consumers is lower than the number of inexperienced consumers. Both firms use persuasive advertising when the brands have similar popularity among experienced users. As the number of experienced consumers gets below $E < N/2$, the entire θ range corresponds to $\langle P,P \rangle$ where both firms use persuasive advertising. That is, for every popularity parameter θ, the unique equilibrium is $\langle P,P \rangle$.

The lower part of Figure 11.4 corresponds to part 2 of Proposition 11.5, where the number of experienced consumers is more than twice the number of inexperienced consumers. In this case both firms use informative advertising unless one brand is very popular among the experienced consumers compared with the other brand. Then, a firm would use persuasive advertising only if its brand is very unpopular among the experienced consumers. Finally, as the number of experienced

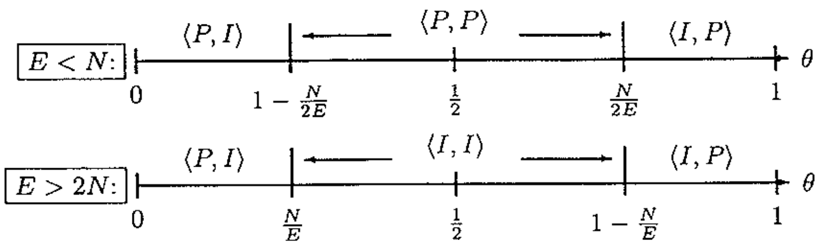

Figure 11.4: Informative versus persuasive advertising

consumers increases with no bounds, the entire popularity parameter θ range corresponds to having both firms using only informative advertising.

Proof of Proposition 11.5.
Part 1: We look at firm 1. In this equilibrium, $\pi^1(P, P) = N/2$. If firm 1 deviates and chooses $s^1 = I$, then $\pi^1(I, P) = \theta E$. Therefore, a deviation is not profitable for firm 1 if $N/2 > \theta E$, or if $\theta < N/(2E)$. Similarly, firm 2 will not deviate if $1-\theta < N/(2E)$, or $\theta > 1-N/(2E)$. In order for this region to be nonempty, we must have it that $1-N/(2E) < N/(2E)$, implying that $E < N$.

Part 2: In this equilibrium, $\pi^1(I, I) = \theta E$. If firm 1 deviates and chooses $s^1 = P$, then $\pi^1(P, I) = N$. Hence, firm 1 will not deviate if $\theta E > N$, or if $\theta > N/E$. Similarly, firm 2 will not deviate if $1 - \theta > N/E$, or $\theta < 1 - N/E$. In order for this region to be nonempty, we must have it that $N/E < 1 - N/E$, implying that $E > 2N$.

Part 3: For firm 1, $\pi^1(P, I) = N$. If firm 1 deviates to $s^1 = I$, then $\pi^1(I, I) = \theta E$. Hence, firm 1 will not deviate if $N > \theta E$ or if $\theta < N/E$. For firm 2, $\pi^2(P, I) = (1 - \theta)E$. If firm 2 deviates to $s^2 = P$, then $\pi^2(P, P) = N/2$. Hence, firm 2 will not deviate if $(1 - \theta)E > N/2$, or $\theta < 1 - N/(2E)$. Altogether, $\theta < \min\{N/E; 1 - N/(2E)\}$.

Finally, Part 4 can be proved in a similar way, and we leave it as an exercise to the reader. ■

11.4 Comparison Advertising

Comparison advertising is defined as one in which the advertised brand and its characteristics are compared with those of the competing brands.

11.4.1 Comparison advertising: an overview

In the United States, no law ever prevented the use of comparison advertising. However, advertisers were reluctant to use it (Boddewyn and Marton 1978). Only in the early 1970s, did television networks begin to (extensively) broadcast comparison advertisements. Since then, comparison ads have become popular in the printed media as well as in the broadcast media.

The EEC also began to address the issue of comparison advertising in the late 1970s, suggesting that comparison advertising should be legal as long as it compares material and verifiable details and is neither misleading nor unfair.

The main advantage of comparison advertising is that the information contained in a comparison advertisement provides consumers with low-cost means of evaluating available products (Barnes and Blakeney 1982). In addition, comparison advertising makes the consumers more conscious of their responsibility to compare before buying. It also forces the manufacturer to build into the products attributes consumers want and eventually to produce a better product. There are arguments suggesting that comparison advertising does not assist consumer comparisons because the comparison will lack objectivity since the advertiser will select only those aspects of his brand that are superior to those of the competitors. The critics consider that the risk of consumer confusion and deception is great in comparison advertising, partly because of information overload.

In most countries where comparative advertising is legal, it is closely monitored and regulated by government agencies. Different studies suggest different figures on the relative use of comparative advertising. Muehling, Stoltman, and Grossbart note that around 40 percent of all advertising is comparative. Others (Pechmann and Stewart 1990, and references) suggest that the majority of all ads are indirectly comparative (60 percent, as opposed to 20 percent that contain direct comparative claims; the rest are noncomparative).

11.4.2 Strategic use of comparison advertising

The model developed in section 11.3 can be modified to capture the effects of comparative advertising. Assume that each firm has an action set given by $S \equiv \{A, C\}$, where C means that a firm uses comparison advertising, and A means that the firm advertises its product without comparing it to the competing brand.

Following Assumption 11.1, we assume that

ASSUMPTION 11.2

1. *Plain (noncomparative) advertising (A) attracts only the inexperienced consumers.*

2. *Comparison advertising (C) attracts only the experienced consumers who are oriented toward the advertised brand.*

Thus, plain (noncomparative) advertising is intended to inform consumers about the existence of the product by informing the consumer about a specific brand. The drawback of plain advertising is that it also attracts new consumers of the wrong type.

In contrast, comparison advertising informs the experienced misplaced consumers (wrong-brand users) about the difference between the brand they have purchased in the past and their ideal brand. Thus, a firm uses the comparison-advertising strategy to attract experienced users who are oriented toward its brand.

The intuition behind Assumption 11.2 is simple. It is likely that a comparison advertisement is meaningless for the inexperienced consumer simply because a nonuser may not understand the way the product and its features operate. Thus, an inexperienced consumer will not comprehend an ad involving a comparison of the brands' attributes. Assumption 11.2 suggests that the relevance of comparison advertising is a consequence of prior experience with the product itself. Assumption 11.2 also suggests that plain advertising is not very relevant (irrelevant in our extreme case) to the experienced user, since an experienced user definitely knows about the existence of the product and its basic features. Although Assumption 11.2 sounds very intuitive, it has not been tested. In fact, many experiments cited in the references (e.g., chapter 7 of Boddewyn and Marton 1978) tend to find very little difference in the effects produced by comparative and by noncomparative advertising. However, none of these tests attempted to test them on experienced and first-time buyers separately.

Applying Proposition 11.5 to the present case yields

Proposition 11.6

1. *Comparison advertising is used by both firms when the majority of the potential consumers are experienced. That is, when $E > 2N$.*

2. *Comparison advertising will not be used if the number of inexperienced consumers is larger than the number of inexperienced consumers. That is, when $E < N$.*

3. *Comparison advertising is used by the popular firm producing the more popular brand among the experienced consumers. That is, a firm would use comparison advertising when the fraction of experienced consumers oriented toward its brand is large.*

11.5 Other Issues Concerning Advertising

11.5.1 Advertising and quality

Information about prices of products is often easier to acquire than information about the quality of products. It is relatively easy (although costly) to find out the distribution of prices for TV sets. However, it is difficult to find out the frequency of repair of various TV brands, for the simple reason that producers do not release these data to consumers.

Several authors questioned whether information on quality can be transmitted via advertising. That is, can advertising correctly inform consumers on the quality of the product? If the answer is yes, then one should ask what the exact relationship is between advertising and the quality of the advertised product.

Advertising a search good (if it occurs) is likely to be honest because lies will be detected immediately. Thus, false advertising of search goods may hurt firms' reputations rather than enhance them. This need not be the case for experience goods, for which producers may gain from false advertising (at least in the short run). Producers of experience goods will attempt to develop all kinds of persuasive methods to get consumers to try their products.

There are few analytical models attempting to find the link between advertising and quality. Schmalensee (1978) finds that low-quality brands are more frequently purchased, and that firms producing low-quality products advertise more intensively. Thus, there is a negative correlation between the intensity of advertising and the quality of the advertised product.

Kihlstrom and Riordan (1984) develop a two-period model in which high- and low-quality products are sold and high quality firms have an incentive to advertise in order to "trap" the consumers seeking to purchase high-quality products in the second period, (i.e., trap repeat buyers). Their model finds a positive correlation between advertising intensity and the quality of the advertised product. On this line, which is similar to the signaling model of subsection 8.4.6, Milgrom and Roberts (1986) develop a signaling model in which a high level of advertising is used as a signal sent by high-quality-producing firms to those consumers who desire to purchase high-quality products. Bagwell (1994) and Bagwell and Ramey (1994) argue that efficient firms operating under increasing returns tend to spend large amount on advertising to convince buyers that large sales will end up with lower prices (due to lower cost). Thus, efficient firms would spend more on advertising than less efficient firms to reveal their cost identity to the buyers.

11.5.2 Advertising and concentration

Basic intuition may lead us to think that in a (near) competitive industry with a large number of firms, no firm would have an incentive to advertise, since (persuasive) advertising may boost the demand facing the industry, but may have only a small effect on the demand facing the advertising firm. Thus, a "free rider" effect will generate little advertising. Recognizing this effect leads advertising associations in some countries to advertise how good advertising can be.

This kind of argument generates the testable hypothesis that intensive advertising (high advertising-expenditure-to-sales ratio) is associated with the more concentrated industries, (concentration measures are analyzed in section 8.1). Orenstein (1976) summarized early empirical tests that attempted to investigate a connection between advertising and concentration. From a theoretical point of view, this hypothesis can be explained by an increasing-returns type of argument. Kaldor claimed that if one takes an industry in which advertising is prohibited, and then allows advertising, the larger firms would increase their advertising expenditure at a faster rate than the smaller firms, thereby increasing industry concentration. However, Telser (1964) demonstrated very little empirical support for an inverse relationship between advertising and competition. In addition, Orenstein (1976) tested for increasing returns in advertising (say, resulting from a falling advertising cost associated with an increase in advertising volume) but showed very little evidence in favor of this hypothesis. For a very comprehensive recent empirical and theoretical study of the association between industry structure, concentration, and advertising intensity, the reader is referred to Sutton 1991.

Several authors, including Sutton (1974), suggested that the relationship between advertising and concentration need not be always monotonically increasing and that there can exist a certain concentration level at which advertising is most intensive. That is, the relation between advertising and concentration may take the form of an (upside-down) U-shaped function. Sutton suggested that industries with low concentration are associated with low incentives to advertise together with low opportunity, (by "incentive" Sutton meant the extra profit generated by extra advertising; whereas, by "opportunity" he meant the success of the advertising). However, Sutton suggested that in highly concentrated industries both the incentives and the opportunity are lower than in medium-level concentrated industries because profit expectations tend to be higher in medium-concentration industries.

11.5.3 Advertising and prices

Despite the fact that there is no significant evidence for the association between concentration and advertising intensity, there is, however, some evidence on how advertising affects prices. Benham (1972) found that the average price of eyeglasses in states where advertising eyeglasses is prohibited is around twice the average price of eyeglasses in states where eyeglasses are advertised. A similar test regarding the introduction of toy advertising on television suggests a sharp price reduction following this introduction.

How can we explain this observation that high advertising intensity is associated with lower price, but not necessarily in a reduced market concentration? We demonstrate it by the following simple example. Let us first assume that there is only one firm (monopoly) selling a particular good, whose period 0 demand is given by $Q = a_0 - p$, where a_0 is (or is positively related to) the period 0 level of advertising by the monopoly. Let A denote the advertising cost. We assume a simple form of increasing-returns technology represented by the following cost function:

$$TC(A,Q) = \begin{cases} A + c_H Q & \text{if } Q \le Q^* \\ A + c_L Q & \text{if } Q > Q^* . \end{cases} \tag{11.21}$$

Thus, for a given advertising level A, the variable cost is discontinuous at the output level Q^*. Figure 11.5 illustrates that the marginal production cost falls to c_L at output levels exceeding Q^*, reflecting a situation where at high output levels, the firm uses a different production method, say employing assembly lines to assemble products or shipping production overseas to low-wage countries. We saw in section 5.1 that the period 0 monopoly equilibrium is at a production level of $Q_0^M = (a_0 - c_H)/2$ and a price level of $p_0^M = (a_0 + c_H)/2$.

Now, suppose that in period 1, the monopoly intensifies its advertising effort and spends $A_1 > A_0$ on advertising. We assume that a higher level of advertising shifts the demand to $Q = a_1 - p$, where $a_1 > a_0$. Figure 11.5 shows that the new equilibrium is associated with an output level $Q_1^M = (a_1 - c_L)/2$ and a price of $p_1^M = (a_1 + c_L)/2$. Comparing the prices associated with the two advertising levels yields

Proposition 11.7 *Monopoly price $p_1^M < p_0^M$ if and only if $c_H - c_L > a_1 - a_0$. That is, advertising reduces the monopoly price if and only if the reduction in marginal cost associated with a higher production level exceeds the level of change in the demand.*

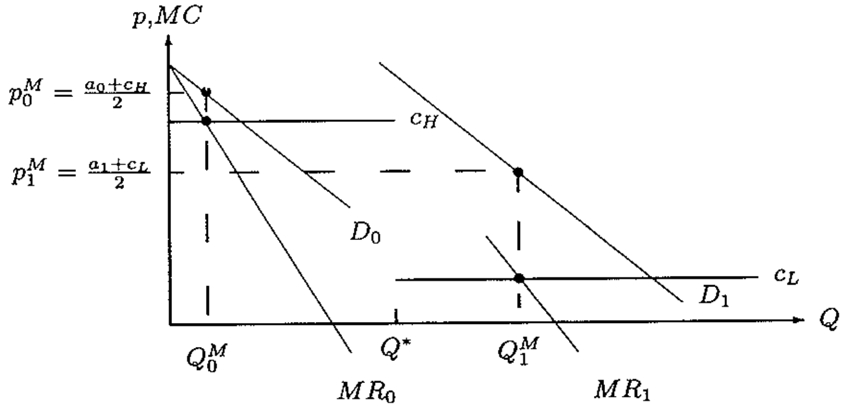

Figure 11.5: Advertising-induced demand increase and falling prices

We have ignored the question of whether advertising is profitable for this monopoly, since it simply depends on how the period 1 advertising expenditure relates to the period 0 advertising level (i.e., on the magnitude of $A_1 - A_0$). If this difference is relatively low, then the monopoly will advertise and price will fall if the condition in Proposition 11.7 is fulfilled. If the difference in advertising expenditure is large, then the monopoly may choose not to increase its advertising level. In any case, we have shown that it is possible to have a situation where prices fall (or rise) when advertising increases but the industry concentration level remains unchanged (in this case concentration remains at the level of 100 percent).

The conclusion from this experiment demonstrates a very well known econometric problem in which looking at data on prices and quantities cannot reveal what has happened to concentration, since prices and quantities may be affected by demand and production cost changes at the same time.

11.6 Appendix: Advertising Regulations

Advertising regulation has two purposes:

1. Regulation prevents firms from using advertising in a way that limits the competition among the firms in the industry.
2. Regulation is intended to protect consumers from false advertising and misrepresentations. In addition, some (negative) advertising, such as the labels on clothes, or smoking-alert labels on cigarettes, is sometimes mandated by governments.

The main difficulty in establishing advertising regulations stems from the fact that these two goals may in some cases conflict with one another. That is, in order to protect the consumer against misrepresentations, the FTC or the local government have to limit the scope of advertising. However, restricting advertising may hamper the operation of the competitive process. A second difficulty in regulating advertising stems from the fact that many countries allow free speech (including commercial free speech), implying that producers are free to advertise their products and services. Yet producers of product or services tend to misrepresent their products and services, thereby leading some consumers to believe that they buy what they want, although they actually do not.

In the following subsections we discuss some advertising regulations in two large markets: The United States and the EC. The interested reader is referred to Barnes and Blakeney 1982, McManis 1988, and Maxeiner and Schotthöfer 1992 for extensive discussions and analysis of country-specific advertising regulations.

