However, one does not observe y_{it}^* directly. All one actually observes is whether consumer i at time t buys *Yoplait 150* ($Y_{it} = 1$) or does not ($Y_{it} = 0$). The standard assumption in this case then is that we observe $Y_{it} = 1$, when $y_{it}^* \geq 0$, and $Y_{it} = 0$ when $y_{it}^* < 0$. This implies that the probability of observing a purchase $Y_{it} = 1$ is given by

$$\begin{aligned}\text{Prob}(Y_{it} = 1) &= \text{Prob}\left[\sum_{j=1}^{k} \beta_j x_{jit} + \varepsilon_{it} \geq 0\right] \\ &= \text{Prob}\left[\varepsilon_{it} > -\left(\sum_{j=1}^{k} \beta_j x_{jit}\right)\right] = 1 - F\left[-\left(\sum_{j=1}^{k} \beta_j x_{jit}\right)\right] \end{aligned} \tag{19.19}$$

where $F()$ is the cumulative distribution of ε_{it}. It is convenient to assume that $F()$ has a symmetric distribution so that $1-F(-Z_{it}) = F(Z_{it})$. Then we have

$$\text{Prob}(Y_{it} = 1) = F\left(\sum_{j=1}^{k} \beta_j x_{jit}\right) \tag{19.20}$$

Clearly, much depends on the choice of the distribution of the random term ε_{it}. If ε_{it} is assumed to be distributed normally[19] one gets the Probit estimation procedure. A popular alternative is to assume instead that ε_{it} has a logistic cumulative distribution in which case:

$$F(Z_{it}) = \frac{e^{Z_{it}}}{1 + e^{Z_{it}}} \tag{19.21}$$

The reason for the popularity of this distribution is that this transformation implies that: $\ln\left[\frac{F(Z_{it})}{1 - F(Z_{it})}\right] = Z_{it}$

In other words,

$$\ln\left[\frac{F\left(\sum_{j=1}^{k} \beta_j x_{jit}\right)}{1 - F\left(\sum_{j=1}^{k} \beta_j x_{jit}\right)}\right] = \ln \frac{\text{Prob}(Y_{it} = 1)}{\text{Prob}(Y_{it} = 0)} = \sum_{j=1}^{k} \beta_j x_{jit} \tag{19.22}$$

The ratio of the probability $Y_{it} = 1$ to the probability that $Y_{it} = 0$ is known as the odds ratio. By assuming a logistic distribution for ε_{it}, the logit estimation procedure assumes that the log of the odds ratio is a linear function of the key exogenous variables. This is a very convenient feature for estimation purposes.

Ackerberg (2001) presents a number of regressions based on the above logit procedure. The independent variables X_{it} include: 1) the amount (in time) of *Yoplait 150* advertising the household has seen up to that time divided by the total time spent watching television,

19 This assumption was made in the empirical applications in Chapters 13 and 19.

ADS; 2) the price of *Yoplait 150* in the relevant market at that time, OWN PRICE; 3) a comparable measure of the average competitor's price, RIVAL PRICE; 4) the number of times (possibly zero) the household had purchased *Yoplait 150* previous to that time, NUMBER PREV; and 5) the key 1,0 variable indicating whether the household had any previous purchases of *Yoplait 150*, EXPERIENCED or INEXPERIENCED.[20] Some of his main results are summarized in Table 19.3 below.

Consider the first regression results. Advertising has an important impact, but only for those who have not yet tried the new product. Again, this implies that advertising mostly plays an informative role. Specifically, the coefficient on the interactive term, ADS*EXPERIENCED captures the impact of advertising on consumers who know the quality of *Yoplait 150* and therefore should reflect only complementary prestige or recognition effects. This coefficient is not statistically different from zero. In contrast, the coefficient on ADS*INEXPERIENCED reflects both prestige and information effects. It *is* statistically different from zero and this suggests that the information effect is behind this because our estimate of prestige effects is not distinct from zero.[21]

The second regression tries to discriminate more between the two types of information that advertising provides. In the first regression, the assumption is that a household becomes fully informed after just one purchase of *Yoplait 150*. This would likely be the case if the important information provided by advertising were simply knowledge of the good's existence and availability. Once a household has bought the product, it presumably knows these features of the product. Learning brand characteristics such as taste, calories, and so on may take a little longer and may be facilitated by continuing advertisements. For this reason, the regression includes ADS alone as an independent regressor, but then also includes this variable in an interaction term with NUMBER PREV, the number of prior purchase of the *Yoplait 150*. The idea is that the pure effect of advertising measured by ADS will decline as the consumer's experience grows. The more rapidly this decline occurs, the more likely it

Table 19.3 Effect of advertising in the low-fat yogurt market dependent variable: purchase (or not) of *yoplait 150* by household i at time t

Independent Variable	*Coefficient*	*Std. Error*	*Coefficient*	*Std. Error*
ADS*INEXPERIENCED	2.306	(0.776)*	—	—
ADS*EXPERIENCED	0.433	(1.212)	—	—
ADS	—	—	2.014	(0.790)*
ADS*(NUMBER PREV)	—	—	–0.356	(0.108)*
NUMBER PREV	–0.267	(0.093)*	–0.270	(0.092)*
(NUMBER PREV)^2	0.009	(0.001)*	–0.001	(0.001)
OWN PRICE	–5.584	(0.350)*	–5.616	(0.356)*
RIVAL PRICE	0.761	(0.217)*	0.768	(0.219)*

*Indicates significant at the five percent level.

[20] Household size and income and, as before, a market dummy for Springfield households were also included. Ackerberg (2001) also includes a random, household-specific intercept to control for household heterogeneity in time-persistent preferences for the product.

[21] To be precise, the difference between the two coefficients, ADS*INEXPERIENCED and ADS*EXPERIENCED is a direct estimate of the pure information effects. Standard techniques yield a t-statistic for this difference of about 1.5.

is that the primary information obtained from advertising is existence and availability. The more slowly it declines, the more likely that the information provided concerns product attributes that take time to learn. Sure enough, the coefficient on ADS*(NUMBER PREV) is negative but a relatively small –0.36. This implies that it takes six or seven purchases of *Yoplait 150* before the advertising information is no longer useful. As noted, this implies that part of the information advertising provides concerns product attributes.

Are these coefficient estimates sensible? It is difficult to say immediately because the coefficients in the logit model relate to the effect of advertising on the *probability* of purchase and not directly to demand. However, there are some aspects of the results that give us confidence in the findings. First, in each case, the price of *Yoplait 150* had a strong negative impact and the rival's price a strong positive effect on a household's purchase decision. Second, one can simulate the model to see what overall demand features the price and advertising coefficients imply. When Ackerberg (2001) conducts such simulations with the full model he finds that, taken at the mean, the own-price elasticity of demand is 2.8—a fairly elastic response. He also finds that the elasticity of demand with respect to advertising is 0.15. Taken together, the advertising and price elasticities would imply, by virtue of the Dorfman-Steiner condition, an advertising-to-sales ratio of $0.15/2.8 = 0.054$ or 5.4 percent. This is a quite reasonable result given that Yoplait's overall advertising-to-sales ratio was reported at the time to be about 7 percent. Overall, then, Ackerberg's (2001) findings seem to be quite plausible.

In short, the evidence from Ackerberg (2001) is that the primary role of advertising is to provide consumers with information. Some of this information is simply making consumers aware of the product's availability, but some of it concerns educating consumers about the product's key features. There is little evidence that in this particular market advertising provides prestige or complementary effects. The data are based, however, on a perishable consumer food product purchased with some frequency. Whether it applies to other more durable consumer goods, or to goods such as medications that consumers buy less frequently, merits further investigation.

Summary

Advertising plays a role in informing consumers of the availability of a product, its brand image, and sometimes product attributes. This role can be played even when the actual information content of the advertising message is low. When consumers are uncertain about product quality, the very fact that a product is advertised heavily may convey information. When advertising informs consumers of the availability of substitute products, it also tends to increase price competition. In competing for customers, advertising in equilibrium may be socially wasteful or excessive if each firm spends substantial but mutually offsetting amounts that, on net, leaves customer patronage unchanged.

However, we should not infer that advertising is socially excessive from the fact that consumers often receive advertising at zero cost. Advertising may be viewed as a complement to the product advertised. As with any complement, an increase in the supply of advertising raises the demand for the promoted product. In this view, what really matters is the total price that consumers pay for the product and the advertising together. It is important, however, to recognize as well the competition-enhancing effects of advertising in assessing its welfare effects.

There is considerable empirical evidence that advertising does in fact provide information to consumers—especially for search goods. Nevertheless, we should not necessarily take this to imply that advertising will foster less concentrated industries. We should in fact expect more advertising by firms with market power and also by firms with low costs. In fact, the more intense competition that advertising may induce can serve to deter entry and to enable more efficient firms to grow large so that industrial concentration rises.