11.6.1 The United States

We focus most of our discussion on the United States since advertising is used most intensively in the United States, and (paradoxically) advertising is heavily regulated in the United States. In the United States, federal, state, and local governments independently regulate advertising. Concurrent regulation is not contradictory, since state laws should not conflict with federal laws. In practice, advertising laws differ from state to state.

Federal advertising legislation is found in two major laws: the Federal Trade Commission Act and the Trademark (Lanham) Act. In practice, the FTC issues advertising guidelines to the industry. States create their own versions of the FTC Act. Finally, the private sector is also active as a self-regulator by imposing many rules via organizations such as the Consumers' Union and Better Business Bureaus.

Under the First Amendment to the U.S. Constitution, freedom of speech is protected. However, freedom of speech applies only to truthful advertising; that is, false advertising is not protected.

The Trademark (Lanham) Act prohibits the use of false designations of origin and false or misleading descriptions of fact and the representation of a fact. This includes the prohibition of the creation of confusion about the origin, sponsorship, and approval of goods and services. The FTC Act prohibits any unfair methods of competition, including dissemination of false advertisements. There is the question of which ads constitute false or misleading advertising. First, misleading advertising has to be material (i.e., it should affect the consumers' decisions). Second, the claims (or implied claims) made in the ad have to be false, where omissions do not constitute false advertising. Third, the ads have to mislead a substantial fraction of the audience, where the audience is expected to have a "reasonable" interpretation.

The FTC requires that the advertisers (advertising agencies) will have bases for their advertised objective claims. Subjective claims such as "this product has changed my life (for the good or for the bad)" need not be substantiated. This guide is particularly important for the case of comparison advertising, which is perfectly legal (even somewhat encouraged) in the United States, but all claims must be substantiated.

Finally, in the United States there are special federal laws that address special products and services. For example, advertising cigarettes on TV is prohibited. Special regulations prevail for advertising financial investments and drugs.

11.6.2 The European Community

Advertising in Europe is generally regulated by national laws. Regulation by the EC takes the form of directives to governments, meaning that the member countries would have to adopt their own laws in order to achieve the (directed) results.

The EC Treaty guarantees the freedom of movement of goods and services across member states. This implies the freedom of transnational advertising. Thus, the idea is to promote a market favorable to all member states' products. The EC directive toward TV and radio advertising is intended to limit the ads separable from the programs to a maximum of 20 percent of the broadcasting time. The ads should not be discriminative on the basis of nationality or any other basis. Cigarette advertising is prohibited, and advertising alcoholic beverages on TV is restricted. In addition, comparison advertising is legal as long as it is based on substantiated grounds. (Australia also allows comparison advertising based on testable claims [see Barnes and Blakeney 1982]).

Finally, the EC has also issued some directives concerning misleading advertising, thereby encouraging member states to adopt measures in order to prevent it.

11.7 Exercises

1. Congratulations! You have been appointed to become a CEO of UGLY, Inc., the sole producer of facial oil skin-life extender. Your first as-

signment is to determine the advertising budget for next year. The marketing department provides you with three important information items: (a) The company is expected to sell \$10 million worth of the product. (b) It is estimated that a 1 percent increase in the advertising budget would increase quantity sold by 0.05 percent. (c) It is also estimated that a 1 percent increase in the product's price would reduce quantity sold by 0.2 percent.

   (a) How much money would you allocate for advertising next year?
   (b) Now, suppose that the marketing department has revised its estimation regarding the demand price elasticity to 1 percent increase in price, resulting in a reduction in quantity sold by 0.5 percent. How much money would you allocate to advertising after getting the revised estimate?
   (c) Conclude how a change in the demand price elasticity affects advertising expenditure.

2. In Future City there are two fortune-tellers: Ms. α and Mr. β. Each fortune-teller charges a fixed (regulated) fee of \$10 for one visit. Let A_i denote the advertising expenditure of fortune-teller i, $i = \alpha, \beta$. The number of clients visiting each teller (per unit of time) is denoted by n_i, $i = \alpha, \beta$. We assume that n_i depends only on the advertising expenditure of both tellers. Formally, let

$$n_\alpha \equiv 6 - 3\left(\frac{A_\beta}{A_\alpha}\right) \quad \text{and} \quad n_\beta \equiv 6 - 3\left(\frac{A_\alpha}{A_\beta}\right).$$

   Thus, the number of clients visiting teller α increases with α's advertising expenditure and decreases with β's advertising expenditure. Altogether, assume that each fortune-teller i has only one choice variable, which is the advertising level, and therefore chooses A_i to maximize the profit given by

$$\pi_i(A_\alpha, A_\beta) = 10n_i - A_i = 10\left[6 - 3\frac{A_j}{A_i}\right] - A_i, \quad i = \alpha, \beta.$$

   (a) Compare the number of visitors and the profit level of each fortune-teller when $A_\alpha = A_\beta = \$1$ and for $A_\alpha = A_\beta = \$2$. What can you conclude about the role of advertising in this city?
   (b) Calculate and draw the best-response function of teller β as a function of the advertising expenditure of teller α. (In case you forgot how to define best-response functions, we first used them in section 6.1).
   (c) Calculate the tellers' advertising level in a Nash equilibrium.
   (d) In view of your answer to (a), is the Nash equilibrium you found in (c) optimal for the fortune-teller industry?
   (e) Is the equilibrium you found stable?

3. Prove part 4 of Proposition 11.5. *Hint:* Follow the same steps as in the proof of part 3.

11.8 References

Adams, W., and J. Brock. 1990. "The Automobile Industry." In *Structure of American Industry,* edited by W. Adams. New York: Macmillan Publishing Company.

Bagwell, K. 1994. "Advertising and Coordination." *Review of Economic Studies* 61: 153–172.

Bagwell, K., and G. Ramey. 1994. "Coordination Economics, Advertising, and Search Behavior in Retail Markets." *American Economic Review* 84: 498–517.

Barnes, S., and M. Blakeney. 1982. *Advertising Regulation.* Sydney: The Law Book Company.

Benham, L. 1972. "The Effects of Advertising on the Price of Eye-Glasses." *Journal of Law and Economics* 15: 337–352.

Boddewyn, J. J., and K. Marton. 1978. *Comparison Advertising.* New York: Hastings House Publishers.

Butters, G. 1977. "Equilibrium Distributions of Sales and Advertising Prices." *Review of Economic Studies* 44: 465–491.

Demsetz, H. 1979. "Accounting for Advertising as a Barrier to Entry." *Journal of Business* 52: 345–360.

Dixit, A., and V. Norman. 1978. "Advertising and Welfare." *The Bell Journal of Economics* 9: 1–17.

Dorfman, R., and P. Steiner. 1954. "Optimal Advertising and Optimal Quality." *American Economic Review* 44: 826–836.

Grossman, G., and C. Shapiro. 1984. "Informative Advertising With Differentiated Products." *Review of Economic Studies* 51: 63–81.

Kaldor, N. 1950. "The Economic Aspects of Advertising." *Review of Economic Studies* 18: 1–27.

Kihlstrom, R., and M. Riordan. 1984. "Advertising as a Signal." *Journal of Political Economy* 92: 427–450.

Maxeiner, J., and P. Schotthöfer. 1992. *Advertising Law in Europe and North America.* Deventer, The Netherlands: Kluwer Law and Taxation Publishers.

McManis, C. 1988. *Unfair Trade Practices in a Nutshell.* St. Paul, Minn.: West Publishing Co.

Meurer, M., and D. Stahl. 1994. "Informative Advertising and Product Match." *International Journal of Industrial Organization* 12: 1-9.

Milgrom P., and J. Roberts. 1986. "Price and Advertising Signals of Product Quality." *Journal of Political Economy* 94: 796–821.

Muehling, D., J. Stoltman, and S. Grossbart. 1990. "The Impact of Comparative Advertising on Levels of Message Involvement." *Journal of Advertising* 19: 41–50.

Nelson, P. 1970. "Information and Consumer Behavior." *Journal of Political Economy* 78: 311–329.

Nelson, P. 1974. "Advertising as Information." *Journal of Political Economy* 82: 729–754.

Orenstein, S. 1976. "The Advertising - Concentration Controversy." *Southern Economic Journal* 43: 892–902.

Pechmann C., and D. Stewart. 1990. "The Effects of Comparative Advertising on Attention, Memory, and Purchase Intentions." *Journal of Consumer Research* 17: 180–191.

Schmalensee, R. 1972. *The Economics of Advertising.* Amsterdam: North-Holland.

Schmalensee, R. 1978. "A Model of Advertising and Product Quality." *Journal of Political Economy* 86: 485–503.

Schmalensee, R. 1986. "Advertising and Market Structure." In *New Developments in the Analysis of Market Structure,* edited by J. Stiglitz and G. Frank Matthewson. Cambridge, Mass.: MIT Press.

Shapiro, C. 1980. "Advertising and Welfare: Comment." *Bell Journal of Economics.* 11: 749–752.

Sutton, J. 1974. "Advertising Concentration, Competition." *Economic Journal* 56–69.

Sutton, J.. 1991. *Sunk Costs and Market Structure.* Cambridge, Mass.: MIT Press.

Telser, L. 1964. "Advertising and Competition." *Journal of Political Economy* 72: 537–562

Chapter 12

Quality, Durability, and Warranties

> Anybody can cut prices, but it takes brains to make a better article.
> —Philip D. Armour (1832-1901)

We observe that products within the same category are distinguished by a wide variety of characteristics. Cars, for example, are differentiated by engine size, horse power, gas consumption, body size, number of doors, body shape (sedan vs. hatchback), transmission (manual vs. automatic), and luxurious components such as air conditioning, radio, seat covers, electric windows, electric seats.

We tackled the issue of product differentiation in chapter 7, where we analyzed markets with firms' target brands for different consumer populations and showed that product differentiation facilitates price competition. In this chapter, we wish to focus on one aspect of product differentiation that we call *quality*. The only aspect of quality not explicitly analyzed is the risk (health hazard) involved in using the product (see Oi 1973).

We also confine part of the analysis in this chapter to one particular aspect of quality that we call *durability*. The reason for focusing on durability separately from quality is that durability is related to the time dimension, which has a direct impact on the frequency of repeated purchase by consumers. For this reason some economists have argued that market structure has a strong effect on the durability aspect of the product but not necessarily on other quality aspects of the product.

In general, it is hard to point out what constitutes the quality of a certain product since quality has many dimensions. Using the example of the car, we note that quality could mean acceleration, frequency of maintenance, frequency of repair, comfort, and safety. Any reader of the consumer magazines will notice that consumer magazines rarely recommend one brand over all others for the simple reason that quality has many dimensions. That is, recommendations for choosing a certain brand are generally given conditionally on the specific needs of the user. In most cases, consumer magazines provide the readers with tables for comparing from ten to thirty features among the popular brands. Hence, in general, brands are noncomparable on the basis of quality since each brand can be highly ranked because it has some features that are not available with other brands.

For this reason, since multidimensional modeling of quality is very difficult, we will follow the literature and assume that quality can be measured by a real number. Thus, we assume that a higher-quality product is indexed by a higher real number. Using this simplified measure of quality, we analyze in section 12.1 (Personal Income and Quality Purchase) the relationship between consumer-income distribution and the quality of products they purchase. Section 12.2 (Quality as Vertical Product Differentiation) explains why firms produce brands with different qualities. Section 12.3 (Market Structure, Quality, and Durability) discusses a thirty-year-old, still ongoing debate, about whether monopoly firms produce a less durable product than firms under competition. Section 12.4 (The Innovation-Durability Tradeoff) analyzes the effect of product durability on the frequency of introduction of new, improved products. Section 12.5 (The Market for Lemons) analyzes the market for used cars and demonstrates how the existence of bad cars can drive good cars from the used-car market. Section 12.6 (Quality-Signaling Games) demonstrates how high-quality firms can set their price structure in order to signal the quality of their products. Section 12.7 (Warranties) analyzes the role that warranties can play when the quality of the product is unknown prior to the actual purchase. In the appendix, section 12.8 provides a short summary of products-liability laws.

12.1 Personal Income and Quality Purchase

We provide now a short illustration of how the level of personal income affects the quality of brands purchased by different-income consumers. In a series of models, Gabszewicz and Thisse (1979, 1980) and Shaked and Sutton (1982) use the following model to determine what the levels of qualities are and the number of different quality brands that are produced in an industry with free entry and exit. For the sake of brevity, we skip the analysis of the firms and concentrate only on consumers.

Consider an industry with two firms producing brands with different qualities: quality level $k = H$ and quality level $k = L$, $(H > L > 0)$. There are two consumers denoted by i, $i = 1, 2$. The income of consumer 1 is given by I_1, and the income of consumer 2 by I_2, where $I_1 > I_2 > 0$. Thus, consumer 1 is the high-income consumer and consumer 2 is the low-income consumer. Each consumer buys only one unit of the product. The utility level of consumer i, $i = 1, 2$ is given by

$$U_i \equiv \begin{cases} H(I_i - p_H) & \text{if he buys the high-quality brand} \\ L(I_i - p_L) & \text{if he buys the low-quality brand.} \end{cases} \tag{12.1}$$

This utility function has the property that for given prices, the marginal utility of quality rises with an increase in the consumer's income.

The following proposition demonstrates how different-income consumers are assigned to different quality products under the utility function given in (12.1).

Proposition 12.1

1. *If the low-income consumer buys the high-quality brand, then the high-income consumer definitely buys the high-quality brand.*
2. *If the high-income consumer buys the low-quality brand, then the low-income consumer definitely buys the low-quality brand.*

Proof. To prove part 1, let $U_i(k)$ denote the utility level of consumer i when he buys the brand with quality k. We want to show that

$$U_1(H) = H(I_1 - p_H) > L(I_1 - p_L) = U_1(L).$$

From (12.1) we have it that since consumer 2 buys the high-quality brand then it must be that

$$U_2(H) = H(I_2 - p_H) > L(I_2 - p_L) = U_2(L).$$

Hence,

$$(H - L)I_2 > Hp_H - Lp_L.$$

Since $I_1 > I_2$, we have it that

$$(H - L)I_1 > (H - L)I_2 > Hp_H - Lp_L.$$

Therefore,

$$H(I_1 - p_H) > L(I_1 - p_L).$$

This concludes the proof for the first part. The second part is left as an exercise in section 12.9. ■

There have been several applications for the model presented above. Gabszewicz and Thisse (1979, 1980) and Shaked and Sutton (1982) present models based on the utility function (12.1) with more than two possible quality levels and show that even under free sequential entry, only a small number of different-quality brands will be produced.

12.2 Quality as Vertical Product Differentiation

In subsection 7.3.1 we introduced the Hotelling location (address) approach to product differentiation. We interpreted the location of each consumer as his preference for, say, a certain degree of sweetness desired in a chocolate bar, where distance between a consumer and the firm is proportional to the consumer's disutility from the specific brand it sells. Another interpretation for the Hotelling model is simply the physical location of two stores, where consumers must bear per-unit-of-distance transportation cost. In this section we modify the Hotelling model to capture quality differences among differentiated brands.

12.2.1 Vertical differentiation in the basic Hotelling model

The Hotelling model developed in subsection 7.3.1 was classified as a model of horizontal differentiation for the simple reason that, given that the firms are located in the same street as the consumers, there always exist consumers who would rank the two brands differently. That is, in the Hotelling model, assuming that all brands are equally priced, the consumer who is closer to firm A than to firm B would purchase brand A, whereas a consumer who is closer to firm B would purchase brand B. Thus, given equal prices, brands are not uniformly ranked among all consumers, and for this reason we say that the brands are horizontally differentiated.

Phlips and Thisse (1982) emphasized the distinction between horizontal and vertical product differentiation in the following way:

DEFINITION 12.1

1. *Differentiation is said to be* **horizontal** *if, when the level of the product's characteristic is augmented in the product's space, there exists a consumer whose utility rises and there exists another consumer whose utility falls.*

2. *Differentiation is said to be* **vertical** *if all consumers benefit when the level of the product's characteristic is augmented in the product space.*

Figure 12.1 illustrates a simple diagrammatic comparison between horizontal and vertical-quality differentiation (for a comprehensive discussion of horizontal and vertical differentiation see Beath and Katsoulacos 1991). In Figure 12.1 all consumers are located between points 0 and 1.

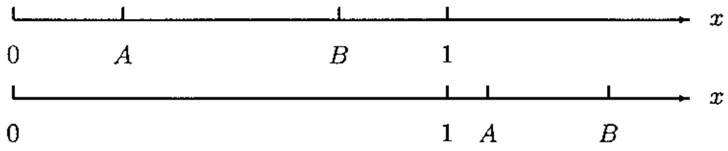

Figure 12.1: Horizontal versus vertical differentiation. *Up:* horizontal differentiation; *Down:* vertical differentiation

The upper part of Figure 12.1 is the same as the Hotelling horizontal-differentiation model displayed in Figure 7.7. In this case, given equal prices, the consumers located near firm A prefer brand A over brand B, whereas consumers located near brand B prefer brand B over brand A. In contrast, the lower part of Figure 12.1 illustrates an industry with vertically differentiated brands where all consumers prefer brand A over brand B (since all consumers are located closer to A than to B).

12.2.2 A modified Hotelling vertical-differentiation model

The basic Hotelling model developed in subsection 7.3.1 is based on preferences given in (7.17) and refers to the "street" illustrated in Figure 12.1. In what follows, we modify the utility function (7.17) so that instead of having consumers gain a higher utility from the nearby brand, all consumers would have their ideal brand located at point 1 on the $[0,1]$ interval. This modification would allow us to model product differentiation where firms still locate on the $[0,1]$ interval (and not outside this interval).

There is a continuum of consumers uniformly distributed on the interval $[0,1]$. There are two firms, denoted by A and B and located at points a and b ($0 \le a \le b \le 1$) from the origin, respectively. Figure 12.2 illustrates the location of the firms on the $[0,1]$ interval.

The utility of a consumer located at point x, $x \in [0,1]$ and buying brand i, $i = A, B$ is defined by

$$U_x(i) \equiv \begin{cases} ax - p_A & i = A \\ bx - p_B & i = B \end{cases} \qquad (12.2)$$

where p_A and p_B are the price charged by firm A and B, respectively.

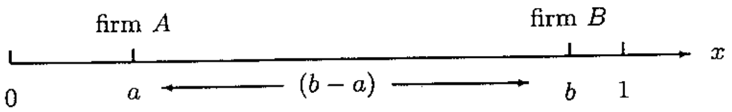

Figure 12.2: Vertical differentiation in a modified Hotelling model

We seek to define a two-period game, where firms choose location in the first period, and choose price in the second period, after locations have been fixed. Before defining the game, let us solve for a Nash-Bertrand equilibrium in prices, assuming fixed locations as illustrated in Figure 12.2.

Let $\hat{x}$ denote a consumer who is indifferent to whether he or she buys from firm A or firm B. Assuming that such a consumer exists, and that consumer $\hat{x}$ locates between the two firms, that is $a \le \hat{x} \le b$, the location of the indifferent consumer is determined by

$$U_{\hat{x}}(A) = a\hat{x} - p_A = b\hat{x} - p_B = U_{\hat{x}}(B). \tag{12.3}$$

Thus, the utility of consumer indexed by $\hat{x}$ from buying brand A equals his utility from buying brand B. Therefore, assuming that $a \le \hat{x} \le b$, the number of consumers buying from firm A is $\hat{x}$, whereas the number of consumers buying from firm B is $(1 - \hat{x})$. Solving for $\hat{x}$ from (12.3) yields

$$\hat{x} = \frac{p_B - p_A}{b - a} \quad \text{and} \quad 1 - \hat{x} = 1 - \frac{p_B - p_A}{b - a}. \tag{12.4}$$

Figure 12.3 provides a graphic illustration of how $\hat{x}$ is determined. The left-hand side of Figure 12.3 illustrates the utility for a consumer located at any point $0 \le x \le 1$ when he or she buys brand A and when he or she buys brand B, assuming that $p_B > p_A$. By definition, for the consumer located at $\hat{x}$, the utility from buying A equals the utility from buying B. Moreover, Figure 12.3 shows that all consumers located on $[0, \hat{x}]$ gain a higher utility from purchasing brand A (lower quality) than from purchasing brand B. Similarly, all consumers located on $[\hat{x}, 1]$ gain a higher utility from purchasing brand B (higher-quality brand) than from purchasing brand A.

Note that, as in subsection 7.3.1, we assume here that consumers always buy one unit from firm A or from firm B. In contrast, assuming a reservation utility of zero would generate a group of consumers who do not purchase any brand. Formally, if a reservation utility of zero is assumed, all consumers indexed on $[0, z]$ (where consumer z, $z = p_A/a$ is drawn in Figure 12.3) will not purchase any brand. In this case, the

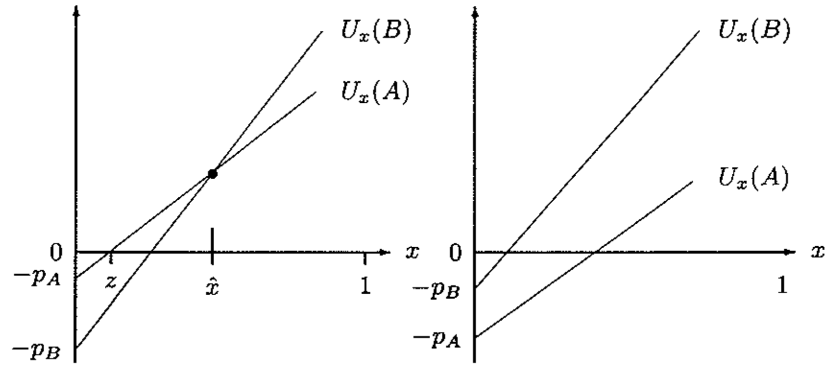

Figure 12.3: Determination of the indifferent consumer among brands vertically differentiated on the basis of quality. *Left:* $p_A < p_B$, *Right:* $p_A > p_B$

number of A buyers would be reduced to the size of the interval $[z, \hat{x}]$. Exercise 1 in Section 12.9 addresses the case of reservation utility.

It is also clear from the right-hand side of Figure 12.3 that if the price of the lower-quality brand (brand A) is higher than the price of the high-quality brand (brand B), ($p_A > p_B$), then all consumers purchase only the high-quality brand (brand B).

For given locations of firms (a and b), in the second period each firm takes the price set by its rival firm as given and chooses its price to maximize its profit level. Formally, firm A and B solve

$$\begin{aligned} \max_{p_A} \pi_A(a, b, p_A, p_B) &= p_A\hat{x} = p_A\left[\frac{p_B - p_A}{b - a}\right] \\ \max_{p_B} \pi_B(a, b, p_A, p_B) &= p_B(1 - \hat{x}) = p_B\left[1 - \frac{p_B - p_A}{b - a}\right]. \end{aligned} \tag{12.5}$$

After introducing all the assumptions for this model, we now pause to give a precise definition for this two-period game. We simply look for a subgame perfect equilibrium as described in Definition 2.10 on page 27.

DEFINITION 12.2 *The quadruple* $< a^e, b^e, p_A^e(a, b), p_B^e(a, b) >$ *is said to be a vertically differentiated industry equilibrium if*

Second period: *For (any) given locations of firms (a and b), $p_1^e(a, b)$ and $p_2^e(a, b)$ constitute a Nash equilibrium.*

First period: *Given the second period-price functions of locations $p_A^e(a, b)$, $p_B^e(a, b)$, and $\hat{x}(p_A^e(a, b), p_B^e(a, b))$, (a^e, b^e) is a Nash equilibrium in location.*

Definition 12.2 is a subgame perfect equilibrium (see Definition 2.10 on page 27) in which, in the first period, firms choose locations taking into account how their choice of location will affect the second-period equilibrium prices and hence profit levels. It is important to note that the equilibrium actions of the firms in the second periods are functions (not scalars) of all possible given locations of firms.

We now proceed to solve the model, starting from the second period. The first-order conditions to (12.5) are given by

$$0 = \frac{\partial \pi_A}{\partial p_A} = \frac{p_B - 2p_A}{b-a} \quad \text{and} \quad 0 = \frac{\partial \pi_B}{\partial p_B} = 1 - \frac{2p_B - p_A}{b-a}. \tag{12.6}$$

Hence,

$$p^e_A(a,b) = \frac{b-a}{3} \quad \text{and} \quad p^e_B(a,b) = \frac{2(b-a)}{3}. \tag{12.7}$$

Note that both equilibrium prices exceed marginal cost despite the fact that one firm produces inferior quality.

Equation (12.7) reveals that

Proposition 12.2 *The firm producing the higher-quality brand charges a higher price even if the production cost for low-quality products is the same as the production cost of high-quality products.*

Substituting (12.7) into (12.5) yields that

$$\begin{aligned} \pi_A(a,b) &= \frac{1}{b-a}\left[\frac{2(b-a)^2}{9} - \frac{(b-a)^2}{9}\right] = \frac{b-a}{9} \\ \pi_B(a,b) &= \frac{1}{b-a}\left[\frac{2(b-a)^2}{3} - \frac{4(b-a)^2}{9} + \frac{2(b-a)^2}{9}\right] = \frac{4(b-a)}{9}. \end{aligned} \tag{12.8}$$

We now move to the first period, where firm A takes b^e as given and maximizes $\pi_A(a, b^e)$ given in (12.8), whereas firm B takes a^e as given and maximizes $\pi_B(a^e, b)$.

It is easy to see that firm A would choose to produce the lowest possible quality and locate at $a^e = 0$, whereas firm B would choose to produce the highest possible quality and locate at $b^e = 1$. This result is known as the *principle of maximum differentiation*. Formally,

Proposition 12.3 *In a vertically differentiated quality model each firm chooses maximum differentiation from its rival firm.*

Are you confused? Well, you should be confused since in the horizontal differentiation model of subsection 7.3.1 we showed that when transportation costs are linear, firms tend to move toward the center (minimum differentiation). However, in a vertical (quality) differentiation model the principle of maximum differentiation applies. The reason for this difference is that in a vertically differentiated products model firms specialize in the production of quality for a certain consumer group. Maximum differentiation implies that firms can increase their market power in their targeted consumer group.

12.3 Market Structure, Quality, and Durability

There is an extensive literature debating the relationship between the degree of a firm's monopoly power and the quality or durability it chooses to build into a product (see a survey article by Schmalensee [1979]). That is, the main question is whether a monopoly firm that is known to distort prices and quantity produced (see chapter 5) also builds a shorter durability or a lower quality into its product than does a competitive industry.

Earlier writers on this subject, Kleiman and Ophir (1966) and Levhari and Srinivasan (1969), concluded that firms with monopoly power have the incentives to produce goods of lower durability than would be produced by firms in a competitive market.

Contrary to this literature, Swan (1970a, b, 1971) has demonstrated that there is actually no implied relationship between monopoly power and durability. Swan's novel result is known in the literature as the *Swan's independence result*. This result gave rise to an extensive literature examining the robustness of the independence result. Levhari and Peles (1973) demonstrated that durability built in a product produced by a monopoly can be longer or shorter than under competition. In addition, they have shown that partial regulation of a monopoly that chooses strategies of quantity produced (or price) and durability (or quality) can reduce welfare, where partial regulation is defined as a restriction by the regulating authority on either the quantity produced or the quality but not on both.

Kihlstrom and Levhari (1977) examine the robustness of Swan's result by analyzing the effect of increasing returns-to-scale (IRS) technologies on the production of durability. Spence (1975) developed a fixed-cost (implying an IRS technology) model to measure the divergence between the socially optimal quality level and the monopoly's equilibrium quality level.

The debate on Swan's independence result will probably continue forever. However, the reader is advised to learn the arguments given by the authors participating in this long debate.

In this section we provide a simple illustration of the Swan's independence result by considering a monopoly firm selling light bulbs with variable durability. Let us consider a consumer who lives for two periods and desires light services for two periods. Assume that the consumer is willing to pay an amount of $\$V$ ($V > 0$) per each period of light services.

On the supply side, assume that light bulb-producing firms possess the technology for producing two types of light bulbs: a short-durability light bulb yielding light services for one period only, and a long-durability light bulb yielding light services for two periods. The unit cost of producing the short-durability light bulb is denoted by c^S, and the unit cost of producing a long-durability light bulb is denoted by c^L, where $0 < c^S < V$, $0 < c^L < 2V$, and $c^S < c^L$.

For simplicity we ignore discounting and analyze market equilibria under extreme market structures: monopoly and perfect competition.

Monopoly firm producing light bulbs

The monopoly firm has the option of selling short- or long-durability light bulbs and to charge a monopoly price for either type of bulbs. First, suppose that the monopoly sells short-durability light bulbs. Then, since the consumer is willing to pay $\$V$ per period of light services, the monopoly would charge $p^S = V$ per period and would sell two units (one unit each period). Hence, the profit of a monopoly selling short-durability light bulbs is given by

$$\pi^S = 2(V - c^S). \tag{12.9}$$

Now, suppose that the monopoly sells long-durability light bulbs. Since the light bulb lasts for two periods, the monopoly charges a price of $p^L = 2V$. Hence, the profit of the monopoly firm selling long-durability light bulbs is given by

$$\pi^L = 2V - c^L. \tag{12.10}$$

We would like to know under what condition the monopoly produces long- or short-durability light bulbs. Clearly, the monopoly produces short-durability bulbs if $\pi^S \geq \pi^L$. Comparing (12.9) with (12.10) yields

Proposition 12.4 *A monopoly producer of light bulbs would minimize the production cost per unit of duration of the light bulb. Formally, the monopoly would produce short-durability light bulbs if $2c^S < c^L$, and would produce long-durability bulbs if $2c^S > c^L$.*

Proposition 12.4 illustrates Swan's argument that despite the fact that there is only one seller, the monopoly's decision about which type of bulb to produce depends only on cost minimization and not on the market conditions, such as the demand structure. However, to show Swan's complete argument, we investigate which type of light bulbs are produced in a competitive industry.

Competitive light bulb industry

Under perfect competition, the price of each type of light bulb drops to its unit cost. Hence, $p^S = c^S$ and $p^L = c^L$. The consumer who desires two periods of light services would purchase a short-duration light bulb if $2(V - p^S) > 2V - p^L$, or, if, $2c^S < c^L$. Similarly, consumers purchase long-durability light bulbs if $2V - p^L > 2(V - p^S)$, or if $c^L < 2c^S$.

Hence, we can state Swan's independent result by the following proposition:

Proposition 12.5

1. *The durability of light bulbs is independent of the market structure.*
2. *The firms would choose the level of durability that minimizes the production cost per unit of time of the product's services.*

It is important to note that this analysis assumes that our consumer is only concerned with the length of time service is provided by the product and does not attach any other value for durability per se. This is rather an extreme assumption since if, for example, cost minimization yields the decision that light bulbs with durability of five minutes are produced, then this means that our consumer has to replace a light bulb every five minutes. Given that our consumer may attach value for the time it takes to buy and replace a light bulb, it is unlikely that consumers will purchase short-duration light bulbs. Similarly, if cost minimization yields the decision that only single-shave razor blades are produced, then consumers will have to buy a stock of 365 razor blades each year. In this case, it is clear that consumers would be willing to pay more than five times the amount they are willing to pay for a single shave blade for a five-shave blade.

12.4 The Innovation-Durability Tradeoff

All of us often wonder what to do with our old washing machine, black-and-white TV, typewriter, personal computer, turntable, or stereo. When technologies keep changing rapidly, consumers desire new-technology products while they still receive some benefits from the older-technology product that they still own. If all consumers have similar preferences, and hence all desire the new-technology products, old-technology products cannot be sold in a market for used products. Hence, we sometimes get the feeling that with a rapidly changing technology, goods are too durable. That is, we often say to ourselves some variation of: "My old computer does not want to break down, so I don't know what to do with it once I replace it with a newer model."

The question we investigate in this section is whether and under what conditions firms may produce products with excess durability, from a social point of view. In other words, under what conditions do firms find it profitable to produce goods that will last for a very long time so that firms entering with new technologies will not be able to introduce and sell new products owing to the large existing supply of durable old-technology products.

This problem is analyzed in Fishman, Gandal, and Shy 1993 in an infinite-horizon overlapping-generations framework. Here, we merely illustrate their argument in a two-period model, with a simplifying assumption that in each period there is only one firm.

Consumers

In period $t = 1$, there is only one consumer who seeks to purchase computer services for the two periods of his or her life, $t = 1, 2$. In period $t = 2$, one additional consumer enters the markets and seeks to purchase one period of the product's services.