Problems

1. Suppose that the demand for a new wrinkle cream is described by a nonlinear demand function $Q(P, A) = P^{-1/2}A^{1/4}$. Show that the price elasticity of demand is $\eta\mathrm{P} = 1/2$ and that the advertising elasticity of demand is $\eta\mathrm{A} = 1/4$. What do you predict the advertising-to-sales ratio would be in this industry? Does it depend on how costly it is to advertise for this product?

2. A firm has developed a new product for which it has a registered trademark. The firm's market research department has estimated that the demand for this product is $Q\ (P, A) = 11,600–1,000P + 20A^{1/2}$ where Q is annual output, P is the price, and A the annual expenditure for advertising. The total cost of producing the new good is $C(Q) = .001Q^2 + 4Q$. The unit cost of advertising is constant at $m = 1$.
   a. Calculate the optimal output level Q^*, price P^*, and advertising level A^* for the firm.
   b. What is firm profit if it follows this optimal strategy?
   c. What is consumer surplus if the firm adopts this strategy?

3. Imagine that there are 1,000 consumers. For each consumer, the willingness to pay for a widget is distributed uniformly over the interval [0,1] depending on the style of the widget. A retailer with a particular style of the good knows this distribution. Its costs are zero. Consumers do not know the style that the retailer has stocked and each incur a transport or search cost of $T = 0.125$. Once this cost is incurred it is sunk. At that point, a consumer in the retailer's store will purchase the product so long as the consumer's valuation is greater than or equal to the price charged by the retailer.
   a. Show that facing a random selection of customers, the retailer's profit maximizing price is $p = 0.5$.
   b. Show that with $T = 0.125$, all consumers will come to shop expecting a price of 0.5. What would happen if $T = 0.15$?

4. Suppose that the retailer in question 3 could communicate in some way to those customers with valuations less than 0.5 of the style that it has in stock and tell them that it is not worth coming. If the retailer keeps the price at 0.5, how large can the transport cost T now be before the market collapses? Will the retailer keep the price at $p = 0.5$?

5. Let there be two firms, 1 and 2. Each firm sells a product of with material quality $Z = 1$ and each chooses its price, p_1 and p_2, respectively. However, firm 1 also gets to choose an advertising level a_1. Consumers perceive the overall product quality to be the product's advertising level times its material quality. In other words, consumers perceive product 1 to be of quality a_1 and product 2 to be of quality 1. Consumers are indexed by v distributed continuously from zero to 1, where v_i is consumer i's willingness to pay for quality. Consumer i's net gain from consuming product 1 is $v_i a_1 - p_1$, while consuming product 2 generates a net gain of $v_i - p_2$. There is no production cost; however, firm 1 incurs advertising cost of $(a_1/2)^2$.
   a. Assume all N consumers always buy the product of either firm 1 or firm 2, i.e., the market is always covered. Derive the condition for the marginal consumer v^m satisfies and the demand facing each firm.
   b. Derive the equilibrium values of p_1, p_2, and a_1.
   c. Suppose firm 2 is permitted now to advertise at any positive level a_2 between 0 and 0.5. What level of advertising will it choose if it takes firm 1's choice a_1 as given?

6. You have been hired to market a new music recording that is expected to have target sales of $20 million for the coming year. The marketing department has estimated that a 1 percent increase in advertising the recording would increase the recordings sold by about 0.5 percent, and that a 1 percent increase in the price of the recording would reduce the number sold by about 2 percent. How much money should you commit to advertising the recording in the coming year?
7. How could you explain the different advertising-to-sales ratios of the following firms:

Firm	*Main Products*	*Advertising/Sales*
Philip Morris	Tobacco, food, beer	7
Procter & Gamble	Soaps, paper, food	5.3
General Motors	Autos	3.5
Kodak	Photo supplies	9
Johnson & Johnson	Pharmaceuticals	11
Pepsico	Soft drinks, snacks	5.2
Sears, Roebuck	Retailing	3.4
American Home Products	Pharmaceuticals	17.3

References

Anderson, S., and R. Renault. 2006. "Advertising Content," *American Economic Review* 96 (March): 93–113.

Ackerberg, D. 2001. "Empirically Distinguishing Informative and Prestige Effects of Advertising," *Rand Journal of Economics* 32 (Summer): 316–33.

Archibald, R., C. A. Haulman, and C. E. Moody. 1983. "Quality, Price, Advertising, Published Quality Ratings," *Journal of Consumer Research* 9 (March): 347–53.

Bagwell, K., and M. Riordan. 1991. "High and Declining Prices Signal Product Quality," *American Economic Review* 81 (March): 224–39.

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

Bain, J. S. 1956. *Barriers to New Competition: Their Character and Consequences in Manufacturing Industries*, Cambridge, MA: Harvard University Press.

Becker, G., and K. Murphy. 1993. "A Simple Theory of Advertising as a Good or Bad," *Quarterly Journal of Economics* 108 (August): 941–64.

Benham, L. 1972. "The Effect of Advertising on the Price of Eyeglasses." *Journal of Political Economy*, 15 (October), 337–52.

Butters, G. 1977. "Equilibrium Distribution of Sales and Advertising Prices," *Review of Economic Studies*, 44 (June): 465–91.

Cady, J. F. 1976. "An Estimate of the Price Effects of Restrictions on Drug Price Advertising," *Economic Inquiry* 14 (July): 493–510.

Caves, R. E., and D. P. Green. 1996. "Brands' Quality Levels, Prices, and Advertising Outlays: Empirical Evidence on Signals and Information Costs," *International Journal of Industrial Organization* 14 (1996): 29–52.

Clark, C., and I. Horstmann, 2005. "Advertising and Coordination in Markets with Consumption Scale Effects," *Journal of Economics and Management Strategy* 14 (June): 377–401.

Comanor, W. S., and T. A. Wilson. 1967. "Advertising Market Structure and Performance," *Review of Economics and Statistics* 49 (November): 423–40.

Dinlersoz, E., and M. Yorukoglu. 2012. "Information and Industry Dynamics." *American Economic Review*, 102 (April), 884–913.

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

Dorfman, R., and P. O. Steiner, 1954. "Optimal Advertising and Optimal Quality," *American Economic Review* 44 (December): 826–36.

Ellison, G., and S. Fisher Ellison. 2005. "Search, Obfuscation, and Price Elasticities on the Internet," Working Paper, MIT Department of Economics.

Fisher, F., and J. J. McGowan. 1979. "Advertising and Welfare: Comment," *Bell Journal of Economics* 10:726–7

Fluet, C., and P. Garella. 2001. "Advertising and Prices as Signals of Quality in a Regime of Price Rivalry." *International Journal of Industrial Organization* 20 (September): 907–30.

Galbraith, J. K. 1958. *The Affluent Society*, Boston: Houghton-Mifflin.

Geroski, P. 1982. "Simultaneous Equations Models of the Structure-Performance Paradigm," *European Economic Review* 19 (September): 145–58.

Glazer, A. 1981. "Advertising, Information, and Prices: A Case Study," *Economic Inquiry* 19 (October): 661–71.

Grossman, G. M., and C. Shapiro. 1984. "Informative Advertising with Differentiated Products," *Review of Economic Studies* 51 (February): 63–81.

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

Kihlstrom, R., and M. Riordan. 1984. "Advertising as a Signal," *Journal of Political Economy* 92 (June): 427–50.

Kotowitz, Y., and G. F. Mathewson. 1986. "Advertising and Consumer Learning," in P. M. Ippolito and D. T. Schefman, eds., *Empirical Approaches to Consumer Protection Economics,* Federal Trade Commission. Washington, D.C.: U.S. Government Printing Office.

Lambin, J. J. 1976. *Advertising, Competition, and Market Conduct in Oligopoly over Time*. Amsterdam: North-Holland.

Marshall, A. 1890. *Principles of Economics*, London: MacMillan and Co.

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

Milyo, J., and J. Waldfogel. 1995. "The Effect of Price Advertising on Prices: Evidence in the Wake of 44 Liquormart," *American Economic Review* 89 (December): 1081–96.

Nelson, P. 1970. "Information and Consumer Behavior," *Journal of Political Economy* 78 (May): 311–29.

Nevo, A. 2001. Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica*, 69 (March), 307–42.

Nichols, W. H. 1951. *Price Policy in the Cigarette Industry*, Nashville: Vanderbilt University Press.

Pastine, I., and T. Pastine. 2002. "Consumption Externalities, Coordination, and Advertising," *International Economic Review* 43 (August): 919–43.

Pinkse, J., and M. Slade. 2007. "Semi-Structural Models of Advertising Competition," *Journal of Applied Econometrics* 22 (December): 1227–46.

Pope, D. 1983. *The Making of Modern Advertising*, New York: Basic Books.

Reisz, P. 1978. "Price versus Quality in the Marketplace," *Journal of Retailing* 54 (Winter): 15–28.

Round, D. K. 1983. "Intertemporal Profit Margin Variability and Market Structure in Australian Manufacturing," *International Journal of Industrial Organization* 1 (June): 189–209.

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

Schwartz, A., and L.L Wilde. 1985. "Product Quality and Imperfect Information," *Review of Economics Studies* 52: 251–262.

Slade, M. 1995. "Product Rivalry with Multiple Strategic Weapons: An Analysis of Price an Advertising Competition," *Journal of Economics and Management Strategy* 4 (September): 445–76.

Solow, R. M. 1967. "The New Industrial State or Son of Affluence," *Public Interest* 9 (Fall): 100–108.

Stigler, G. 1968. "Price and Non-Price Competition," *Journal of Political Economy* 76 (February): 149–54.

Stigler, G., and G. Becker. 1977. "De Gustibus Non Est Disputandum," *American Economic Review* 67 (March): 76–90.

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

Telser, L. 1964. "Advertising and Competition," *Journal of Political Economy* 72 (December): 537–62.

Appendix

We present here first a model of excessive advertising competition and, second, a simplified version of the advertising information and differentiated product competition model developed by Grossman and Shapiro (1984).

WASTEFUL COMPETITION

Assume a duopoly with each firm's profit given by:

$$\pi_1 = Z(1 + A_1 - bA_1A_2 - 0.5A_1^2)$$
$$\pi_2 = Z(1 + A_2 - bA_1A_2 - 0.5A_2^2) \tag{19.A1}$$

where A_i is the level of advertising effort of firm i measured as the advertising-to-sales ratio, and Z and b are positive parameter a with $0 < b < 1$.

The first-order conditions then yield the following best response functions:

$$A_i = 1 - bA_j \quad i, j = 1, 2 \text{ and } i \neq j \tag{19.A2}$$

Hence in equilibrium:

$$A_1 = A_2 = \frac{1}{1+b} \tag{19.A3}$$

and

$$\pi_1 = \pi_2 = Z\left[1 + \frac{0.5}{(1+b)^2}\right] \tag{19.A4}$$

The cooperative advertising choices that maximize the sum, $\pi_1 + \pi_2$, are:

$$A_1^C = A_2^C = \frac{1}{1+2b} \tag{19.A5}$$

This yields a cooperative π_i^C profit to each firm of

$$\pi_1^C = \pi_2^C = Z\left[1 + \frac{0.5+b}{(1+2b)^2}\right] > \pi_1 = \pi_2 = Z\left[1 + \frac{0.5}{(1+b)^2}\right] \text{for} \quad 0 < b < 1 \tag{19.A6}$$

INFORMATIVE ADVERTISING AND PRICE COMPETITION

We assume a set of N consumers continuously distributed along a line segment of unit length. Firm 1 is at the left end of the segment and firm 2 is at the right end. Each consumer buys at most one unit of the good and receives surplus net of travel cost x_i equal to

$$U_i = V - p_1 - tx_i \text{ if the consumer buys from firm 1; and}$$
$$U_i = V - p_2 - t(1 - x_i) \text{ if the consumer buys from firm 2} \tag{19.A7}$$

Each firm i chooses a level of advertising aimed at informing a fraction θ_i of the N consumers. Hence, $\theta_1(1–\theta_2)N$ know only of brand 1; $\theta_2(1–\theta_1)$ know only of brand 2; $\theta_1\theta_2 N$ know of both goods; and $(1–\theta_1)(1–\theta_2)$, know of either good. We assume that both θ_1 and $\theta_2 < 1$.

Among the $\theta_1\theta_2 N$ consumers informed about both products the marginal consumer indifferent to either good has address $x^m(p_1, p_2) = \dfrac{(p_2 - p_1 + t)}{2t}$. Demand for firm 1 is made up of two parts. The first part is the $\theta_1(1–\theta_2)N$ consumers who know only of firm 1's brand. The second part comes from the $\theta_1\theta_2 N$ consumers who know of both brands. Hence, the total demand for firm 1 is:

$$q_1(\theta_1, \theta_2 p_1, p_2) = \theta_1(1 - \theta_2)N + \theta_1\theta_2 x^m N = \left[\theta_1\left(1 - \theta_2\right) + \theta_1\theta_2 \frac{(p_2 + t - p_1)}{2t}\right] N \tag{19.A8}$$

We assume a quadratic advertising cost function:

$$A(\theta_1)N = \frac{1}{2}\alpha\theta^2 N \tag{19.A9}$$

Firm 1's profit function is therefore:

$$\pi(\theta_1, \theta_2, p, p_2) = (p_1 - c)\left[\theta_1\left(1 - \theta_2\right) + \theta_1\theta_2 \frac{(p_2 + t - P_1)}{2t}\right] N - \frac{1}{2}\alpha\theta_1^2 N \tag{19.A10}$$

The two first-order conditions necessary for profit maximizaiton are:

$$\frac{\theta_1\theta_2(p_2 + - p_1)N}{2t} + \theta_1(1 - \theta_2)N - \frac{\theta_1\theta_2(p_1 - c)N}{2t} = 0$$
$$(p_1 - c)\left[\left(1 - \theta_2\right) + \theta_2\left(\frac{p_2 + t - p_1}{2t}\right)\right] N - \alpha\theta_1 N = 0 \tag{19.A11}$$

By symmetry, $p_1 = p_2 = p^*$ and $\theta_1 = \theta_2 = \theta^*$. Hence, the first of the two-first order conditions is:

$$p^* - c = -t + \frac{2t}{\theta^*} \tag{19.A12}$$

Similarly, the second first-order condition now is:

$$(p^* = c)\left(1 - \frac{\theta^*}{2}\right) = \alpha\theta* \tag{19.A13}$$

These two conditions may then be jointly solved to yield:

$$p^* = c + \sqrt{2\alpha t} \tag{19.A14}$$

and

$$\theta^* = \frac{2}{1 + \sqrt{\dfrac{2\alpha}{t}}} \tag{19.A15}$$

Because we assumed that some consumers remain uninformed, $\theta^* < 1$, we assume that that $\alpha > t/2$. Equation (19.A14) then implies an equilibrium price $p^* > c + t$. Each firm's equilibrium profit π^* is:

$$\pi^* = \frac{2\alpha}{\left(1 + \sqrt{2\alpha/t}\right)^2} \tag{19.A15}$$

20

Research and Development

The final results of the human genome project indicate that we humans are not as complicated as we thought we were. Rather than consisting of the approximately 100,000 genes that were initially predicted, it appears that we have only 30,000 genes, more but less than twice as many as the humble roundworm with its 19,098 genes.[1] This finding is important for many reasons. From our perspective, there is an important economic aspect to this result. It is well understood that genes are a crucial factor in predicting and curing many diseases. Therefore, identifying and understanding the workings of each gene could lead to the creation of a new family of custom-made drugs. The rough equation quoted by the pharmaceutical companies was "one gene, one patent, one drug."[2] If, as initially expected, there were 100,000 genes then there was potentially a vast number of revenue-generating patents. The finding that the actual number of genes is far less than 100,000 has suggested to many that genes hold many fewer of the keys to the treatment of disease. As a result, understanding genes and their functions may offer a much less lucrative source of new patentable treatments.

However, all is not lost. It is being suggested that much of human biology is determined at the protein level rather than at the DNA level, and we have well over 1,000,000 different proteins in our bodies. So now we have a whole new science, proteomics—studying how genes control proteins—as a method for creating tailored drugs. Proteomics is being pursued by an increasingly wide number of companies and institutions: Harvard University, for example, has created a new Institute of Proteomics.

The race to understand the proteomic causes of diseases and to develop new drugs targeted at those diseases will not come as a surprise to anyone familiar with the popular business literature of the past twenty years. That literature is characterized by the dominant theme that the most successful firms find new ways of doing things, or develop new products and new markets.[3] The now prevalent view is that firms become industry leaders

[1] If you are interested, the complete human genome is available as a free download from http://gdbwww.gdb.org/.

[2] "Scientists, Companies Look to the Next Step After Genes," *The New York Times* 13 February 2001.

[3] This is virtually the mantra in the best-selling book by Peters and Waterman, 1982. *In Search of Excellence: Lessons from America's Best Run Companies*. However, the argument is repeated frequently in other business books, including, as noted herein, Porter's (1990) encyclopedic volume.

by conducting research and development (R&D), leading to innovations in their production technologies or the products they provide. Michael Porter's *The Competitive Advantage of Nations* (1990) serves to make the point. Porter writes that any theory of competitive success

> must start from the premise that competition is dynamic and evolving .... Competition is a constantly changing landscape in which new products, new ways of marketing, new production processes, and whole new market segments emerge .... [Economic] theory must make improvement and innovation in methods and technology a central element. (Porter 1990, 20).

Porter's quote could almost have been taken verbatim from Joseph Schumpeter's classic work written almost fifty years earlier. Schumpeter was both an economist and a historian. He brought a historical perspective to his study of competition and the rise and fall of corporate empires. The following dramatic passage appears in his book *Capitalism, Socialism, and Democracy* first published in 1942.