Let V_t denote the per period gain from the quality of the technology imbedded into the product a consumer purchases in period t, and let p_t be the corresponding price. Altogether, the per period utility of each consumer purchasing period t technology is

$$U_t \equiv \begin{cases} V_t - p_t & \text{if purchasing the period } t \text{ technology product} \\ 0 & \text{if not purchasing.} \end{cases} \tag{12.11}$$

Firms

There are two firms. Firm 1 (operating in period 1 only) is endowed with an old technology providing a (per period) quality level of v^O to consumers. Firm 2 (a potential entrant in period 2) can produce the old-technology product (v^O); however, in addition, firm 2 is endowed with the capability of upgrading the technology to a level of v^N, $v^N > v^O$ for an innovation cost of $I > 0$.

On the production side, we assume that the production cost is independent of the technology level but depends on the durability built into the product. Durability affects production costs since long-lasting products are generally made with more expensive material (say, more metal relative to plastic cases and moving parts). We say that the product is nondurable if it lasts for one period only. That is, a nondurable product is assumed to completely disintegrate after one period of usage. We say that the product is durable if it lasts for two periods. The unit production cost of a nondurable is denoted by c^{ND}, whereas the

unit production cost of a durable is denoted by c^D, where we assume that $c^D > c^{ND}$. That is, we assume that durable goods are more costly to produce than nondurables. With no loss of generality, we also assume that the production of a nondurable product is zero ($c^{ND} = 0$).

The two-period, two-firm game is described as follows: In period 1 firm 1 sells the old-technology product and therefore has to decide which price to charge, (p_1) and whether to produce a durable (D) or a nondurable (ND) product. In this second period, firm 2 obviously chooses to produce a nondurable (since the world ends at the end of period 2) and hence has to decide whether to invest in adopting the newer (v^N) technology and the price (p_2). Figure 12.4 illustrates this two-period game.

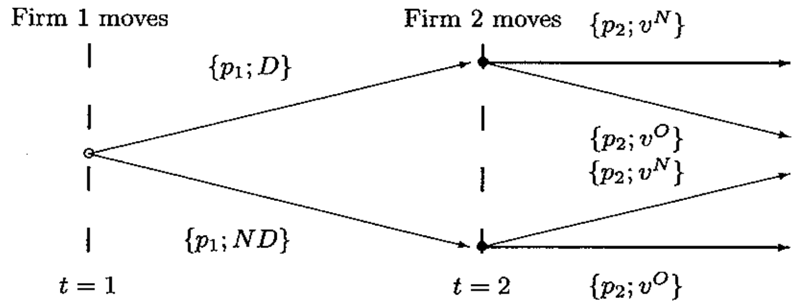

Figure 12.4: Innovation and durability

Below, we analyze two situations based on whether firm 1 produces a durable or a nondurable in period 1.

Second-period pricing, given that first-period production is nondurable

In the second period firm 2 offers either the old-technology v^O product for sale, or invests I for the adoption of its new-technology v^N product. The pricing and innovation decision of firm 2 are summarized by

$$p_2 = \begin{cases} v^N & \text{if } 2(v^N - v^O) \geq I \\ v^O & \text{if } 2(v^N - v^O) < I \end{cases} \qquad \pi_2 = \begin{cases} 2v^N - I & \text{if } 2(v^N - v^O) \geq I \\ 2v^O & \text{if } 2(v^N - v^O) < I. \end{cases} \tag{12.12}$$

That is, when firm 1 produces a nondurable in period 1, then in period 2 both the old and the new consumers seek to purchase the product. If the innovation cost is sufficiently low, firm 2 invests in the improved technology and sells it to the old and new consumers. However, if I

is high, firm 2 sells the old technology to both the old and the new consumers.

Second-period pricing, given that the first-period product is durable

Now, suppose that firm 1 sells a durable in period 1. Then, in period 2, the old consumer already possesses the v^O technology product. In this case, firm 2 has two possibilities: It can price its new-technology product low enough at $p_2^L = v^N - v^O$, which induces the old consumer to discard his old-technology durable and purchase the new product (v^N); in this case, $\pi_2^L = 2(v^N - v^O) - I$. Or, it can price it high at $p_2^H = v^N$, so that only the new consumer purchases the new-technology product, while the old keeps using the old durable product. In this case, $\pi_2^H = v^N - I$.

Comparing π_2^L with π_2^H yields:

Proposition 12.6 *Suppose that firm 1 sells a durable to period 1 consumer. Then, in period 2, firm 2 sells the new-technology product if* $I \le \max\{2(v^N - v^O); v^N\}$. *In this case,*

1. *if* $v^N > 2v^O$, *firm 2 sells its new-technology product to both the old and new consumers;*
2. *if* $v^N < 2v^O$, *firm 2 sells its new-technology product to the new consumer only.*

First-period durability choice

In period $t = 1$, firm 1 chooses a price (p_1) and whether to produce a durable or a nondurable. If firm 1 sells a nondurable, (12.11) implies that the maximum price firm 1 can charge for selling one period of the product service is $p_1^{ND} = v^O$. In this case, $\pi_1^{ND} = v^O - c^{ND} = v^O$.

In contrast, if firm 1 sells a durable, (12.11) implies that the maximum it can charge is given by $p_1^D = 2v^O$, since in this case the product provides a service of v^O for two periods. In this case, $\pi_1^D = 2v^O - c^D$. Therefore, comparing π_1^{ND} with π_1^D yields

Proposition 12.7 *Firm 1 produces a durable if* $v^O > c^D$. *Otherwise, it produces a nondurable.*

Proposition 12.7 is rather simple. Firm 1 would produce a durable if the extra profit from charging for second-period product service exceeds the difference in cost between producing a durable and a nondurable ($c^D - c^{ND} = c^D$).

Durability, innovation, and welfare

We define the social-welfare function as the sum of consumers' utility levels and the firms' profits over the two periods given by

$$W \equiv U_1^1 + U_2^1 + U_2 + \pi_1 + \pi_2 \tag{12.13}$$

where U_1^1 and U_2^1 are the utility levels of period 1 consumer in periods 1 and 2 respectively; U_2 is the utility level of the consumer who lives in period 2 only, and π_t is the profit of the firm operating in period t.

We conclude from the previous analysis that there could be three types of equilibria: (1) firm 1 produces a durable or a nondurable; (2) firm 2 innovates and adopts the new technology or does not innovate; (3) the combination of the two possibilities. The type of equilibrium that obtains is determined by the exact parameter values. In order to restrict the parameter range to interesting cases, we assume that

ASSUMPTION 12.1 $v^O > c^D$, *and* $\max\{2(v^N - v^O); v^N\} < I < 2v^N$.

The first part of Assumption 12.1 implies that the first-period firm would find it profitable to produce a durable product. The second part implies that the innovation cost for the new technology is at an intermediate range.

We now state our main proposition:

Proposition 12.8 *Under Assumption 12.1,*

1. *firm 1 produces a durable, innovation will not occur, and only the old-technology product will be sold; and*
2. *this outcome is dominated, from a social-welfare viewpoint, by an outcome where firm 1 produces a nondurable instead of a durable.*

Proof. Since $v^O > c^D$, Proposition 12.7 implies that firm 1 produces a durable in period 1. Now, by way of contradiction suppose that firm 2 innovates. Then, if firm 2 sells to both consumers, $\pi_2^L = 2(v^N - v^O) - I < 0$ by Assumption 12.1. Similarly, if firm 2 innovates and sells only to the young consumer, $\pi_2^H = v^N - I < 0$, also by Assumption 12.1: a contradiction. Hence, firm 2 will not innovate, which proves part 1 of the proposition.

To prove part 2, we first calculate the social welfare under this outcome (firm 1 produces a durable and firm 2 does not innovate). In this case, $p_1 = 2v^O$, $\pi_1 = 2v^O - c^D$, $p_2 = v^O$, $\pi_2 = v^O$, and $U_1^1 = U_1^2 = U_2 = 0$. Hence, using (12.13)

$$W^D = 3v^O - c^D. \tag{12.14}$$

Now, suppose that for some reason, firm 1 is forced to produce a non-durable. Then Assumption 12.1 implies that firm 2 does not innovate. In this case, $p_1 = v^O$, $\pi_1 = v^O$, $p_2 = v^O$, $\pi_2 = 2v^O$, and $U_1^1 = U_1^2 = U_2 = 0$. Hence, using (12.13)

$$W^{ND} = 3v^O. \tag{12.15}$$

Comparing (12.14) with (12.15) implies that $W^{ND} > W^D$. ∎

The intuition behind part 2 of Proposition 12.8 is as follows: Durability in this model serves as a strategic means to capture future market share. However, durability per se does not serve any purpose to consumers and therefore to the social planner. Since durability is costly to the economy, the social planner can increase welfare by supplying a product of the same quality with no durability.

What policy conclusions can we derive from this model? One recommendation would be for quality-regulating institutions such as standards institutes to allow short durability products into a market with rapidly changing technologies.

12.5 The Market for Lemons

So far, we have analyzed markets where sellers could control the quality of the product they sell. However, there are many markets in which products with predetermined qualities are sold, and therefore sellers are constrained to sell a product with a given quality.

If consumers can determine the precise quality by simply inspecting the product prior to the purchase (if the product is a search good), then the market will be characterized by a variety of qualities of the same product sold at different prices, where higher quality brands will be sold for a higher price. However, in most cases buyers cannot determine the quality before the actual use (the product is an experience good). A natural question to ask is whether markets can function when buyers cannot observe qualities prior to purchase and when experience goods with different qualities are sold. The reason the answer may be negative is that in such markets sellers need not adjust prices to reflect the actual quality of the specific product they sell.

In this section we analyze markets where sellers and buyers do not have the same amount of information about the product over which they transact. That is, we analyze markets with *asymmetric information,* where sellers who own or use the product prior to the sale have a substantial amount of information concerning the particular product they own. By contrast, a buyer does not possess the knowledge about the quality of the particular product he wishes to purchase.

A second feature of the particular markets we analyze here is that

reputation does not play a role. This assumption is unrealistic for certain markets where sellers generate most of their sales from returning customers. In fact, almost all the large retail stores in the United States are now allowing consumers to return the products for a full refund, thereby guaranteeing satisfactory quality. Reputation effects are also present in expensive restaurants where most sales are generated from fixed clientele. Still, there is a substantial number of markets in which reputation does not play a role. For example, our analysis will focus on the market for used cars. Whether the seller is a private owner or a dealer, the issue of reputation is of not of great interest to the seller. Therefore, if the seller possesses a low-quality product, the seller has all the incentives to sell it as a high-quality product.

The problem of asymmetric information between buyers and sellers is perhaps most noticeable in the market for used cars. A buyer has a short time to inspect the car, to check the engine's compression and oil consumption, and to perform other tests that can partially reveal the quality of the car. Since full warranties are not observed in the market for used cars, a buyer has to assume that with some probability the used car he buys may be a lemon. Of course, lemon cars need not be just old cars, since all lemon cars have been initially sold as new cars. However, the difference between new lemon cars and used lemon cars is that the seller of a new car (new-car dealer) does not know the quality of the particular car he sells to a particular customer, whereas a seller of a used car knows whether the particular car is a lemon or a good car. Thus, the markets for used and new cars have substantially different information structures.

12.5.1 A model of used and new car markets

Following Akerlof (1970), let us consider an economy with four possible types of cars: brand-new good cars, brand-new lemon cars (bad cars), used good cars, and used lemon cars. All individuals in this economy have the same preferences for all the four types of cars. We let

N^G = value of a new good car;

N^L = value of a new lemon car;

U^G = value of a used good car; and

U^L = value of a used lemon car.

We make the following assumptions:

ASSUMPTION 12.2

1. *The value of new and old lemon cars is zero; that is,* $N^L = U^L = 0$.

2. *Half of all cars (new and old) are lemons, and half are good cars.*
3. *New good cars are preferred over used good cars; that is,* $N^G > U^G > 0$.

The first and the second items of Assumption 12.2 are merely for the sake of simplifying the model. The third item is clear and is intended to induce good used-car owners to purchase new cars under certain price structure. Assumption 12.2 implies that the expected values of new and used cars are given by

$$\mathrm{E}N \equiv 0.5N^G + 0.5N^L = 0.5N^G, \quad \text{and} \quad \mathrm{E}U \equiv 0.5U^G + 0.5U^L = 0.5U^G. \tag{12.16}$$

Clearly, the expected value of a new car exceeds the expected value of a used car, $\mathrm{E}N > \mathrm{E}U$.

There are four types of agents in this economy: (1) new car dealers who sell new cars for an *exogenously given* uniform price denoted by p^N. Clearly, since there is no knowledge of the quality of new cars, all new cars are sold for the same price; (2) individuals who do not own any car, whom we call buyers in what follows; (3) owners of good used cars, whom we call sellers; and (4) owners of lemon used cars, whom we also call sellers.

We denote by p^U the price of a used car. Since used-car buyers cannot distinguish between lemon used cars and good used cars, all used cars are sold for the same price p^U.

We assume that each buyer maximizes the expected value of a car minus a price, in case the agent is a buyer. Formally, the utility of a buyer (who does not own any car) is assumed to be

$$V^b \equiv \begin{cases} \mathrm{E}N - p^N & \text{if he buys a new car} \\ \mathrm{E}U - p^U & \text{if he buys a used car.} \end{cases} \tag{12.17}$$

The utility of a seller of a good used car who sells his used car for p^U and buys a new car for p^N is given by

$$V^{s,G} \equiv \begin{cases} \mathrm{E}N - p^N + p^U & \text{if he buys a new car (and sells his used car)} \\ U^G & \text{if he maintains his (good) used car.} \end{cases} \tag{12.18}$$

Finally, the utility of a seller of a lemon car who sells his used lemon for p^U and buys a new car for p^N is given by

$$V^{s,L} \equiv \begin{cases} \mathrm{E}N - p^N + p^U & \text{if he buys a new car (and sells his used car)} \\ U^L & \text{if he maintains his lemon used car.} \end{cases} \tag{12.19}$$

That is, each used-car owner has the option to maintain his car, thereby gaining a utility of U^G or U^L, depending on whether he owns a good or a lemon used car, or to buy a new car for p^N and, in addition, get paid p^U for selling his used car.

The problem of the buyers

The buyers do not own any car and therefore have the option of either buying a new car or buying an old car. Thus, in view of (12.17) buyers will buy a used car if $\mathrm{E}U - p^U \geq \mathrm{E}N - p^N$, or if p^U satisfies

$$p^U \leq \mathrm{E}U - \mathrm{E}N + p^N = \frac{U^G - N^G}{2} + p^N. \tag{12.20}$$

The problem of the lemon used-car seller

An owner of lemon used car has the option of keeping his car (gaining zero utility), or selling his used car and buying a new car. In view of (12.19) an owner of a lemon used car sells his car if $0 \leq \mathrm{E}N - p^N + p^U$ or

$$p^U \geq p^N - \mathrm{E}N = p^N - 0.5N^G. \tag{12.21}$$

The problem of the good used-car seller

An owner of a good used car has the option of keeping his car, or selling his used car and buying a new car. In view of (12.18) an owner of a good used car sells his car if $U^G \leq \mathrm{E}N - p^N + p^U$, or

$$p^U \geq p^N + U^G - EN = p^N + U^G - 0.5N^G. \tag{12.22}$$

Figure 12.5 summarizes the cases given in (12.20), (12.21), and (12.22) in the (p^N, p^U) space, where used cars are either demanded or offered for sale. The two regions of interest are the upper one, corresponding to (12.22) where p^U is sufficiently high so that an owner of a good used car offers his or her car for sale, and the lower region, corresponding to (12.20) where buyers (those who do not own cars) find p^U to be sufficiently low and decide to purchase a used car.

Figure 12.5 shows that the combinations of p^N and p^U satisfying the condition in which good used cars are sold do not satisfy the condition in which buyers would demand used cars. That is, the region in which p^U is high enough to induce an owner of a good used car to sell his or her good car does not intersect with the region in which p^U is low enough to induce a buyer to purchase a used car instead of a new car. This proves our main proposition, known as the *Lemons' Theorem*.