> [I]t is not ... [price] competition which counts but competition from the new commodity, the new technology, the new source of supply, the new type of organization ... competition which commands a decisive cost or quality advantage and which strikes not at the margins of the profits and outputs of existing firms but at their foundations and very lives. (Schumpeter 1942, 84).

Interest in the forces behind innovative activity is perhaps stronger today than it was when Schumpeter wrote.[4] An important issue, raised by Schumpeter, concerns the market environment most conducive to R&D activity. Schumpeter conjectured that R&D efforts are more likely to be undertaken by large firms than by small ones. He speculated secondly that monopolistic or oligopolistic firms would more aggressively pursue innovative activity than would firms with little or no market power. Accordingly, Schumpeter argued that the benefits of an economy made up largely of competitive markets populated by small firms reflected the rather modest gains of allocating resources efficiently among a *given set of goods and services produced with given technologies*. In contrast, the benefits of markets dominated by large firms, each with sizable market power, stems from the much larger dynamic efficiency gains of developing new products and new technologies. As Schumpeter wrote, "a shocking suspicion dawns upon us that big business may have had more to do with creating (our) standard of life than with keeping it down" (p. 88).

The validity of Schumpeter's ideas—which have come to be jointly referred to as the Schumpeterian hypothesis—is the key issue addressed in this chapter. Do larger firms do more R&D? Does a concentrated market structure provide a better environment for the development of new innovations than a competitive structure?

Table 20.1 lists the ten companies awarded the most patents by the US Patents and Trademark Office (USPTO) in 2011 as well as their ranks in 2010 and 2009. Each of these is a large company. Most operate in oligopolistic markets with only a few large competitors. Moreover, there is considerable stability in the rank ordering, at least over these three years. It is tempting to conclude on the basis of such data that Schumpeter was right and that large firms in concentrated markets are more innovative. However, great care is needed before

[4] For example, see the story by C. J. Whalen, "Today's Hottest Economist Died 50 Years Ago," *Business Week* 11 December 2000. Ever since Solow's (1956) classic work, macroeconomists studying growth have also focused intensively on technological progress and innovation as the primary source of improved living standards over time. See, e.g., the books by Barro and Sala-I-Martin (1995) and Romer (2006).

Table 20.1 Top ten patent-receiving firms in 2011 and rank in 2010 and 2009

Company	*# of patents 2011*	*Rank in 2010*	*Rank in 2009*
International Business Machines	6,148	1	1
Samsung Electronics	4,868	2	2
Canon Kabushiki Kaisha	2,818	4	4
Panasonic Corporation	2,533	5	5
Toshiba Corporation	2,451	6	6
Microsoft Corporation	2,309	3	3
Sony Corporation	2,265	7	7
Seiko Epson Corporation	1,525	12	9
Hitachi	1,455	11	13
General Electric	1,444	14	15

Source: USPTO

reaching that conclusion. Rather than implying that large firms do more R&D, these results could imply that firms that do more R&D become large.

The most active areas for research activity are likely to vary over time. Table 20.2 lists the top patent-receiving industries or research areas in 2010 and 2011. It also shows the cumulative patents in that area up to that year. While there is some consistency across the three columns there is also considerable variation. Thus, while multiplex communications have led the patent parade in recent years, bio-science drugs, semiconductor devices, solid state devices, and molecular chemistry have accounted for many more patents in total over time.

Introducing a new product can often undermine the marketability of the firm's existing products. Similarly, the development of a new production process requiring new equipment reduces the value of existing productive capacity. Because introducing new products

Table 20.2 Top patent receiving sectors in 2011 and cumulative patents to that year

Industry Class	*2010*	*2011*	*Total*
Multiplex Communications	7415	7720	57275
Active Solid-State Devices (e.g., Transistors, Solid-State Diodes)	6901	7028	74029
Semiconductor Device Manufacturing: Process	6143	5909	81088
Telecommunications	4311	5578	44707
Electrical Computers and Digital Processing Systems: Multicomputer Data Transferring	4647	4877	31663
Drug, Bio-Affecting, and Body Treating Compositions	4712	4597	94335
Data Processing: Database and File Management or Data Structures	4452	4245	27382
Data Processing: Financial, Business Practice, Management, or Cost/Price Determination	4056	4062	20460
Image Analysis	3370	3744	33340
Chemistry: Molecular Biology and Microbiology	3711	3733	64819
Computer Graphics Processing and Selective Visual Display Systems	3321	3525	39756
Pulse or Digital Communications	3024	3073	34123

Source: USPTO

Reality Checkpoint

Creative Destruction in the Pharmaceutical Industry: Will the Prozac Work if the Viagra Fails?

The pharmaceutical industry offers perhaps the best examples of Schumpeter's "creative destruction." Consider the market for antidepressants. For several years after its introduction in 1987 by Eli Lilly & Co., *Prozac* dominated this market. Originally envisioned as a treatment for high blood pressure and, when that failed, an anti-obesity drug, Lilly was pleasantly surprised when hospital tests on mildly depressed patients showed a marked and widespread positive effect. Lilly took its fluoxetine drug as it was then called, and asked Interbrand, one of the major branding companies in the world, to develop a new name and sales campaign. *Prozac* was born and soon it dominated the antidepressant market. By the early 1990s, *Prozac* accounted for nearly a quarter of Lilly's $10 billion revenue.

Competition came quickly. In 1992 Pfizer, Inc., introduced *Zoloft*, which quickly jumped to a third of the market. SmithKline Beecham's *Paxil* followed shortly and soon had 20 percent. All three drugs increased levels of the neurotransmitter, serotonin, but all three had slightly different chemical bases and different side effects. When the *Prozac* patent expired in 2001, prescriptions for generic fluoxetine grew rapidly. Within a year, Lilly had lost 90 percent of its *Prozac* prescriptions. Lilly countered with a new drug, *Cymbalta* that works on two neurotransmitters, serotonin and norepinephrine.

Originally known as sildenafil citrate, *Viagra*, was first envisioned as a treatment for high blood pressure and angina. Even though it failed in that regard, its many male users reported a dramatic increase in sexual function. Pfizer received approval from the Food and Drug Administration (FDA) to market the drug as a treatment for erectile dysfunction (ED) and re-launched the drug under the name, *Viagra*, in 1998. Sales topped $1 billion within a year.

Once again, success brought competition. By 2003, two new ED drugs appeared. *Levitra* (made by Bayer and GlaxoSmithKline) worked twice as fast as *Viagra*, while *Cialis* (developed by ICOS but marketed by Lilly) lasted far longer. Within the United States, the two new drugs combined quickly for as much as 40 percent of the market. In countries such as Australia and France, *Cialis* alone claimed 40 percent of the market within its first year. Once again, generic competition threatens these market positions. Pfizer's UK patent expired in June, 2013. Its US patent survived one court challenge from Teva that barred doctors from prescribing generic substitutes as an ED treatment until 2019 but not as a treatment for hypertension. Some medical analysts are forecasting a significant increase in the number of men reported to be suffering from high blood pressure over the next few years.

Sources: R. Langreth, "High Anxiety: Rivals Threaten Prozac's Reign," *Wall Street Journal*, May 9, 1996, p. A4; A. Pollack, "Lilly Pays Bid Fee Up Front to Share in Rival of Viagra," *New York Times*, October 2, 1998, p. C1; S. Carey, "Lilly Reports 22% Decline in Net Income as Generics Hurt Sales of Prozac," *Wall Street Journal*, April 16, 2002, p. C2; and "Viagra Rival Cialis Wins up to 40 Percent Market Share," *Reuters News Wire*, March 23, 2004; P. Loftus, "Pfizer Viagra Patent Upheld," *Wall Street Journal*, August 15, 2011, p. C1.

or processes inevitably means the destruction of old ones, Schumpeter dubbed such competition by innovation "creative destruction." In addition, because some of the products and processes that are made obsolete may well be those of the innovating firm itself, we can ask our central question in a somewhat different way. Why do firms undermine existing activities (including their own ones) in this way? More generally, what are the incentives to engage in innovative activity and how do these vary with firm size and market structure?

Both the professional and the popular business literature have had much to say on the Schumpeterian hypothesis in recent years. In this chapter, we approach this topic using the tools of economic analysis and strategic interaction that we have built throughout the book. However, before we begin a more formal analysis, we need to establish some definitions or classifications to which we can easily refer.

20.1 A TAXONOMY OF INNOVATIONS

Research and development consists of three related activities. The first is *basic research*. This includes studies that will not necessarily lead to specific applications but, instead, aim to improve our fundamental knowledge in a manner that may subsequently be helpful in a range of activities. The derivation and validation of the theory of laser technology is a good example. A second category is *applied research*. Such research generally involves substantial engineering input and is aimed at a more practical and specific usage than basic research. The creation of the first laser drill for dentistry would be an example of applied research. Finally, there is the *development* component of R&D. Here, the goal is to move from the creation of a prototype to a product that can be used by consumers and that is capable (to some extent at least) of mass production. To continue our analogy, the transformation of the first laser drill into a small, handheld product that is affordable and usable by a large number of dentists would be an example of the development stage. For the most part we shall be concerned with applied research rather than development, but we shall touch upon some of the important issues that characterize the decision to move from research to development.

In considering the output of R&D, it is common to distinguish between two kinds. *Process innovations* are discoveries of new, typically cheaper methods for producing existing goods. *Product innovations* are the creation of new goods. For the most part, we shall concentrate on process innovations, but we shall also present examples showing how the analysis can be extended to product innovations.

Finally, with respect to process innovations, there is a further distinction that can be made. This is the division of innovations into *drastic* or major innovations, and *nondrastic* or minor innovations. Roughly speaking, drastic innovations are ones that reduce a firm's unit cost to such an extent that even if it charges the profit-maximizing monopoly price associated with that low cost, it will still undercut all competitors. Hence, a drastic innovation creates a monopolist unconstrained by any fear of entry or price competition—at least for some time. By contrast, a firm making a nondrastic innovation may gain some cost advantage over its rivals but not one so large that the firm can price like a monopolist without fear of competition.

The formal distinction between drastic and nondrastic innovations is illustrated in Figure 20.1. Assume that demand for a particular product is given by $P = 120 - Q$ and that before the innovation all firms can produce the product at a constant marginal cost of \$80. Assume also that the existing firms are Bertrand competitors so that the price is \$80 and total output is 40 units.

Now suppose that one firm gains access to a process innovation that reduces its marginal costs to \$20 as in Figure 20.1(a) and that, perhaps because of a patent, this firm is the only

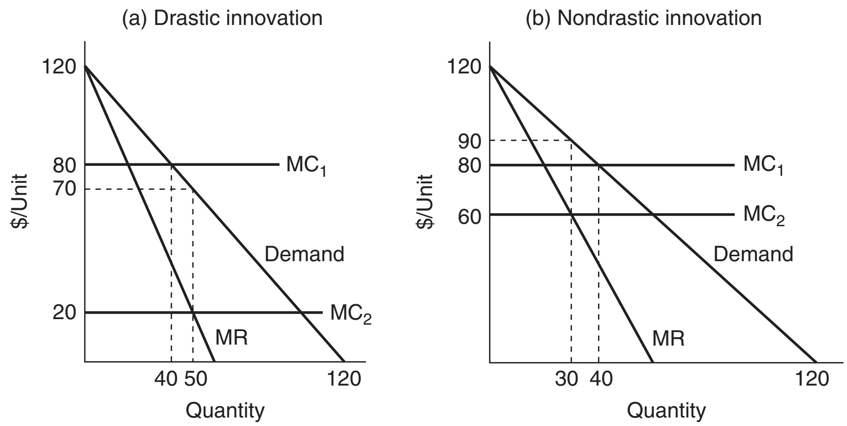

Figure 20.1 Drastic and nondrastic process innovations

one able to use the new low-cost technology. If this innovator were alone in the market, it would set the monopoly price corresponding to its new, lower marginal costs of \$20. Given our demand function we know that marginal revenue is $MR = 120 - 2Q$. Equating this with marginal cost of \$20 gives an output of 50 units and a monopoly price of \$70. Setting this monopoly price forces all the other firms out of the market. The innovation is a drastic one because the reduction in cost is so great that the innovating firm can charge the full monopoly price associated with the new low cost and still be able to undercut the marginal costs of all other firms.

Suppose by contrast that the innovation reduces marginal costs to \$60 as in Figure 20.1(b). By exactly the same argument as above, the innovating firm acting as a monopolist wants to produce an output of 30 units and set a price of \$90. The problem is that this will not work. The remaining firms can profitably undercut this price. So the best that the innovating firm can do now is to set a price of \$80 (more accurately, \$79.99) and an output of 40 units. This still eliminates the other firms but only by the innovator lowering the price that it charges. Hence, this is a nondrastic innovation.

Practice Problem 20.1

Assume that demand in a competitive market is given by the linear function: $P = 100 - 2Q$ and that current marginal cost of production is constant at \$60. Now assume that there is a process innovation that reduces marginal cost to \$28. Show that this is a nondrastic innovation. How much would the innovation have to lower marginal costs for it to be drastic?

20.2 MARKET STRUCTURE AND THE INCENTIVE TO INNOVATE

We now turn to some of the basic questions economists have asked regarding how the incentives to spend on R&D are affected by market structure.[5] We assume the demand for a particular good is linear. Specifically, the inverse demand curve is again assumed to be

[5] This analysis owes much to Nobel Prize winner Kenneth Arrow's path-breaking work (1962).

given by the equation $P = 120 - Q$. We also assume that each producer of the good has a marginal cost of \$80. Accordingly, if the market is competitive and there are many such producers, the current price is also \$80.

20.2.1 Competition and the Value of Innovation

Suppose that a research firm, not involved in the actual manufacture of this good, discovers a new production process by undertaking research at some cost K. Using the notation from above, we consider the case of a nondrastic process innovation that reduces the marginal production cost to \$60. We further assume that the innovation is protected by a patent of unlimited duration that cannot be "invented around" by other potential or actual firms. What benefits does the innovation bring, and does the market mechanism work to convey such incentives to the research firm?

Let us first consider how society as a whole values the innovation. Imagine a social planner whose goal is to maximize total social surplus (producer surplus plus consumer surplus) and, moreover, who has the power to command that prices be set at whatever level the planner requests. Such a benevolent dictator would reason as follows: With or without the innovation, optimality requires that price be set to marginal cost. The per-period value that the social planner places upon the innovation is the increase in consumer surplus when price equals (constant) marginal cost as then there is no producer surplus. Prior to the innovation, consumer surplus at a price of \$80 is \$800. After the innovation, when firms set the price equal to the new lower marginal cost of \$60, consumer surplus increases to \$1,800.[6] The increase in consumer surplus is \$1,000, the shaded area in Figure 20.2(a). This additional surplus will be realized not just in one period but also in all subsequent periods following the innovation. Hence, using the discounting techniques discussed in Chapter 2, the total present value of the additional surplus created by the innovation is $V^p = 1000/(1 - R)$ where $R = (1 + r)^{-1}$ is the discount factor and r is the interest rate. The more this value exceeds the cost K, that is, the more it exceeds the present value of the expenses associated with discovering the process, the more desirable is the innovation.

Of course, we don't have a social planner, and if we did it is doubtful that the planner would succeed in maximizing social welfare. What we have are markets. The issue is how the structure of the market affects the realization of the value of this innovation. What is the incentive of a research firm to pursue the innovation if when it is successful it can auction the rights to the innovation to a competitive industry comprised of many firms? Prior to the innovation all firms sell at a price equal to the marginal cost of \$80 and earn zero profit. Total output each period prior to the innovation is just 40 units.

Now consider the behavior of a firm that has the rights to the innovation. Quite evidently, its best strategy is to undercut its erstwhile competitors just slightly, driving them out of the market and giving it an effective monopoly. The firm that wins the rights will set a price that is one cent less than the old competitive price, \$80. At this price, the industry's total output remains identical to what it was prior to the adoption of the innovation. Consequently, the firm will earn per-period profit of $\$(80 - 60) \times 40 = \$800$. This is illustrated by the shaded rectangle in Figure 20.2(b). The present value that a competitive firm places on the innovation is the maximum amount it willingly bids for the rights to it and this is $V^c = 800/(1 - R)$. This is less than the social value of the innovation. The reason is simple: the competitive firm only considers the profit it can earn as a result of

6 Given our demand function $P = 120 - Q$ and assuming that $P = MC$, consumer surplus is the area of a triangle with height $120 - MC$ and base $120 - MC$. That is, $\text{CS} = (120 - MC)^2/2$.