Proposition 12.9 *Good used cars are never sold. That is, lemon used cars drive good used cars out of the market.*

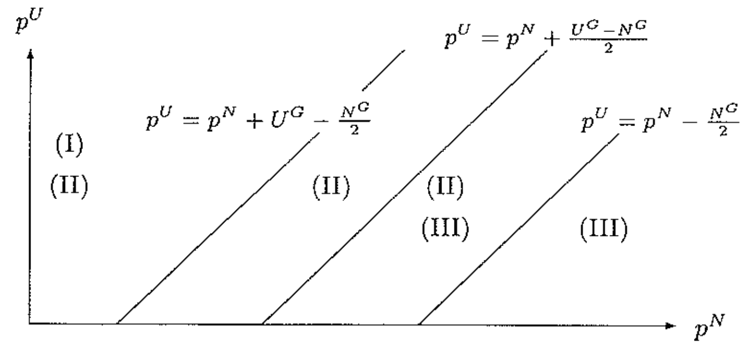

Figure 12.5: The market for lemons: Bad cars drive out the good cars. The prices of new and used cars corresponding to cases where used cars are demanded or offered for sale. (I) Good used-car seller sells. (II) Bad used-car seller sells. (III) Buyers demand used cars.

A reader who may have purchased a good used car may wonder how it happened. That is, we sometimes observe that good used cars are sold in the market despite what Proposition 12.9 predicts. The reason this happens follows from our assumption that used-car owners may sell their cars only if they wish to buy a new one. However, it often happens that used-cars owners sell their cars for different reasons, such as moving to another state or abroad. Thus, the observation that sometimes good used cars are sold does not contradict Proposition 12.9.

12.5.2 Applications of the lemon problem

The model described in the previous subsection can be applied to describe a wide variety of other markets as well. Consider the health insurance market, where both healthy and sick people wish to purchase health insurance from an insurance company or an HMO. The buyer of an insurance policy knows whether he is healthy. However, the insurance company has no prior information on the particular buyer, unless it requires that all buyers go through an extensive medical checkup. If the insurance price reflects the average treatment costs for a certain period, then by the same argument as in the previous subsection, it is clear that only sick people would purchase health insurance. So, the remaining question is how can insurance companies or HMOs make a profit? The answer is probably that insurance companies attempt to discriminate on

the basis of price (charge different rates) according to age and according to health problems the patient had prior to filing the insurance applications. A similar problem occurs in other insurance coverage, namely, risky drivers tend to buy extended coverage for their car. Insurance companies can partially solve this problem by charging different rates according to age, location, and distance to work, following the data they collect on accident frequencies.

Consider now the market for all-you-can-eat restaurants. The buyers are divided into two groups of people, those who eat a lot, and those who eat very little. If the price of a meal reflects the food cost of the average eaters, then it is clear that only very hungry people would go to all-you-can-eat restaurants, whereas less hungry people would generally prefer to pay for each specific dish they order. The question is, then, how can all-you-can-eat restaurants earn a profit? The answer is perhaps that many all-you-can-eat restaurants also serve regular meals, and, as with most restaurants, they earn the profit on side dishes, such as drinks and desserts.

Consider now a labor market in which firms cannot distinguish between productive workers and lazy ones. If the ongoing market wage reflects the average productivity of a worker, it is clear that a good worker who has an alternative wage which does reflect his productivity will not apply for a job at the ongoing wage rate. Thus, the lemon theorem suggests that only less productive workers apply for jobs. Spence (1974) suggests that good workers may take some acts that will distinguish them from the less productive workers, thereby signaling to the firms that they are productive (signaling is discussed in subsections 8.4.6, 12.6, and 12.7). One act would be to go to college. Although college does not necessarily improve the skill of the worker, going to college may signal to firms that the graduate is a productive worker since unproductive workers may not be able to graduate and therefore would not benefit from investing in education.

12.6 Quality-Signaling Games

Consumers are often unable to recognize the quality of a product before they actually purchase and use the product, even if they are aware that both high-quality and low-quality brands are sold in the market. We refer to such goods as experience goods. Producers, however, have more information regarding the brands they sell and in most cases are fully aware of their product. This creates a problem of asymmetric information that we first analyzed in subsection 8.4.6. In that section, we analyzed an entry-deterrence problem in which the potential entrant did not know the production cost of the incumbent, and the incumbent had to signal its production cost by the price it charged prior to the threat of entry.

In this section we analyze a technically similar problem, in which a monopoly firm knows the quality of the brand it sells, but consumers are unable to learn the brand's quality prior to the actual purchase. Our goal is to demonstrate that a monopoly firm can signal the quality it sells by choosing a certain price and by imposing a quantity restriction on the brand it sells. We should note that signaling models are derived from Spence (1974) (for an application to quality signaling, see Wolinsky 1983).

Suppose that there is a continuum of identical consumers. With no loss of generality we normalize the number of consumers to equal 1. Each consumer buys, at most, one unit and knows that the product can be produced in two quality levels: high ($k = H$) and low ($k = L$), where $H > L > 0$. For a given price denoted by p, the utility function of each consumer is given by

$$U \equiv \begin{cases} H - p & \text{if the brands happens to be of high quality} \\ L - p & \text{if the brands happens to be of low quality} \\ 0 & \text{if he does not purchase the product.} \end{cases} \tag{12.23}$$

Suppose that each consumer goes to the monopoly's store and observes a price level of p dollars. Will consumers purchase the product if they find that $p = H$? Clearly not, since there is a possibility that the product may be of low quality, and in this case (12.23) implies that such a purchase results in a utility level below zero (which is the reservation utility level).

We now describe the monopoly producer side. Denote by c_H the unit production cost of the monopoly if it is a high-quality producer, and by c_L if it is a low-quality one, where $c_H > c_L \geq 0$. That is, the unit production cost of a high-quality product exceeds that of a low-quality product. We make the following assumptions:

Assumption 12.3

1. *The monopolist is a high-quality producer.*
2. *Production costs are sufficiently low relative to consumers' valuation of the two qualities. Formally, $L > c_H$.*

The second part of Assumption 12.3 ensures that a high-quality producer can charge $p = L$ for a high-quality product without making a loss.

We assume that the strategy available to the monopolist is two-dimensional so that it can choose the price (p) and the quantity produced (q). Clearly, $0 \leq q \leq 1$ (since the total number of consumers is

normalized to equal 1). We wish to solve the problem how a high-quality monopolist can sell a high-quality product, given that consumers are not sure whether the brand they buy is a high-quality one. In other words, how can a high-quality producer convince the consumers that he or she does not cheat by selling a low-quality brand for a high price? Hence, in choosing the price and quantity levels, the producer needs to signal his or her high quality to the consumer.

Proposition 12.10 *There exists a pair of a price and a quantity level that convinces consumers (beyond all doubts) that the brand they buy is a high-quality one. Formally, if the monopolist sets*

$$p^m = H \quad \text{and} \quad q^m = \frac{L - c_L}{H - c_L},$$

then (a) consumers can infer that the brand is of high quality, (b) q^m consumers will purchase the product and $(1 - q^m)$ consumers will not purchase the brand due to the lack of supply.

Before proving this proposition, we think it is worthwhile to repeat that the essence of signaling is the firm's to choosing a price-quantity combination that would signal to the consumer that the product is of high quality. In order to do that, the monopoly must choose both a price and a quantity produced that a low quality producer would not find profitable to set! Using this action, the monopoly can convince the consumer that it is not a low-quality producer.

Proof. The monopoly has to show that a low-quality producer would not choose p^m and q^m as the profit-maximizing price and quantity. If the monopolist were a low-quality producer, then he or she could clearly sell to all consumers for the price $p = L$ and make a profit of $\pi^L(L, 1) = 1(L - c_L)$. Let us note that this profit level is attainable by a low-quality producer. Clearly, at this price all consumers would purchase the product.

Now, the question is whether a low-quality monopoly could profitably choose p^m and q^m as the profit-maximizing price and quantity? Suppose it does! Then,

$$\pi^L(p^m, q^m) = (p^m - c_L)q^m = (H - c_L)\frac{L - c_L}{H - c_L} = L - c_L = \pi(L, 1).$$

Thus, using these price and quantity levels, a high-quality monopolist is able to demonstrate that, had he or she been a low-quality producer, he or she could earn the same profit by setting $p = L$ and selling to all consumers instead of setting p^m and q^m. That is, by cutting the profit level to that of what a low-quality monopolist could collect under perfect information, the high-quality producer convinces the consumers that he or she is not a low-quality one, since if he or she were a low-quality producer, he or she could make the same profit level. ■

So far, we have showed that a high-quality producer can signal his or her quality level to the consumer by using the price and quantity instruments, so that consumers' uncertainty is completely resolved. However, this signaling mechanism raises two questions: What is the cost paid by the monopoly to resolve consumers' uncertainty? Would this high-quality monopoly find it profitable to signal its (high) quality level to the consumers?

To answer these questions, we need to calculate the profit level of a high-quality monopolist when he or she sets p^m and q^m. Hence,

$$\pi^H(p^m, q^m) = (p^m - c_H)q^m = (H - c_H)\frac{L - c_L}{H - c_L} < H - c_H.$$

Hence, comparing this profit level to the profit under perfect information $(H - c_H)$ yields the cost of revealing information. The answer to our second question depends on whether

$$(H - c_H)q^m = (H - c_H)\frac{L - c_L}{H - c_L} \geq (L - c_H)1. \tag{12.24}$$

Cross-multiplying (12.24) yields that this inequality always holds, since $H > L > c_H > c_L$.

Criticism of the quality-signaling model

The quality-signaling model developed in this section is used only for the sake of illustration. Note that if a firm can choose whether to become a low-quality or a high-quality producer, it would choose to be a low-quality producer. That is, since a high-quality producer needs to signal his quality, and since production cost is higher, it becomes more profitable to be a low-cost producer. This model can be modified to capture profitable signaling by adding consumers who purchase only high quality goods.

12.7 Warranties

One common method of insuring the consumer against defects in the product is to bundle the product with a warranty. There are many kinds of warranties. Some warranties restrict the manufacturer's liability only to parts, others to labor and parts, in case that repair is needed. Most warranties are limited to a certain time period after the purchase, whereas few provide a lifetime warranty. We shall not discuss in the

present section why most warranties are limited. The reason has to do with the *moral hazard* phenomenon, a situation where a full warranty will provide the consumer with the incentives to misuse the product, or not to take proper care of it (see Cooper and Ross 1985). Therefore, in order to demonstrate the role of warranty in market behavior, we make the following assumption:

ASSUMPTION 12.4

1. *The product can be either fully operative or fully defective. A defective product has no value to the buyer and cannot be resold for scrap.*

2. *At the time of purchase neither sellers nor buyers know whether the specific product is defective.*

3. *The manufacturer/seller has two options regarding the sale of the product:*

   (a) *He or she can sell the product without a warranty. In this case, if the specific product is found to be defective, the buyer loses the entire value of the product.*

   (b) *He or she can sell the product with a full replacement warranty, which guarantees full replacement of a defective product with no loss of value to the buyer. That is, if the replacement product is also found to be defective, the monopoly is obligated to replace the replacement product, and so on.*

In the literature, Grossman (1980) provides a comprehensive analysis of a monopoly that can offer a warranty for the product it sells. Spence (1977) builds on a signaling argument and shows that higher-quality firms offer a larger warranty than do low-quality firms. In what follows, we confine our analysis to a monopoly selling a product to a competitive consumer, where the product has a certain probability of being defective. The next subsection discusses the monopoly optimal provision of warranty under symmetric information between the buyer and the seller. A subsequent subsection analyses a market in which warranties can serve as a (partial) signal of the product's quality.

12.7.1 Warranties under symmetric information

Consider a product whose value to the consumer is V if the product is operative, and 0 if the product is defective, where $V > 0$. Suppose that there is a known probability for products of this type to be functional. We denote this probability by ρ, where $0 < \rho < 1$. Thus, with probability

$(1-\rho)$, the product produced by the monopoly will be found to be defective. In this subsection we assume that the seller and the buyer have symmetric information regarding the product's reliability, meaning that both the seller and the buyer know the product is reliable with an exogenously given probability ρ.

Let p denote the monopoly price and $c > 0$ denote the unit production cost of the product. We assume that the utility function of the consumer is the expected value of the product minus the product's price, if he buys the product, and zero if he does not buy the product. Formally,

$$U \equiv \begin{cases} V - p & \text{if he buys the product with full replacement warranty} \\ \rho V - p & \text{if he buys the product without any warranty} \\ 0 & \text{if he does not buy.} \end{cases} \tag{12.25}$$

Finally, we assume that $\rho V > c$, which implies that the expected utility from the product exceeds the unit production cost. Assuming otherwise would yield that the product will not be produced since the monopoly will not be able to get consumers to pay a price exceeding unit cost.

The profit-maximizing monopoly has the option of selling the product with or without a warranty.

No warranty

With no warranty, (12.25) implies that the maximum price the monopoly can charge is the expected value of the product. Thus, if we assume one consumer, then under no warranty the monopoly price and profit level are given by

$$p^{NW} = \rho V \quad \text{and} \quad \pi^{NW} = \rho V - c. \tag{12.26}$$

Warranty

When the monopoly provides the consumer with a full replacement warranty, under Assumption 12.4 the consumer is assured of gaining a value of V from the product. We need the following Lemma.

Lemma 12.1 *The expected unit production cost for a firm providing a full replacement warranty is c/ρ.*

Proof. The cost of producing the product is c. If the product is defective, expected cost increases by $(1-\rho)c$. If the replacement product is defective, then expected cost increases again by $(1-\rho)^2 c$, and so on. Hence, expected cost is given by

$$c + (1-\rho)c + (1-\rho)^2 c + (1-\rho)^3 c + \cdots = \frac{c}{1-(1-\rho)} = \frac{c}{\rho}. \tag{12.27}$$

■

Thus, Lemma 12.1 implies that the expected production cost is c when $\rho \to 1$ (zero failure probability), and becomes infinite as $\rho \to 0$ since in this case the product is produced and replaced infinitely many times. Altogether, the maximum price a monopoly can charge and the profit level are given by

$$p^W = V \quad \text{and} \quad \pi^W = V - \frac{c}{\rho}. \tag{12.28}$$

Will the monopoly sell with a warranty?

Comparing (12.26) with (12.28) yields the conclusion that $\pi^W > \pi^{NW}$ if $V > c/\rho$, which must hold for the monopoly to make profit under any warranty policy. Hence, we can conclude the analysis with the following proposition:

Proposition 12.11 *Under symmetric information where the reliability parameter ρ is common knowledge, a monopoly will always sell the product with a warranty.*

The intuition behind Proposition 12.11 is as follows. When the monopoly provides a warranty, the monopoly can increase the price by $(1-\rho)V$ above the price selling with no warranty. The associated increase in cost is

$$\frac{c}{\rho} - c = \frac{(1-\rho)c}{\rho} < (1-\rho)V = p^W - p^{NW}$$

by assumption. Hence, by providing a warranty, and given that the monopoly extracts all consumer surplus, the monopoly can increase its price by more than its increase in the cost associated with replacing the products with a certain probability of failure. In other words, consumers are willing to pay more for the warranty than what it costs the seller.

12.7.2 The role of warranties under asymmetric information

In section 12.5 we encountered the problem of asymmetric information between sellers and buyers, where we assumed that sellers are generally better informed about the product's quality than the buyers. Since consumers are not informed, they cannot distinguish between highly reliable products (products with a high probability of not breaking down) and products with a high defective rate.

In this subsection, we continue with the exploration of markets with asymmetric information and analyze a duopoly in which one firm produces a reliable product (high probability of being operative) and one firm produces an unreliable product (with a low probability of being operative). However, the consumer does not have any way of knowing which one of the firms produces the more reliable product. That is, the consumer cannot distinguish between the two products. We show that by providing a warranty with the product and choosing a certain price, the high-quality firm can signal to the consumer that it is selling the more reliable product. In this case, the consumer can conclude beyond all doubt that the high-quality firm is indeed a high-quality producer and not a low-quality producer masquerading as a high-quality firm.

The signaling principle always remains the same: if a high-quality producer wants to prove to the consumer that he or she is a high-quality producer, he or she has to carry an act that is unprofitable for a low-quality producer. From this act, the consumer will conclude that the producer does produce a high-quality product and will be willing to pay for the product accordingly.

Consider an economy with two producers. A high-quality producer selling a product with probability ρ_H of being operative, and a low-quality producer producing a product with probability ρ_L of being reliable, $0 < \rho_L < \rho_H < 1$.