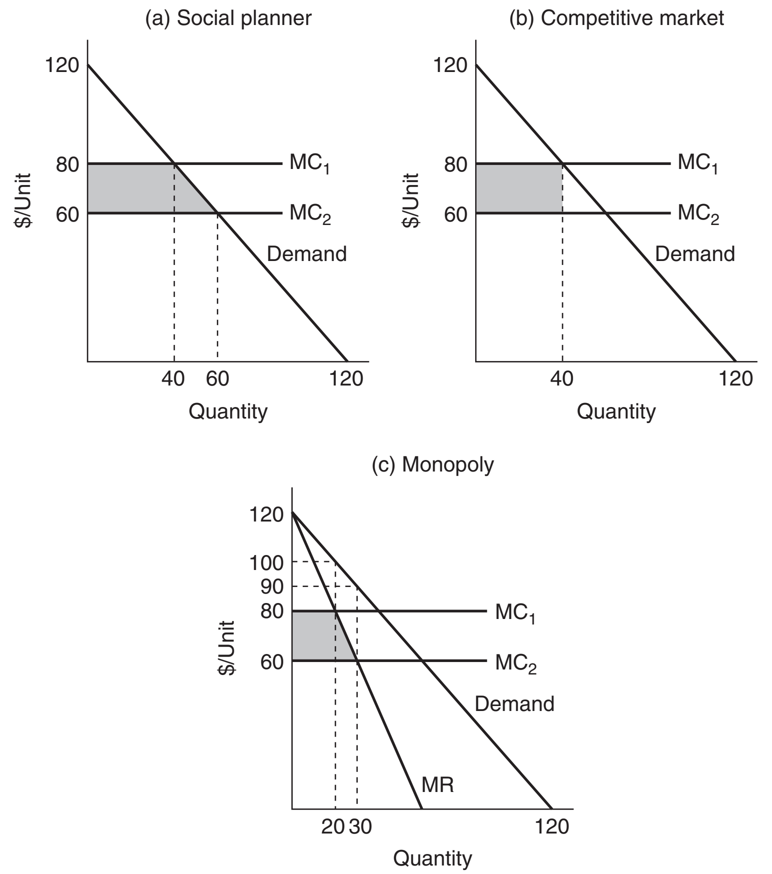

Figure 20.2 Market structure and the incentive to innovate

the innovation. It ignores the additional benefit from increased consumer surplus that the innovation brings.

Now consider the potential value when it is a monopolist who has the rights to the innovation and who faces no threat of entry. For such a firm, the gain from introducing the innovation is the additional profit it makes as a result of being able to produce at a lower marginal cost. Because the monopolist maximizes profit by setting marginal revenue equal to marginal cost, we can measure this gain by comparing the monopolist's per-period profit at its current marginal cost with its per-period profit at the lower marginal cost that the innovation permits. This is illustrated in Figure 20.2(c).

Given our demand function we know that marginal revenue is $MR = 120 - 2Q$. So, prior to the innovation, the monopolist produces an output of 20 units, sets a price of \$100 and earns profit per period of \$400. After the innovation output is increased to 30 units, price is reduced to \$90 and per-period profits are \$900. As a result, the per-period value placed by the monopolist on the innovation is \$500—the difference between profits with and without the innovation. This is illustrated by the shaded area in Figure 20.2(c).[7] In turn, the total present value the monopolist places on the innovation is $V^m = 500/(1 - R)$.

[7] This is derived from the property that one way to represent the monopolist's profit is the area between the monopolist's MR and MC curves.

From the foregoing analysis, it is obvious that $V^p > V^c > V^m$. Both the competitive firm and the monopolist undervalue the innovation relative to the social planner interested in maximizing total welfare. However, a competitive firm values the innovation more than the monopolist.

The reason that the value placed upon the innovation by the monopolist is smaller than the value of the innovation to a competitive firm and to society is again easily explained. A competitive firm is just breaking even prior to adopting the innovation and so values the innovation at the full additional profits it will generate. Like the competitive firm, the monopolist ignores the increase in consumer surplus. In contrast to a competitive firm, however, the monopolist is already earning a profit with its existing technology. Introducing the new process displaces and therefore undermines that investment. This is often referred to as the *replacement effect* but the term is misleading. After all, society also values the innovation by comparing it to the technology that it is replacing. The important reason the monopolist undervalues the innovation is because the monopolist restricts output to less than the socially optimal level. To see why suppose, by contrast, that the monopolist could employ first-degree price discrimination. Then the monopolist's valuation of the innovation would exactly equal society's valuation.

While the comparison just drawn is between a monopolist and a firm in a perfectly competitive market, the results would be the same if we instead compare a monopolist with a firm in an oligopoly market characterized by Bertrand competition. (Why?) Moreover, the same qualitative result will be obtained in a comparison of a monopoly firm with firms engaged in Cournot competition. The basic reason remains. While the Cournot firm does enjoy some positive, pre-innovation profits, these are much smaller than those of a monopolist. Therefore, the Cournot competitor has much less to lose than does the monopolist from pursuing the innovation. While the case just described considered a nondrastic process innovation, the same ordering, $V^p > V^c > V^m$, holds for a drastic one. In other words, the social gain from a drastic innovation exceeds the gain to a firm engaged in Bertrand (or Cournot) competition, which in turn exceeds the gain to a monopolist. Finally, while our analysis assumes a specific linear demand, the same results are obtained for any demand function even if it is nonlinear.[8]

20.2

Practice Problem

Assume that demand for a homogeneous good is $P = 100 - Q$, where P is measured in dollars, and that a process innovation reduces marginal costs of production from \$75 to \$60 per unit. Assume that the discount factor is $R = 0.9$.

a. Confirm that this is a nondrastic innovation and that marginal costs would have to be reduced to less than \$50 per unit for the innovation to be drastic.
b. Calculate the maximum amount that a monopolist is willing to pay for the innovation. Now assume that the market is served by Cournot duopolists who have identical marginal costs of \$75 prior to the innovation.
c. Confirm that the pre-innovation price is \$83.33 and that at this price each firm has profits per period of \$69.44.
d. Confirm that if one of these firms is granted use of the innovation, the price will fall to \$78.33.
e. Show that this firm is willing to pay more for the innovation than the monopolist.

[8] Gilbert (2006) shows that our results generalize to any demand function.

20.2.2 Preserving Monopoly Profit and the Efficiency Effect

The analysis in the previous section assumed that the only innovator is a lab outside the industry. If that laboratory company does not innovate, no one does. This does not truly capture the spirit of Schumpeter's contention. Instead, Schumpeter's point is precisely that firms *compete* by means of innovation. This means that firms have their own labs and that each firm is a potential innovator. As a result, even if one firm does not innovate, another might. This can reverse the previous results.[9]

Suppose that demand is given by $P = 120 - Q$ and that the current technology allows production at a marginal cost of \$60. An incumbent monopolist and a potential entrant play the following three-stage game. In stage 1 the incumbent decides whether or not to undertake R&D, which we assume reduces marginal cost to \$30. In stage 2 a potential entrant decides whether or not to enter. If the incumbent has not undertaken R&D, the entrant then chooses whether or not to undertake R&D. Without R&D, either firm's marginal cost is \$60 and with it marginal cost is \$30. No matter who innovates, the innovation is protected by a patent of unlimited duration that cannot be "invented around" by other potential or actual firms. If entry occurs then in stage 3 the entrant and the incumbent act as Cournot competitors. Using the standard Cournot equations gives the extensive form of this game illustrated in Figure 20.3.

As usual we solve this game "backwards." Suppose that the incumbent has undertaken R&D. Then the entrant will enter with a cost of \$60. The incumbent earns per period profit of \$1,600 and the entrant earns per-period profit of \$100. (This assumes, of course, that there are no sunk costs of entry. We return to this point below.) Now suppose that the incumbent does not innovate. The entrant will certainly enter. Innovation by the entrant gives the entrant per-period profit of \$1,600 while no innovation leads to per-period profit of \$400.

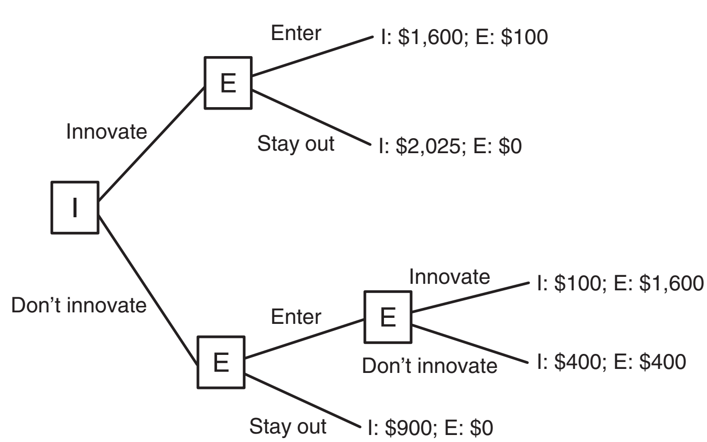

Figure 20.3 Extensive form for the innovation and entry game

[9] The underlying analysis can be found in Gilbert and Newbery (1982). Reinganum (1983) shows, however, that this conclusion might not hold when the timing of the successful breakthrough is uncertain. The incumbent monopolist might delay innovation in order to enjoy its current profits. A potential entrant has no such incentive to delay its innovative activity.

We can now calculate how much the innovation is worth to the incumbent and to the entrant. For the entrant, innovation increases per-period profit from \$400 *to* \$1,600. Accordingly the present value of the innovation to the entrant is $V^e = \$1{,}200/(1 - R)$. What about the incumbent? No innovation by the incumbent will lead to innovative entry provided only that the cost of the innovation is less than $\$1{,}200/(1 - R)$. In that case, the incumbent then earns per-period profit of \$100. By contrast, if the incumbent innovates and pre-empts innovation by the entrant the incumbent earns per-period profit of \$1,600. As a result, the value of the innovation to the incumbent is $V^i = \$1{,}500/(1 - R)$. Clearly this exceeds the value placed on the innovation by the entrant. Hence, the monopolist has the stronger incentive to innovate.