No warranties

Since the consumer cannot distinguish between the producers before the purchase, both products (high and low reliability) are sold for the same price. In this case, since from the consumer's point of view the products are homogeneous before the purchase, a Bertrand price competition (see section 6.3) leads to a unique equilibrium where prices equal the unit cost, hence, zero profits. That is, $p^{NW} = c$ and $\pi_i^{NW} = 0$, $i = H, L$. Therefore, with equal production cost, both high- and low-quality products are produced and the high quality manufacturer cannot be identified by the consumer.

Warranty as a signal

We now show that by providing a warranty and choosing an appropriate price, the high-quality producer can signal to the consumer that he or she sells a reliable product.

Proposition 12.12 *Let $V > c$. The high-quality producer can push the low quality producer out of the market by setting $p^W = c/\rho_L$ and providing a warranty. In this case the consumer will buy only the more reliable product, and the high-quality producer will make a strictly positive profit.*

Proof. We first show that a low-quality producer will not find it profitable to sell his or her product with a warranty at this price. To see

that, using (12.27), we calculate that

$$\pi_L^W(p^W) = p^W - \frac{c}{\rho_L} = 0.$$

This concludes the main part of the proof. To complete the proof, we need to verify that first, the consumer will indeed prefer purchasing the more reliable product with a warranty instead of the less reliable product at the lowest possible price, $p = c$; and second, the high-quality producer makes an above zero profit. To see this, observe that the profit of the high-quality firm is given by

$$\pi_H^W = p^W - \frac{c}{\rho_H} = \frac{c}{\rho_L} - \frac{c}{\rho_H} > 0.$$

Finally, the utility of a consumer buying the more reliable product exceeds the utility of buying the less reliable product without a warranty even if the less reliable product has the lowest possible price, c, since

$$U_H^W = V - p^W = V - \frac{c}{\rho_L} > U_L^{NW} = \rho_L V - c.$$

■

12.8 Appendix: The Legal Approach to Products Liability

In this section, we briefly describe the legal approach to products liability, which is concerned with defective products and trades. The reader interested in learning all the legal issues concerning liability should consult Howard 1983 and Phillips 1988 for a comprehensive analysis of product liability. *Liability* refers to the obligation of the producer or the merchant seller to those who were damaged as a result of a defective product. Note that those damaged need not be only the buyers, but could also be bystanders and owners of property.

12.8.1 Defects and liability

In general, there are four types of defects: production defects, design defects, erroneous operating instructions and warnings, and mislabeling and misrepresentations of products. Thus, liability law extends the liability beyond what are purely understood as manufacturing flaws. Clearly, these distinctions are hard to make, but they seem to be important in deciding what standard of liability (strict liability or negligence) is assumed for the manufacturer. For example, it seems more likely that strict liability is generally imposed for production defects than for design defects. Misrepresentation defects may or may not be judged under strict liability.

Under these classifications, it is necessary to determine whether the product is defective. The most common way to make that determination is to rely on consumer expectations, meaning that the product sold must be more dangerous than the "ordinary consumer" with the "ordinary" knowledge common to the community would expect it to be. A problem may arise when the consumer buys products known to be dangerous, since an ordinary consumer should expect the danger associates with this product. Another way of determining defectiveness is to ask whether the seller would have sold the product had he or she known the potential harm resulting from the sale. Thus, in this case, defectiveness is defined as a presumed knowledge by the seller about the quality of the product. Defectiveness can also be determined by determining whether the producer invested a sufficient amount in preventing a risk, where sufficiency estimated by balancing the cost of preventive investment and the monetary value of the inflicted damage or risk caused by the product in the condition it was sold.

Liability is not limited solely to the producer. Liability may be assumed to rest on any commercial seller such as dealers, vendors, constructors, stores, and so on. However, strict liability is less likely to be imposed on them, since the presumed knowledge of the seller is smaller than that of the maker of the product.

12.8.2 Warranties

The Uniform Commercial Code states that unless excluded or modified, a warranty is implied in the contract of sale. However, a warranty is not implied if the seller is not a merchant. The implied warranty, which attaches strict liability to the seller is important since it reduces the chances that written agreements (such as warranty certificates or disclaimers) would always be effective in reducing the seller's liability. In order for the seller to reduce his or her liability to a level below that assumed in the implied warranty, he or she has to provide a disclaimer; however, a disclaimer is not always accepted by courts. A disclaimer is generally accepted in the case of negligence on the part of the consumer. Also, a disclaimer is valid only with respect to the trading parties (buyers), not, for example, with respect to bystanders.

Since warranties have been recognized as a special source of the deception of consumers, the FTC has issued several rules, some of which have been adopted as laws, that require that the terms of the guarantee will be clear and presented in a clear fashion. All of us who have been

given warranties can imagine that the amount of information that has to be included in a warranty must be enormous. For example, what is the time interval corresponding to a life time warranty that we often see on back of our packages? What is meant by full warranty? Does a full warranty include labor cost, parts, freight, or the loss of time associated with the loss of use?

12.9 Exercises

1. Consider the modified Hotelling vertical-differentiation model of subsection 12.2.2, but suppose that consumers have a reservation utility, in the sense that a consumer prefers not to buy any brand if his or her utility falls below zero. Recall that the preferences exhibited in (12.2) imply that there is no lower bound on utility from consumption. Figure 12.3 implies that this modification in preferences would not affect the number of high-quality-brand buyers, since all consumers indexed on $[\hat{x}, 1]$ gain a strictly positive utility from buying the high-quality brand. However, point z in Figure 12.3 shows that no consumers indexed on $[0, z]$ will purchase any brand, since otherwise their utility falls below zero.
   Perform the following exercises:
   (a) Show that for given a, b, p_A and p_B, the number of consumers who do not purchase any brand equals to $z = p_A/a$.
   (b) Conclude that the market share of firm A is
   $$\hat{x} - z = \frac{p_B - p_A}{b - a} - \frac{p_A}{a}.$$
   (c) Using the same procedure as in (12.5), show that for given a and b, the second-period equilibrium prices (and profit levels) are given by
   $$p_A(a,b) = \frac{a(b-a)}{4b-a}, \quad \text{and} \quad p_B(a,b) = \frac{2b(b-a)}{4b-a}$$
   $$\pi_A(a,b) = \frac{ab(b-a)}{(4b-a)^2} \quad \text{and} \quad \pi_B(a,b) = \frac{4b^2(b-a)}{(4b-a)^2}.$$
   (d) Show that in the first period, firm A would choose to locate at $a^e = 4/7$, whereas firm B would locate at $b^e = 1$.
2. Prove the second part of Proposition 12.1 using the same procedure as the one used in the proof of the first part.
3. Consider the lemon model described in section 12.5 and suppose that the owner of the good used car must sell his or her car because he or she is leaving the country. Assume that the market prices of used and new cars are exogenously given by $0 < \bar{p}^U < U^G/2$, and $\bar{p}^N = N^G/2$, respectively. Characterize the demand and supply patterns of the four types of agents under these prices.

4. Consider the monopoly's warranty problem under symmetric information analyzed in subsection 12.7.1 but assume that for some reason the monopoly cannot guarantee more than one product replacement in case the product purchased is found defective. That is, if the product is found defective, the monopoly can provide a warranty to replace the product with a new product; however, if the replacement product fails, then the monopoly cannot replace the replacement product.
   (a) What is the monopoly's expected cost if it provides this type of warranty?
   (b) What is the maximum price the monopoly can charge for the product sold with this type of warranty?
   (c) Conclude whether Proposition 12.11 holds for this type of warranty.

12.10 References

Akerlof, G. 1970. "The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 89: 488–500.

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

Cooper, R., and T. Ross. 1985. "Product Warranties and Double Moral Hazard." *Rand Journal of Economics* 16: 103–113.

Fishman, A., N. Gandal, and O. Shy. 1993. "Planned Obsolescence as an Engine of Technological Progress." *Journal of Industrial Economics* 41: 361–370.

Gabszewicz, J., and J. Thisse. 1979. "Price Competition, Quality and Income Disparities." *Journal of Economic Theory* 20: 340–359

Gabszewicz, J., and J. Thisse. 1980. "Entry (and Exit) in a Differentiated Industry." *Journal of Economic Theory* 22: 327–338.

Grossman, S. 1980. "The Role of Warranties and Private Disclosure about Product Quality." *Journal of Law and Economics* 24: 461–483.

Howard, M. 1983. *Antitrust and Trade Regulation Selected Issues and Case Studies.* Englewood Cliffs, N.J.: Prentice-Hall.

Kihlstrom, R., and D. Levhari. 1977. "Quality, Regulation, Efficiency." *KYKLOS* 30: 214–234.

Kleiman, E., and T. Ophir. 1966. "The Durability of Durable Goods." *Review of Economic Studies* 33: 165–178.

Levhari, D., and Y. Peles. 1973. "Market Structure, Quality, and Durability." *Bell Journal of Economics* 4: 235–248.

Levhari, D., and T. N. Srinivasan. 1969. "Durability of Consumption Goods: Competition versus Monopoly." *American Economic Review* 59: 102–107.

Oi, W. 1973. "The Economics of Product Safety." *Bell Journal of Economics* 4: 3–28.

Phillips, J. 1988. *Products Liability in a Nutshell.* 3rd ed. St. Paul, Minn.: West Publishing Co.

Phlips, L., and J. Thisse. 1982. "Spatial Competition and the Theory of Differentiated Products: An Introduction." *Journal of Industrial Economics* 31: 1–11.

Schmalensee, R. 1979. "Market Structure, Durability, and Quality: A Selective Survey." *Economic Inquiry* 17: 177–196.

Shaked, A., and J. Sutton. 1982. "Relaxing Price Competition Through Product Differentiation." *Review of Economic Studies* 49, 1–13.

Spence, M. 1974. *Market Signaling.* Cambridge, Mass.: Harvard University Press.

Spence, M. 1975. "Monopoly, Quality, and Regulation." *Bell Journal of Economics* 6: 417–429.

Spence, M. 1977. "Consumer Misperceptions, Product Failure, and Producer Liability." *Review of Economic Studies* 44: 561–572.

Swan, P. 1970. "Durability of Consumer Goods." *American Economic Review* 60: 884–894.

Swan, P. 1970. "Market Structure and Technological Progress: The Influence of monopoly on Product Innovation." *Quarterly Journal of Economics* 84: 627–638.

Swan, P. 1971. "The Durability of Consumer Goods and the Regulation of Monopoly." *Bell Journal of Economics* 2: 347–357.

Wolinsky, A. 1983. "Prices as Signals of Product Quality." *Review of Economic Studies* 50: 647–658.

Chapter 13

Pricing Tactics: Two-Part Tariff and Peak-Load Pricing

> People want economy, and they'll pay any price to get it.
> —Attributed to Lee Iacocca

> You'd be surprised how much it costs to look this cheap.
> —Attributed to Dolly Parton

The pricing techniques discussed in this chapter are generally studied under the subject of public-utility pricing, where a regulating agency (such as the state, city, or any other local government) controls the prices and quality of service provided by the public utility. However, as the reader will discover, these pricing techniques are also used by unregulated and privately owned firms. The major difference between regulated public-utility pricing and prices chosen by privately owned firms is that a regulator attempts to choose prices intended to maximize consumer welfare, whereas unregulated firms choose prices to maximize profit. As it turns out, in many cases the regulator and an unregulated monopoly will choose to set similar price structures that may differ only by a lump-sum transfer from consumers to firms.

In what follows we study several pricing techniques employed by unregulated, profit-maximizing firms. Section 13.1 (Two-Part Tariff) analyzes why sports clubs tend to charge annual membership fees instead of (or in addition to) fixing a price per visit. Two-part tariffs are also charged by some cable TV companies and by wholesale club stores. Section 13.2 (Nonuniform Pricing) generalizes the two-part tariff to the case of heterogeneous consumers and demonstrates how quantity discounts can increase firms' profit by extracting higher surplus from different consumer groups. Section 13.3 (Peak-Load Pricing) analyzes firms' choices of capacity and prices when the demand is seasonal, for example, the choices of airline firms, car rental companies, hotels, resorts, regulated and unregulated phone and electricity companies, universities (day versus evening classes), movie theaters, restaurants, and many others. Section 13.4 (Can Firms "Control" the Seasons?) concludes with an extension of the peak-load pricing problem by having firms set prices to manipulate the relative quantity demanded between seasons.

13.1 Two-Part Tariff

It has been observed that many commercial enterprises charge annual membership dues instead of (or in addition to) pricing each unit of consumption separately. This phenomenon is observed mostly in entertainment industries—such as amusement parks, most sports clubs, and some theaters—and recently in wholesale clubs. Oi (1971) proposed an explanation for this observation. Given downward-sloping demand, when a monopoly charges a fixed price per unit of consumption, if consumers purchase the product, then they gain positive consumer surplus (see subsection 3.2.3 on page 52). Thus, even when a monopoly charges its profit-maximizing price, it is unable to extract the entire consumer surplus. Therefore, in addition to the per unit price, a monopoly firm needs to set a second pricing instrument in order to be able to extract the entire consumer surplus.

13.1.1 Club-visiting consumers

Suppose that a consumer gains satisfaction from club visits and from other goods which we term as money. We denote by Q the number of club visits, and by m the amount of money spent on other goods. Let the consumer earn a fixed income of $\$I$ to be spent entirely on club visits and other goods. We denote by ϕ the membership dues and by p the price per visit. Thus, the consumer's budget constraint is given by

$$m + \phi + pQ \le I. \tag{13.1}$$

The utility of our fun-loving consumer is a function of the number of club visits (Q) and the consumption of other goods (m). We assume a *quasi-linear* utility function given by

$$U \equiv m + 2\sqrt{Q}. \tag{13.2}$$

Figure 13.1 illustrates a set of indifference curves derived from this utility function. In Figure 13.1, the indifference curve U_0 originating

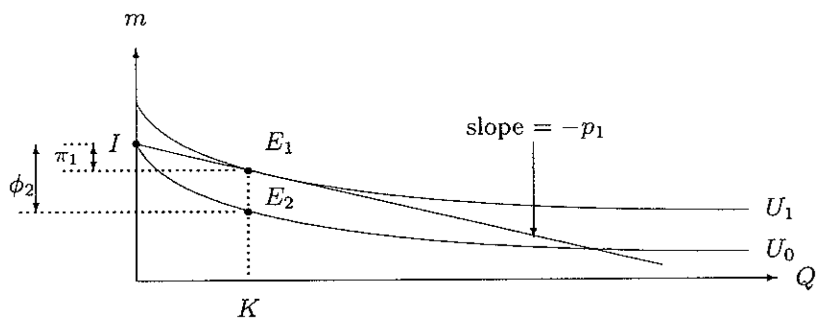

Figure 13.1: Quasi-linear utility indifference curves

from the income level I is associated with the initial utility from spending all the income I on other goods. This indifference curve shows the combinations of club visits and spending on other goods that leave the consumer neither better off nor worse off than spending all the income on other goods.

We now derive the consumer's demand curve for club visits. Substituting (13.1) into (13.2) for m yields the consumer-utility-maximization problem. Hence, for given p and ϕ at a sufficiently low level, the consumer chooses the number of visits Q that solves

$$\max_Q U = I - \phi - pQ + 2\sqrt{Q} \tag{13.3}$$

yielding a demand function

$$p = \frac{1}{\sqrt{Q^d}}, \quad \text{i.e.} \quad Q^d = \frac{1}{p^2}. \tag{13.4}$$

13.1.2 No club annual membership dues

Suppose that the club has a limited capacity. Formally, assume that the club's capacity is limited to K visitors, $K > 0$. We now suppose that the club has only one method of collecting money from the club visitors, which is charging a price p per visit where club membership is not required. That is, the club sets the annual membership dues to $\phi = 0$.

When $\phi = 0$ the monopoly club chooses Q to maximize

$$\pi \equiv pQ = \frac{1}{\sqrt{Q}}Q = \sqrt{Q}. \qquad (13.5)$$

Proposition 13.1 *Under the preferences given in (13.2), the monopoly club sets the price so that the demand for club visits equals its capacity. Formally,*

$$p_1 = \frac{1}{\sqrt{K}}, \quad \textit{and } Q_1 = K, \quad \textit{and hence } \pi_1 = \sqrt{K}.$$

Proof. The preferences (13.2) yield an elastic demand curve (13.4), implying that the club's profit rises with the number of visits. Therefore, the club will operate under full capacity. The consumption point is illustrated in Figure 13.1 at the point E_1, where the price line (budget constraint) is tangent to the indifference curve labeled U_1, $U_1 > U_0$. Hence, under a price-per-visit structure with no membership fees, the welfare of the consumer must increase compared with the no-club-visits allocation. As we show below, this is not necessarily the case when club charges involve annual fixed dues. ■

13.1.3 Annual membership dues

Annual membership fees (fixed-part tariff) is in fact a bundling method discussed in section 14.1, in which the club offers the consumer the opportunity to pay a fixed amount of $\phi > 0$ and to receive a package containing a fixed number of "free" visits.