Our analysis illustrates the potential for innovation to deter entry, protecting the incumbent's monopoly position and profit. Suppose that sunk entry costs S are such that $\$100/(1 - R) < S < \$400/(1 - R)$. That is, imagine that sunk costs are greater than the profit that the entrant expects to make if the incumbent innovates but less than the profit that the entrant expects to make if neither entrant nor incumbent innovates. The value of the innovation to the entrant is unchanged at $\$1{,}200/(1 - R)$. This is not the case for the incumbent. Now innovation deters entry, allowing the incumbent to maintain its monopoly position with per-period profit of \$2,025. Failure to innovate, by contrast, leads to innovative entry and per-period profit to the incumbent of \$100. The value of the innovation to the incumbent is now even greater at $V^m = \$1{,}925/(1 - R)$.

The foregoing result is not peculiar to the numbers we have assumed. It is in fact quite general. Suppose first that innovation by the incumbent does not deter entry. Denote the per-period duopoly profit of the incumbent as $\pi_i^d(c_i, c_e)$ and of the entrant as $\pi_e^d(c_i, c_e)$, where c_i is marginal cost of the incumbent and c_e is marginal cost of the entrant. Innovation reduces marginal cost from c_h (high) to c_l (low). The incumbent knows that innovation gives per-period profit $\pi_i^d(c_l, c_h)$ while failure to innovate leads to innovative entry and profit $\pi_i^d(c_h, c_l)$. The entrant innovates is possible only if the incumbent has not innovated. Innovation then gives per-period profit of $\pi_e^d(c_h, c_l)$ while the failure to innovate gives per-period profit of $\pi_e^d(c_h, c_h)$. The per-period value of the innovation to the incumbent is $\pi_i^d(c_l, c_h) - \pi_i^d(c_h, c_l)$ while the per-period value of the innovation to the entrant is $\pi_e^d(c_h, c_l) - \pi_e^d(c_h, c_h)$.

Symmetry between the two firms tells us that $\pi_i^d(c_l, c_h) = \pi_e^d(c_h, c_l)$ and $\pi_e^d(c_h, c_h) = \pi_i^d(c_h, c_h)$. As a result, for the incumbent to place a higher value on the innovation than the entrant requires that $\pi_i^d(c_h, c_l) < \pi_i^d(c_h, c_h)$. This condition is always satisfied. The profit of the incumbent firm when it faces a low-cost rival is less than its profit when it faces a high-cost rival, no matter what the incumbent's marginal costs are.

Now suppose that innovation by the incumbent deters entry. The per-period value of the innovation to the entrant is unchanged. By contrast, the per-period value of the innovation to the monopolist is now $\pi^m(c_l) - \pi_i^d(c_h, c_l)$. This is clearly greater than the value of the innovation with entry because $\pi^m(c_l) > \pi_i^d(c_l, c_h)$. A low-cost incumbent always prefers monopoly to sharing the market, even when the sharing is with a high-cost rival.

To summarize, no matter whether innovation by an incumbent monopolist maintains that monopoly or not, the incumbent firm values the innovation more highly than a potential entrant. Replacing oneself is better than being replaced by a newcomer. This effect is called the *efficiency effect*.

20.3 A MORE COMPLETE MODEL OF COMPETITION VIA INNOVATION

What drives the efficiency effect is the fact that the cost of not innovating becomes higher once we recognize that it is precisely in this case that a rival may innovate. Such an increase in the opportunity cost of non-innovation makes the incumbent monopolist much more willing to pay for the innovation. Clearly, the strategic interaction from potential entry through innovation seems closer to the view Schumpeter (1942) presents.

We can get even closer to the Schumpeterian spirit by making the decision to spend on R&D an explicit part of a firm's strategy. The simplest model in this spirit is one due to Dasgupta and Stiglitz (1980). Their model is attractive both for its key insights and because it builds on the Cournot model developed in Chapter 9. We present the essentials of their analysis here.

Dasgupta and Stiglitz assume an industry comprised of n identical Cournot firms, each of which has to determine the level of output, q_i it will produce and the amount, x_i, that it will spend on R&D. While R&D is costly, the benefit of R&D spending is that it lowers the firm's unit cost of production, c. Specifically, each firm's unit cost is a decreasing function of the amount it spends on R&D, $c_i = c(x_i)$ and $dc(x_i)/dx_i < 0$. Total net profit for any firm, π_i, is:

$$\pi_i = P(Q)q_i - c(x_i)q_i - x_i \tag{20.1}$$

Suppose that each firm spends a specific amount, x^*, on research. Each firm then has a unit cost of $c(x^*)$. Accordingly, if we know the value of x^*, we know each firm's unit cost, and we can work out the equilibrium output for the individual firm and the industry in total using the analysis from Chapter 9.[10] In particular, we know that the outcome in this symmetric, n-firm Cournot model is an equilibrium price-cost margin, or Lerner Index, given by

$$\frac{(P - c(x^*))}{P} = \frac{s_i}{\eta} \tag{20.2}$$

Here, P is the industry price, s_i is the ith firm's share of industry output, η is the elasticity of market demand, and x^* is the amount that each firm spends on R&D in equilibrium. We have dispensed with the subscript on the term x^* because for identical firms the amount chosen is the same in equilibrium for each firm. We can simplify further by recognizing that because all firms are identical, s_i is just $1/n$. So, equation (20.2) can be written as

$$P\left(1 - \frac{1}{n\eta}\right) = c(x^*) \tag{20.3}$$

Equation (20.3) does not by itself tell us the amount of R&D expenditure, x^* that each firm will find optimal. To determine that value we must add a second equilibrium condition. That condition must reflect a choice x^* consistent with profit maximization. Let $-\Delta c$ be the

[10] If we set the derivative of equation (20.1) with respect to q_i to zero, taking the production of all firms other than the ith, Q_{-i} as given, and then solve for q_i we obtain each firm's best response function.

reduction in unit cost that results from a small increase Δx in R&D spending. Since that unit cost reduction applies to all q_i units, the marginal benefit of R&D spending is:

$$\text{Marginal Benefit of R\&D Spending} = -\frac{\Delta c}{\Delta x_i} q_i \tag{20.4}$$

Of course, the marginal cost of an extra dollar of R&D spending is just 1. Profit maximization then requires:

$$\text{Marginal Benefit} = \text{Marginal Cost OR} - \frac{\Delta c}{\Delta x_i} q_i = 1 \tag{20.5}$$

No equilibrium can exist that has obvious ways in which firms—or even just one firm—can raise profit. So, the profit-maximizing condition expressed in equation (20.5) must hold for each firm i, in any equilibrium. Symmetry then implies that we can drop the i subscripts and simply write $-\Delta c/\Delta x = 1$ as the generic representation at each firm.

What are the implications of the equilibrium conditions of equations (20.3) and (20.5)? The most obvious conclusion is that an increase in the number of firms in the industry will decrease the amount that each firm is willing to spend on R&D. An increase in the number of firms in the industry decreases the amount that each firm will choose to produce. This is, actually, a direct implication of equation (20.3). But equation (20.5) makes clear that the marginal benefit of extra R&D spending is directly proportional to a firm's output. Hence, the reduction in a firm's output that results from increasing the number of firms also reduces the marginal benefit that R&D spending yields to an individual firm. It follows that the equilibrium level of such spending per firm, x^*, will fall as the number of firms rises.

This does not necessarily imply, however, that the total industry spending on R&D, which is nx^*, will also fall. It is perfectly possible that each firm spends less on R&D but total R&D spending increases. Dasgupta and Stiglitz show that aggregate spending on R&D may actually either increase or decrease as the number of firms in the industry increases. The key point is that for aggregate R&D spending to increase, the elasticity of market demand must be fairly large. When demand is relatively elastic, the expansion of industry output resulting from a greater number of firms will not decrease the price too much and, as a result, will not decrease the marginal revenue of equation (20.3) very much either. Because this difference between price and cost is what finances a firm's R&D expenditure, such expenditure can be expected to rise in total with the number of industry firms so long as η is relatively large. If, however, the elasticity of market demand declines as output expands (as is the case with linear demand curves), then increasing the number of firms will, beyond some point, lead to a reduction in total R&D efforts. Even for a relatively small number of firms in the market adding one more firm induces a decline in total R&D spending. Therefore, the Dasgupta and Stiglitz model may be taken as partial support for the Schumpeterian hypothesis that concentration fosters innovation.

The foregoing does not explain what determines the number of firms in an industry. Dasgupta and Stiglitz invoke a third equilibrium condition that, in the long run, free entry will lead to an increase in the number of firms until each firm makes zero profit. In other words, industry structure is determined endogenously by the firms' output and R&D expenditure decisions. The zero profit condition, when applied to equation (20.1), tells us that

$$P(Q^*)q^* - c(x^*)q^* - x^* = 0 \tag{20.6}$$

Aggregating this over the equilibrium number of firms in the industry, n^*, gives

$$P(Q^*)Q^* - c(x^*)Q^* - n^*x^* = 0 \tag{20.7}$$

which implies that $(P(Q^*) - c(x^*))Q^* = n^*x^*$. Now because each of the n firms is of the same size, each has a market share equal to $1/n$. By equation (20.2) we know that $P - c(x^*) = P/n^*\eta$. Using this substitution, the equilibrium R&D outcome derived by Dasgupta and Stiglitz is

$$\frac{n^*x^*}{P(Q^*)Q^*} = \textit{industry R\&D spending as a share of industry sales} = \frac{1}{n^*\eta} \tag{20.8}$$

Comparing across industries, equation (20.8) suggests that the share of an industry's total sales revenue that will be devoted to R&D is likely to be smaller in less concentrated industries. In other words, those industries with a naturally more competitive structure will undertake less R&D effort, all else equal. This may then be seen as offering fairly strong theoretical support for Schumpeter's basic claim that imperfect competition is good for technical progress and more imperfect competition is even better.

20.4 EVIDENCE ON THE SCHUMPETERIAN HYPOTHESIS

The debate over the Schumpeterian hypothesis cannot be resolved by an appeal to economic theory alone. We must also consider empirical evidence. To date, a number of statistical studies relating R&D effort to firm size and industry structure have been conducted. While these studies are far from uniform in their results, one general finding does emerge. R&D intensity does appear to increase with increases in industrial concentration but only up to a rather modest value after which R&D efforts appear to level off or even decline as a fraction of firm revenue.

Some of the earliest studies exploring the link between industry structure and R&D were those of Scherer (1965, 1967). His basic finding was that while firm size and concentration are each positively associated with the intensity of R&D spending, these correlations diminish beyond a relatively low threshold. That is, once firms reach a relatively small size and/or markets reach a relatively low level of concentration, any positive effects of firm size or market concentration on innovative activity tend to vanish. Subsequent studies, including those of Levin and Reiss (1984); Levin, Cohen and Mowery (1985 and 1987); Levin, Klevorick, Nelson, and Winter (1987); Lunn (1986); Scott (1990); Geroski (1990); and Blundell, Griffith and Van Reem (1995) have tended to confirm Scherer's (1965) basic finding.[11]

In examining the influence of firm size and market structure on innovative activity, a number of important issues must be addressed. The first of these is that in comparing R&D efforts across markets, we should control for the "science-based" character of each industry: recall the very different patent activity by industry category noted in Table 20.2. Markets in which the member firms produce goods such as chemical products or computer hardware have such a strong technical base that general advances in scientific understanding can rapidly translate into either product or process innovations. Other markets, however, such as

[11] See Cohen and Levin (1989) for an early summary.

those for haircuts or hairstyling, have limited ability to make use of scientific breakthroughs and have less direct contact with universities and research laboratories. It turns out that measures of such technological opportunities tend to be highly correlated with the degree of industry concentration. In other words, while the simple correlation between concentration and innovation may be positive, this correlation reflects the positive effects on innovation that come with increases in an industry's opportunity for technical advances. The more recent studies cited above demonstrate that controlling for this factor is very important.

A second factor is the distinction between R&D expenditures and true innovations as perhaps measured by patents. While innovative effort can be measured by the ratio of R&D spending to sales, this approach really measures the inputs to the innovative process. Presumably though, what we are really interested in are the outputs of that process—the true number of innovations as perhaps measured by the number of patents a firm acquires. Even though different firms do the same amount of R&D spending, the Schumpeterian hypothesis might be validated if size or concentration leads that spending to be more productive. The studies cited above do in fact look at the patent output of firms. Here again, however, little evidence is found in support of the Schumpeterian claims. Cohen and Klepper (1996) conclude that the general finding is that large firms do proportionately more R&D than smaller firms but get fewer innovations from these efforts. A notable exception in this regard, however, is Gayle (2002) who finds that firms in concentrated industries do generate many more patents when patents are not simply counted but, instead, are measured on a citation-weighted basis.[12]

Finally, a third issue is the endogeneity of market structure. Some firms, for example Alcoa or Microsoft, came to dominate their industry on the basis of a dramatic innovation. In the case of Alcoa, it was its unique process for refining aluminum. In the case of Microsoft, it was its unique *Windows* operating systems for personal computers. In these and other cases, the key technology that led to the firm's dominant position was associated with a number of patents. If this experience is pervasive, a naïve researcher may find that large, dominant firms are also firms with many patents and wrongly conclude that the Schumpeterian hypothesis is validated. In these cases, it is the innovative activity that leads to market power and not the other way around. If the firms that come to dominate their markets start out as small operations and then grow on the basis of entrepreneurial skill and technical breakthroughs, the implication would be quite to the contrary of Schumpeter's model.[13]

20.5 R&D COOPERATION BETWEEN FIRMS

Our final topic in this chapter is the issue of cooperation on R&D efforts among firms. Two features of the innovative process make such efforts attractive from the viewpoint of economic efficiency. First, modern technology is very complicated and often draws on different areas of expertise and experience. Because it is doubtful that the scientists

[12] When a patent application is filed, the applicant must cite all the prior patents related to the new process or product. It is plausible that the most important patents are those that are cited most frequently. Hence, in evaluating a firm's true innovative output, one may want to control for how often the firm's patents are cited.

[13] Generally, market structure and innovative activity evolve together. For example, if experience raises R&D productivity then older firms will tend to do more innovation because it has a higher return for them, so that early entrants will tend to dominate an industry over time. See Klepper (2002) for an analysis along these lines.

and engineers in one firm possess all this know-how, it is desirable that firms share their experiences, experimental results, and design solutions with each other so as to realize fully the benefits from scientific study. Second, there is a potential for wasteful R&D spending as firms duplicate each other's efforts in a noncooperative R&D race.

We do have explicit evidence on this score. In the 1980's, steel minimills emerged as one of the most dynamic sectors in the US economy. These firms rely on small-scale plants using electric arc furnaces to recycle scrap steel. They became widely regarded as world leaders and have outperformed even the Japanese steel firms once thought to be invincible. Through a series of interviews, von Hippel (1988) found that these firms regularly and routinely exchanged technical information with each other. In fact, sometimes workers of competing firms were trained (at no charge) by a rival company in the use of specific equipment. Such exchanges of information and expertise were made with the knowledge and approval of management even though they had the effect of strengthening a competitor.

To analyze the implications of research spillovers, we again make use of the Cournot duopoly model, similar to the Dasgupta and Stiglitz (1980) model except that we now explicitly permit one firm's research to benefit others.[14] We address three issues. First, how do technical spillovers affect the incentives firms have to undertake R&D? Second, what is the impact of such spillovers on the effects of R&D? Finally, what are the benefits to be gained from allowing firms to cooperate in their research, for example, by forming research joint ventures (RJVs)? Are these benefits worth the risk that cooperation in R&D might facilitate collusion between the same firms in the final product markets?

To begin, suppose that the demand for a homogeneous good is linear and given by $P = A - BQ$. Two firms, each of which has constant marginal costs of c per unit, manufacture the good. These costs can be reduced by research and development activity, but it is possible that the knowledge developed by one firm can spill over to its rival. This can happen, for example, because the firms fund common sources of basic research such as universities or research laboratories; or because the research direction that one firm is taking becomes known to its rival; or because some of the preliminary results of research effort leak out; or, of course, because of industrial espionage.

Specifically, if firm 1 undertakes R&D at intensity x_1 and firm 2 undertakes R&D at intensity x_2, the marginal production costs of the two firms become

$$\begin{aligned} c_1 &= c - x_1 - \beta x_2 \\ c_2 &= c - x_2 - \beta x_1 \end{aligned} \tag{20.9}$$

Here $0 \leq \beta \leq 1$ measures the degree to which the R&D activities of one firm spill over to the other firm.[15] If $\beta = 0$, there are no spillovers—firm 1's research effort x_1 yields benefits only to firm 1 itself. If $\beta = 1$, spillovers are perfect—every penny of cost reduction that x_1 brings to firm 1, it also brings to firm 2. For the intermediate case of $0 < \beta < 1$, spillovers are only partial—if firm 1's research lowers its own cost by one dollar per unit, it will lower firm 2's cost by some fraction of a dollar per unit.

14 The model is developed in d'Aspremont and Jacquemin (1988). A more general version of this type of investigation can be found in Kamien et al. (1992).

15 We confine our attention to the case in which the spillovers are positive. It is, however, possible that there might be negative spillovers. For example, firms might spread misinformation about their research or claim that they have made a breakthrough in order to discourage rivals from continuing with a particular line of research.