Figure 13.1 shows that for a package containing $Q = K$ number of visits, the consumer is willing to pay a maximum amount of ϕ_2. That is, consuming a package of K visits for a fixed fee of $\phi \le \phi_2$ would leave the consumer no worse off than he or she would be with the no-club-visits case.

We now calculate the maximum annual fee that the club can charge for K number of visits that make it worthwhile for the consumer to purchase. To do that, let us observe that in Figure 13.1, by construction, the point E_2 lies on the initial indifference curve U_0. That is, the club sets ϕ just about the level where the consumer is neither better off nor worse off by joining the club. Formally, the club sets ϕ_2 that solves

$$\max_{\phi} \pi(\phi) = \phi \ \textbf{s.t.} \quad I - \phi + 2\sqrt{K} \ge I = U_0, \qquad (13.6)$$

implying that $\phi_2 = 2\sqrt{K}$ and hence $\pi_2 = 2\sqrt{K}$. Hence,

Proposition 13.2 *A fixed fee for a bundle of visits yields a higher profit to the club than any profit generated with a per unit price with no annual fee. Formally,*

$$\pi_2 \equiv \pi(\phi = \phi_2, p = 0) = 2\sqrt{K} > \sqrt{K} = \pi(\phi = 0, p = 1/\sqrt{K}) = \pi_1.$$

13.1.4 Two-part tariff

In practice, a club would hesitate charging exactly ϕ_2 as a membership fee mainly because a small mistake in estimating the exact location of the indifference curve U_0 or the consumer's income may result in no sales at all. A second reason why a firm would not use only a fixed membership fee is that consumers may have heterogeneous preferences so that a high membership fee may induce only a partial participation. We therefore conclude that clubs would generally charge a lower fee than the maximum fee calculated in the earlier subsection. For example, Figure 13.2 demonstrates a possible "package" of Q_3 club visits for an annual fee equal to ϕ_3. Clearly, the consumer buys such a package.

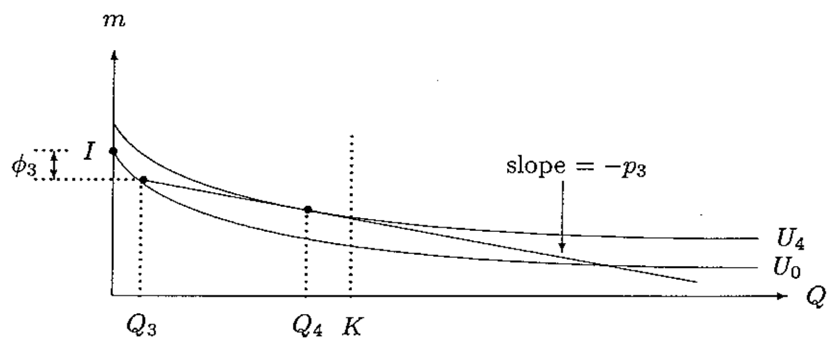

Figure 13.2: Pure two-part tariff club charges

However, Figure 13.2 also shows that the club can further increase its profit by supplementing the membership fee ϕ_3 with an option to purchase additional visits for a price of p_3 per unit. In this case, for p_3 that is not too high, the consumer purchases additional visits, bringing the total number of visits to Q_4, as illustrated in Figure 13.2.

13.2 Nonuniform Pricing

Section 13.1 demonstrated how a two-part tariff can increase firms' profit above the monopoly's per-unit-price profit level by employing two price instruments: the conventional per unit price and the lump-sum (consumption independent) fixed membership dues. The profit gains from using the two-part tariff are due to the monopoly's ability to extract higher surplus from a given group of homogeneous consumers.

In this section we demonstrate a price strategy commonly used by firms to price discriminate among heterogeneous groups of consumers. The *nonuniform price schedule* is a tariff for one or more goods in which the consumer's total outlay does not simply rise proportionately with the amounts of goods the consumer purchases. That is, a nonuniform price schedule consists of quantity discounts and quantity premiums, (for extensive analysis of nonuniform pricing, see Brown and Sibley 1986).

Figure 13.3 illustrates the (inverse) demand for local phone calls by two different groups: households and business, given by $p_H = 12 - 2q_H$ and $p_B = 6 - q_B/2$, respectively, where prices are given in cents. Assuming zero marginal cost in providing phone services, section 5.3 on page 75 shows that a monopoly selling in two segmented markets (markets in which arbitrage cannot take place), would set quantity produced in each market by equating

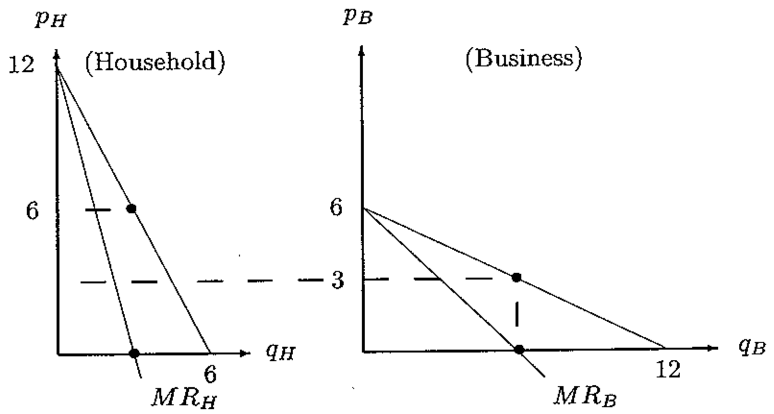

Figure 13.3: Nonuniform pricing and price discrimination

$$MR_H(Q_H) = MR_B(Q_B) = MC(Q_H + Q_B) = 0,$$

thereby charging different prices in the two markets given by $p_H = 6$ and

$p_B = 3$ and producing $q_H = 3$ and $q_B = 6$. Therefore, if the monopoly can price discriminate, it would charge business lower rates than it would charge households for local phone calls.

The problem facing the monopoly is how to set the price schedule in a way that would induce the two different groups of consumers to pay different prices and to consume different quantities. In general, there are many reasons why a firm may not be able to charge different prices to different groups of consumers, for example, price discrimination is illegal under the Clayton Act (see subsection 5.6.3); also a monopoly may not be able to identify the consumers belonging to a particular group. Altogether, we now demonstrate that nonuniform pricing can generate the price discrimination monopoly outcome even when the monopoly does not directly discriminate among the different groups of consumers or cannot simply identify these groups.

We now investigate the price schedule illustrated in Figure 13.4.

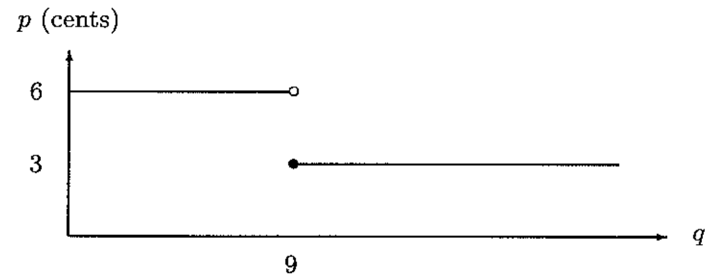

Figure 13.4: Nonuniform price schedule

Proposition 13.3 *Consider the price-per-call schedule illustrated in Figure 13.4 and formally given by*
Regular Rate Program: *Pay 6 cents per phone call.*
Quantity Discount Program: *Pay a reduced rate of 3 cents per phone call but be charged for at least 9 phone calls.*
Then, this price schedule yields the same market prices as those charged by a discriminating monopoly.

Proof. Clearly, Figure 13.3 implies that when $p_H = 6$, households demand $Q_H = 3$ phone calls, and given $p_B = 3$, business customers demand $Q_B = 6$ phone calls. We need to show to show that households will not benefit from adopting the quantity-discount price scheme. If households adopt the regular rate, their consumer surplus (subsection 3.2.3 on page 52) is $CS(6) = (6 \times 3)/2 = 9$.

If households adopt the discount rate, then they are "forced" to buy 9 phone calls (and actually use only 6), which makes the gross-consumer surplus equals the entire area under the demand curve given by $(12 \times 6)/2$. Since households are required to pay for 9 phone calls, their net consumer surplus is

$$CS_H(\text{discount}) = \frac{12 \times 6}{2} - 3 \times 9 = 9 = CS_H(6).$$

Given that the households are indifferent between the two plans, we can assume that they do not purchase the discount plan.

Clearly, when $p = 6$, businesses will purchase zero on the regular payment program. However, when they choose the discount plan

$$CS_B(\text{discount}) = \frac{(6 - 1.5)9}{2} + 1.5 \times 9 - 3 \times 9 = 6.75 > 0.$$

Hence, businesses will choose the discount plan.

Finally, it can be shown that this monopoly phone company makes a higher profit under nonuniform pricing than under uniform pricing. ■

13.3 Peak-Load Pricing

The problem of peak-load pricing is generally studied in the context of optimal governmental regulations for public companies such as public utilities, including phone, transportation and electricity companies (see Brown and Sibley 1986; Joskow 1976; Sherman 1989; and Steiner 1957). However, it should be emphasized that unregulated firms also benefit from setting peak-load pricing, simply because peak-load pricing tends to be efficient and profitable when demand is periodic, and when the investment in capacity is irrevocable in the short run. For example, private firms such as hotels, restaurants, sports clubs, movie theaters, and airlines and other transportation companies are all subject to seasonal demand schedules that vary between yearly seasons, days of the week, or the hours of the day. We therefore focus our analysis on a private-sector monopoly firm (which could represent an airline, a hotel, or a restaurant) and then conclude with a discussion on the role of the regulator in controlling the prices.

Three factors characterize the peak-load pricing problem: First, the levels at which demand varies between periods. Second, capital has to be rented or leased for a long period. That is, since the firm must commit in advance to the level of the plant's capacity, and since this commitment cannot be reversed between periods, the duration of these contracts affect firms' seasonal pricing decisions. Third, the firm's output (products or services) is too costly or impossible to store. Otherwise, if the output is storable, then the firm could produce equal amounts in each period (or all the output in a single period) and then allocate the output across periods according to demands.

Consider a monopoly airline company flying on a single route during high (H) and low (L) seasons.

13.3.1 Seasonal passengers

We let p^H, Q^H, p^L, and Q^L denote the price and quantity of tickets in the high and low seasons, respectively. The demand for flights in each season is given by

$$p^H = A^H - Q^H \quad \text{and} \quad p^L = A^L - Q^L, \quad A^H > A^L > 0. \tag{13.7}$$

Figure 13.5 illustrates the seasonal demand structure.

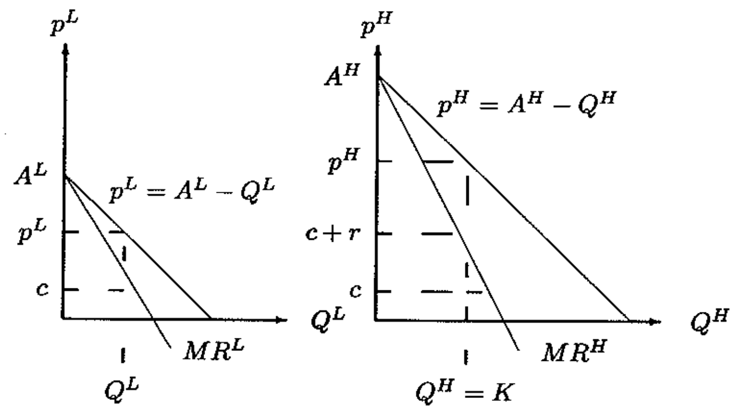

Figure 13.5: Seasonal demand structure and monopoly peak-load pricing

13.3.2 Seating capacity and the airline's cost structure

The monopoly airline faces two types of costs: Capacity cost, which is the number of airplane seats the airline rents for the entire year, and variable cost, which is the cost associated with handling each passenger, which includes check-in, luggage, and food services. For simplicity we ignore other costs commonly associated with airline operations, such as airport charges (see section 17.2 of an analysis of the airline industry).

We denote by r, $r > 0$, the unit capacity cost. Thus, if the airline rents aircraft capacity that can fly K passengers throughout the year, its total capacity investment cost is rK. We denote by c the operational (variable) cost per passenger. Thus, assuming that seating capacity cannot be rented for less than one year (high and low seasons together), the airline's total cost when it flies Q^H passengers in the high season, and Q^L in the low season is

$$TC(Q^H, Q^L, K) = c(Q^H + Q^L) + rK \quad \text{for} \quad 0 < Q^L, Q^H \le K. \quad (13.8)$$

Equation (13.8) highlights the difference between a two-market, discriminating monopoly analyzed in section 5.3 and the present problem, in which a monopoly airline faces two independent seasonal markets. The difference between the analysis of section 5.3 and this problem follows from the fact that investment in capacity for the high season implies that no investment in seating capacity is needed for the low season. Thus, (13.8) implies that the airline monopoly cost structure exhibits *joint production* where production cost in one market also (partially) covers the cost of producing in a different market (different season).

13.3.3 Profit-maximizing seasonal airfare structure

In section 5.3 we proved that a monopoly discriminating between markets determines the price charged and quantity produced for each market by equating the marginal revenue in each market to its marginal cost. However, how should we calculate the airline's marginal cost in the present case? Clearly, the operational cost (c) is part of the unit cost, but how do we allocate the unit-capacity cost between the markets? The following proposition assumes that the low-season demand is significantly lower than the high-season demand, see Steiner (1957).

Proposition 13.4 *Suppose that the low-season demand is significantly lower than the high-season demand. Then, the monopoly's profit maximizing seasonal pricing and output structure is determined by*

$$MR^H(Q^H) = c + r \quad \textit{and} \quad MR^L(Q^L) = c, \quad \textit{where} \ Q^H > Q^L; \quad \textit{and}$$

$$p^H = \frac{A^H + c + r}{2} > \frac{A^L + c}{2} = p^L.$$

That is, capacity is determined only by the high-season demand, where the high-season marginal revenue equals the sum of the operational and capacity (marginal) costs.

Proof. Clearly, given the linear shift of demand between the seasons, the profit-maximizing output levels satisfy $Q^H > Q^L$. Hence, $K \ge Q^H > Q^L$, meaning that in the low season the airline does not fly at full capacity. Consequently, the marginal cost of flying one additional passenger in the low season is independent of k. Hence, according to section 5.3,

the profit-maximizing low-season number of serviced passengers is determined by the condition $MR^L(Q^L) = c$. Therefore, investment in capacity is determined only by the high-season demand, so if we follow section 5.3, the monopoly sets is $MR^H(Q^H) = c + r$. ∎

13.3.4 Peak-load pricing and efficiency

Many utility companies (gas, local phone, electricity, and transportation) are regulated in most states, and they have to adhere to price schedules determined by the corresponding government. Most states require that utility companies (especially electricity) submit variable-load price structures based on the (efficient) marginal-cost pricing principle. If we move to the regulator's problem, we discover that the fact that the monopoly faces periodic demand schedules does not complicate the problem beyond the regulator's problem when the monopoly faces a stable demand. Thus, given that marginal-cost pricing is efficient, Proposition 13.4 tells us that the regulator should set the price in the high season to $p^H = c + r$ and in the low season to $p^L = c$. Thus, efficient pricing requires that high-season consumers pay the marginal operational plus the marginal capacity costs, whereas low-season consumers pay only the marginal operational cost.

13.3.5 Peak-load pricing over longer periods

So far our analysis has concentrated on a time period where there is only one low season and only one high season. Suppose that the airline firm is required to invest in capacity for n years, $n > 1$, so that capacity holds for n low seasons and n high seasons. In this case, what would be the profit-maximizing pricing structure for this monopoly airline?

Proposition 13.5 *The monopoly's profit-maximizing seasonal pricing and output structure over n low and n high seasons is determined by*

$$MR^H(Q^H) = c + r/n \quad \textit{and} \quad MR^L(Q^L) = c.$$

Thus, if the monopoly expects that the capacity would be maintained for n high seasons, the effective unit capacity cost in each period should be taken as k/n.

13.3.6 Limitation of our peak-load pricing analysis

Some limitations of the traditional approach to peak-load pricing analysis are listed in Bailey and White 1974 and Bergstrom and MacKie-Mason 1991. A serious limitation of this analysis is that we neglected to analyze the markets with periodic demand schedules when the different seasonal prices induce consumers to substitute high-season consumption for low-season consumption. High substitutability between peak and off-peak hours is most noticeable in the telephone industry, where individuals postpone making personal phone calls until late at night, early in the morning, and on weekends. Thus, our analysis is incomplete, since it assumes that the demand for peak-season service is independent of the off-peak price.

13.4 Can Firms "Control" the Seasons?

Peak-load prices are generally calculated by assuming that peak and off-peak periods are exogenously given. Although this assumption may describe some public utilities where the regulating authority decides on which periods are considered peak and which off-peak (such as electricity and the telephone), most firms get to control the quantity demanded in each period by simply adjusting the relative prices in the different periods/seasons. For example, by substantially reducing winter airfare, airline firms can potentially turn a low season into a high season. Restaurants control the flow of customers by substantially reducing the price of lunch compared with the price of a dinner. Car rental companies can turn the weekend into a high-demand period by substantially reducing weekend rents to attract nonbusiness-related renters during the weekends.

All these examples lead to one conclusion, namely, peak and off-peak periods should be regarded as economic variables and therefore should not be assumed.

In this section we calculate peak-load prices in an environment where the selling firm can use the pricing structure to manipulate which period will be the peak and which will be off-peak. We analyze what would be the profit-maximizing pricing structure chosen by a service-providing monopoly. There are two reasons why we should analyze the monopoly case: First, analyzing the monopoly case helps us to capture the intuition about the tradeoff between consumers' preferences towards certain period services and the cost of maintaining capacity. Second, many utility and transportation companies are (regulated or unregulated) monopolies. Examples include most transportation companies (buses, trains, and airline), PTTs (public telegraph and telephone companies), and gas and electric utility companies.

Let us consider an industry selling a particular service in two time periods, say, during the day (denoted by D), or during the night (denoted by N). We denote by p_D the price of the service sold during the day and by p_N the price of the service sold during the night.

Consumers and seasonal demand

Let us consider a continuum of consumers indexed and uniformly distributed on the closed interval $[a, b]$, where $b > a \geq 0$ and $b > 1$. We denote by δ a particular consumer indexed on $[a, b]$. The utility of consumer δ, $\delta \in [a, b]$, is assumed to be given by

$$U^\delta \equiv \begin{cases} \beta\delta - p_D & \text{if she buys a day service} \\ \beta - p_N & \text{if she buys a night service} \\ 0 & \text{if she does not buy any service} \end{cases} \tag{13.9}$$

where $\beta > 0$ is the reservation utility for a night service.

Recalling Definition 12.1 on page 310, we can use the following definition to provide the terminology for characterizing consumers' attitudes toward purchasing the service in the different periods (seasons).

DEFINITION 13.1 *Day service and night service are said to be*

1. **vertically differentiated** *if, given equal prices ($p_D = p_N$), all consumers choose to purchase only the day service;*
2. **horizontally differentiated** *if, given equal prices ($p_D = p_N$), consumers indexed by a high δ choose to purchase the day service whereas consumers indexed by a low δ choose to purchase the night service.*

Using (13.9), we can see that all day and night services are vertically differentiated if $a \geq 1$, since in this case $\delta\beta \geq \beta$. In contrast, when $0 \leq a < 1$, the two services are horizontally differentiated according to Definition 13.1.

Finally, the consumer indexed by $\hat{\delta}$ denotes the consumer who is indifferent about whether to buy a day service or a night service at the given market prices for these services. Clearly, from (13.9), $\hat{\delta}$ is determined by

$$\hat{\delta} = \frac{\beta + p_D - p_N}{\beta}. \tag{13.10}$$

Thus, given prices, all consumers indexed by $\delta \in [a, \hat{\delta})$ purchase the night service, whereas all the consumers indexed by $\delta \in (\hat{\delta}, b]$ buy the day service.

Production of services

We denote by n_D the number of consumers buying a daytime service and by n_N the number of consumers buying a nighttime service. Clearly, $n_D + n_N \leq b - a$, which is the total number of consumers in the economy.

Production of services requires an investment in capacity and, in addition, bears operation costs. For example, in transportation industries, capacity determines the upper limit on the number of passengers that can be transported in each of the time periods. In the telecommunication industry, capacity determines the upper limit on the number of phone calls (switchboards) that can be simultaneously made in each time period.

Therefore, we denote by K the capacity of a service-producing firm. Then, the number of day or night users cannot exceed this capacity; that is, $n_D \le K$ and $n_N \le K$. We denote by r the cost of a unit capacity facing the firm(s).

In addition to capacity cost (number of aircraft seats, etc.), service-producing firms bear operation costs. Therefore, we denote by c_D the per customer operation cost of producing a day service, and by c_N the per customer operation cost of producing a night service. With no loss of generality we assume that $c_D \ge 0$ and $c_N \ge 0$. That is, the operation-per-customer cost of producing a night service is not higher than the operation-per-customer cost of producing a day service.

Clearly, by varying the relative price of the daytime service and the nighttime service, the monopoly service-producing firm can shift the peak demand from day to night or night to day. For this reason, we refrain from using the terminology *peak* and *off-peak* periods (commonly used in the literature) and confine the terminology to *daytime* or *nighttime* periods. That is, peak and off-peak periods are endogenously determined by the selling firm.

In order to find the profit-maximizing pricing scheme set by the monopoly firm, in what follows we decompose the analysis into a cost analysis and a revenue analysis.

The monopoly's cost structure

Assuming that all consumers are served (either by day or night service), we have it that $n_N = \hat{\delta} - a$, and $n_D = b - \hat{\delta}$. Then, the total cost as a function of the indifferent consumer defined in (13.10), is given by

$$TC(\hat{\delta}) = r \max\left\{\hat{\delta} - a, b - \hat{\delta}\right\} + (\hat{\delta} - a)c_N + (b - \hat{\delta})c_D. \qquad (13.11)$$

Figure 13.6 illustrates the monopoly's production cost as a function of the location of the indifferent consumer. Figure 13.6 shows that the cost is minimized when the market is equally divided between daytime users and nighttime users, that is, $\hat{\delta} = (a+b)/2$, because when the market is equally divided, half of the total population buys a day service and the other half buys a night service, which implies that the amount of capacity needed by the firm is $K = (b-a)/2$, which is at minimum

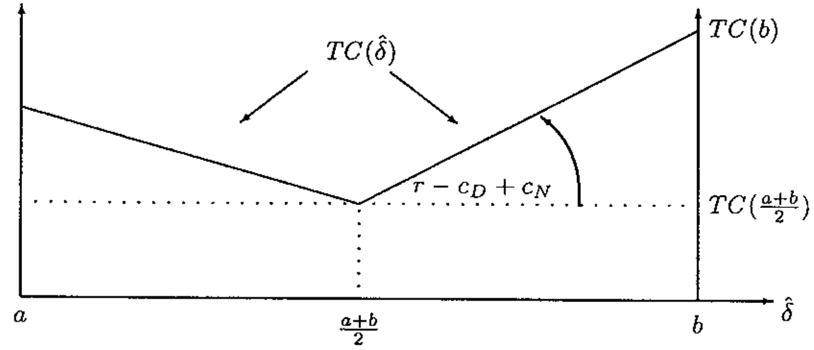

Figure 13.6: Cost structure of a monopoly selling services in two periods. *Notes:* (i) The figure assumes that $r > |c_D - c_N|$. (ii) $TC((a+b)/2) = (b-a)(r + c_D + c_N)/2$.

under this equal division. As $\hat{\delta}$ increases, the amount of capacity must increase to accommodate a larger number of nighttime users. Hence, any deviation from the equal division of consumers, either by increasing the number of night users (an increase in $\hat{\delta}$), or by increasing the number of day users (a decrease in $\hat{\delta}$), will result in an additional investment in building capacity.

If we assume that all consumers are served, an increase in $\hat{\delta}$ means that the monopoly switches consumers from day service to night service. Hence, for each consumer switching from day to night, the monopoly saves an operation cost of $c_D - c_N$. Similarly, for each consumer being switched from night to day service (a decrease in $\hat{\delta}$), the operation cost increases by the difference $c_D - c_N$.

Altogether, in view of (13.11), the marginal cost as a function of the indifferent consumer is given by

$$MC(\hat{\delta}) = \begin{cases} -r + c_N - c_D & \text{if } \hat{\delta} < (a+b)/2 \\ +r + c_N - c_D & \text{if } \hat{\delta} > (a+b)/2. \end{cases} \qquad (13.12)$$

Monopoly's revenue

The monopoly seeks to extract maximum surplus from consumers. Hence, in view of (13.9), the monopoly would charge a price of $p_N = \beta$ for a night service. Then, according to (13.10), determining the price for the day service, p_D is equivalent to determining the location of the indifferent consumer, $\hat{\delta}$. Hence, we can assume that the monopoly's choice variable is $\hat{\delta}$, while p_D is determined according to $p_D = \beta\hat{\delta}$. Consequently, we can define the monopoly's revenue as a function of the location of the indifferent consumer by

$$TR(\hat{\delta}) \equiv p_N n_N + p_D n_D = \beta(\hat{\delta} - a) + \beta\hat{\delta}(b - \hat{\delta}). \tag{13.13}$$

The marginal revenue as a function of the indifferent consumer is given by

$$MR(\hat{\delta}) = \beta(1 + b) - 2\beta\delta. \tag{13.14}$$

Figure 13.7 illustrates the revenue functions for the cases of vertical and horizontal differentiation. The bottom figure shows that under vertical differentiation ($a > 1$), the revenue is maximized when the indifferent consumer locates to the left of the midconsumer. This is because when the products are vertically differentiated, all consumers prefer day over night services, and given that they are willing to pay more for a daytime service, the monopoly will choose prices so that the majority of the consumers will be daytime users.

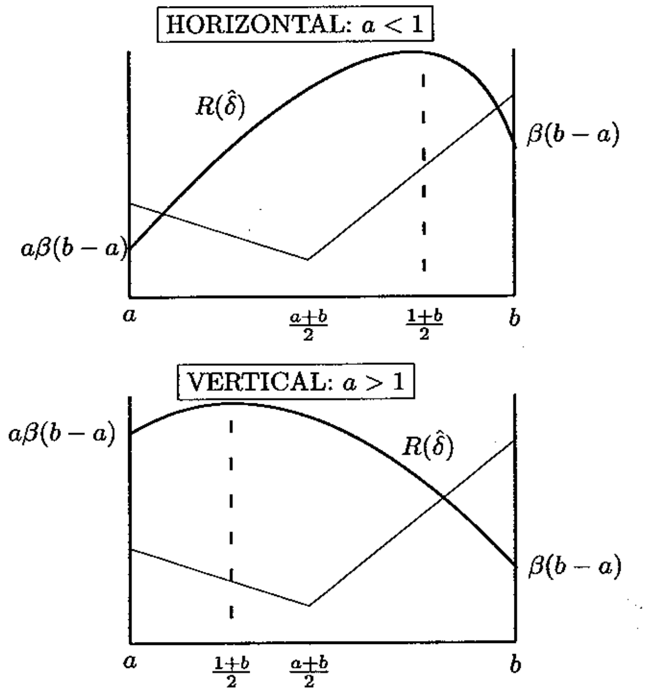

Figure 13.7: Revenue functions for the vertical and horizontal differentiation cases

The top figure shows that when the products are horizontally differentiated ($a < 1$), revenue is maximized when the indifferent consumer locates to the right of the midconsumer. The last case, when $a = 1$, is not illustrated, but in this case the revenue is maximized when the indifferent consumer locates exactly at the midpoint, implying that the monopoly allocates half of the consumers to day services and half to night services.

Monopoly's profit-maximizing pricing structure

Before we proceed with the calculations of the profit-maximizing pricing structure, let us note that the monopoly's profit is measured by the distance between the revenue and the cost functions in Figure 13.7. Figure 13.7 (bottom) reveals that under vertical differentiation the monopoly will never choose to price the service, so that the indifferent consumer would locate to the right of the midconsumer. Figure 13.7 (top) reveals that under horizontal differentiation the monopoly will never choose to price the service, so that the indifferent consumer would locate to the left of the midconsumer.

DEFINITION 13.2 *The daytime period is called a* **peak period** *if* $\hat{\delta} < (a + b)/2$, *and* **off-peak** *otherwise. Similarly, the nighttime period is called a peak period if* $\hat{\delta} > (a + b)/2$, *and off-peak otherwise.*

Hence, Figure 13.7 and Definition 13.2 imply that

Proposition 13.6 *If the two time-period services are vertically differentiated, then the monopoly will turn the daytime period into the peak period. If the two time-period services are horizontally differentiated, then the monopoly will turn the nighttime period into the peak period.*

We therefore can state the main proposition concerning monopoly behavior:

Proposition 13.7 *Given that* $r > |c_D - c_N|$, *a monopoly that maximizes profit will set prices so that services are purchased in both periods such that*

1. *under vertical differentiation*

$$\hat{\delta} = \min\left\{\frac{\beta(1+b) + r + c_D - c_N}{2\beta}; \frac{a+b}{2}\right\}, \text{ and}$$

2. *under horizontal differentiation*

$$\hat{\delta} = \max\left\{\frac{\beta(1+b) - r + c_D - c_N}{2\beta}; \frac{a+b}{2}\right\}.$$

Proof. The monopoly seeks to choose $\hat{\delta}$ to maximize $TR(\hat{\delta}) - TC(\hat{\delta})$. By Proposition 13.6, under vertical differentiation $\hat{\delta} < (a+b)/2$. Hence, (13.12) and (13.14) imply that

$$\beta(1+b) - 2\beta\hat{\delta} = -r + c_N - c_D.$$

Under horizontal differentiation $\hat{\delta} > (a+b)/2$. Hence, (13.12) and (13.14) imply that

$$\beta(1+b) - 2\beta\hat{\delta} = r + c_N - c_D.$$

■

13.5 Exercises

1. Congratulations! You have been appointed to be the chairperson of the Economics department at Wonderland University. Since that old photocopy machine broke down three years ago, the department has been deprived of copying services, and therefore, your first task as a chairperson is to rent copying services from KosKin Xeroxing Services, Inc.. The KosKin company offers you two types of contracts: The Department can simply pay 5 cents per page, or, the department can pay a yearly fee of $300 and in addition pay 2 cents per page.
   (a) Draw the department's total photocopying expenses as a function of the number of copies made each year under the two types of contracts.
   (b) Conclude which contract is less costly, given the number of copies made each year.
2. SouthNorthern Airlines is the sole provider of flights between City A and City B. During the winter, the inverse demand for flights on this route is given by $p_W = 10 - q_W$, where p_W is the airfare charged during the winter and q_W is the number of passengers flown on this route during the winter. Similarly, during the summer the inverse demand function is given by $p_S = 5 - q_S/2$. Denote by K the airline's capacity, defined by the number of airplane seats SouthNorthern intends to acquire, and assume that the average cost of an airplane seat is $r > 0$. Also, suppose that the cost of flying each passenger is $c > 0$.
   (a) Calculate the number of passengers flown in each season and SouthNorthern's profit level, assuming that $r = c = 1$.
   (b) Calculate the number of passengers flown in each season and SouthNorthern's profit level, assuming that $r = 3$ and $c = 1$.

13.6 References

Bailey, E., and L. White. 1974. "Reversals in Peak and Off-Peak Prices." *Bell Journal of Economics* 5: 75–92.

Bergstrom, T., and J. MacKie-Mason. 1991. "Some Simple Analytics of Peak-Load Pricing." *Rand Journal of Economics* 22: 241–249.

Brown, S., and D. Sibley. 1986. *The Theory of Public Utility Pricing.* Cambridge: Cambridge University Press.

Joskow, P. 1976. "Contributions to the Theory of Marginal Cost Pricing." *Bell Journal of Economics* 7: 197–206.

Oi, W. 1971. "A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly." *Quarterly Journal of Economics* 85: 77–96.

Sherman, R. 1989. *The Regulation of Monopoly.* Cambridge: Cambridge University Press.

Steiner, P. 1957. "Peak-Loads and Efficient Pricing." *Quarterly Journal of Economics* 585–610.

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