

Research is, of course, costly. Indeed, not only is it costly, but we assume that R&D activity exhibits *diseconomies of scale*, i.e., it becomes more costly the more research the firm does. Specifically, we assume that research costs are the same for both firms and given by the research cost function

$$r(x_i) = \frac{x_i^2}{2}, \quad i = 1, 2. \quad (20.10)$$

Thus, if the R&D intensity is $x_i = 10$, then the research budget $r(x_i) = 10^2/2 = \$50$. If the R&D intensity doubles to $x_i = 20$, the budgetary expense climbs to $20^2/2 = \$400$. A doubling of R&D effort therefore leads to a quadrupling of the R&D cost. This is an example of what we mean by a scale diseconomy.

20.5.1 Noncooperative R&D: Profit, Prices, and Social Welfare

Consider first what happens when firms do not cooperate on research. Suppose that we have a two-stage game and in the first stage, each firm chooses its research intensity x_i . In the second stage, each firm acts as a Cournot competitor in choosing its output. As usual, this game is solved backwards. Again, from Chapter 9 we know the Cournot equilibrium outputs for given values of c_1 and c_2 are

$$\begin{aligned} q_1^C &= \frac{(A - 2c_1 + c_2)}{3B} \\ q_2^C &= \frac{(A - 2c_2 + c_1)}{3B} \end{aligned} \quad (20.11)$$

and the firm profits after paying the research costs are

$$\begin{aligned} \pi_1^C &= \frac{(A - 2c_1 + c_2)^2}{9B} - \frac{x_1^2}{2} \\ \pi_2^C &= \frac{(A - 2c_2 + c_1)^2}{9B} - \frac{x_2^2}{2} \end{aligned} \quad (20.12)$$

We also know from equation (20.9) that $c_1 = c - x_1 - \beta x_2$ and $c_2 = c - x_2 - \beta x_1$. This allows us to express the final equilibrium outputs directly as a function of each firm's choice of R&D effort and the degree of spillover from one firm's findings to the other firm's costs. The resultant Cournot–Nash equilibrium outputs for each firm are

$$\begin{aligned} q_1^C &= \frac{(A - c + x_1(2 - \beta) + x_2(2\beta - 1))}{3B} \\ q_2^C &= \frac{(A - c + x_2(2 - \beta) + x_1(2\beta - 1))}{3B} \end{aligned} \quad (20.13)$$

and their profits are

$$\begin{aligned} \pi_1^C &= \frac{(A - c + x_1(2 - \beta) + x_2(2\beta - 1))^2}{9B} - \frac{x_1^2}{2} \\ \pi_2^C &= \frac{(A - c + x_2(2 - \beta) + x_1(2\beta - 1))^2}{9B} - \frac{x_2^2}{2} \end{aligned} \quad (20.14)$$

Equation (20.13) indicates that the output of each firm is an increasing function of its own R&D expenditures x_i . Such expenditures reduce a firm's costs and thereby make higher output more profitable. By contrast, the effect of the *rival's* R&D effort on a firm's production can go either way. Consider firm 1. On the one hand, the R&D activity of firm 2 spills over and lowers firm 1's costs, which has an expansionary effect on firm 1's own output. On the other hand, firm 2's R&D reduces firm 2's cost. This makes firm 2 more competitive and permits it to expand its output leaving less market available to firm 1. The net result of these two countervailing forces is ambiguous. This ambiguity is reflected in the coefficient on x_2 in the q_1 equation and the coefficient of x_1 in the q_2 equation. In both cases, this coefficient, $2\beta - 1$, is positive only when the degree of spillover is large, that is, when $\beta > 0.5$. When spillovers are small, that is, when $\beta < 0.5$, a firm's output and profit are decreasing functions of the R&D expenditures of its rival. The same ambiguity appears in the profit equations (20.14).

We know that each firm will choose the level of research activity that maximizes its profit given the research effort of its rival. For every choice of effort that firm 2 makes, firm 1 will choose its own profit-maximizing response. The same is true for firm 2 in reverse. So we can in principle identify the best response or *research intensity reaction function* for each firm.

This is done in the Appendix to this chapter, allowing us to specify the research intensity reaction functions as follows:

$$R_1 : x_1 = \frac{2(2 - \beta)[A - c + x_2(2\beta - 1)]}{[9B - 2(2 - \beta)^2]} \text{ and } R_2 : x_2 = \frac{2(2 - \beta)[A - c + x_1(2\beta - 1)]}{[9B - 2(2 - \beta)^2]} \quad (20.15)$$

Inspection of these functions confirms an intuitive result that follows from our previous discussion. When research spillovers are low, the research intensity reaction functions for the two firms are downward sloping, indicating that the research expenditures of the two firms are *strategic substitutes*—more research by one firm reduces the amount done by the other. That is, research activity by one firm substitutes for research activity by the other. The intuition is that in this case the increased research effort by one firm primarily reduces its costs and so gives it a competitive advantage with respect to the other rival firm. In turn, this results in a reduction in the profitability of the rival firm, which can be offset only by the rival reducing its expenditure on research.

By contrast, when spillovers are high, the research intensity reaction functions are upward sloping, meaning that the research expenditures of the two firms are *strategic complements*. When spillovers are this high, an increase in research intensity by one of the firms induces an increase in research intensity by the other. In this case the intuition is that if one firm opts for a high level of R&D effort, the benefits of that activity spill over to the other firm to such an extent that the other firm's profit increases, providing that firm with the funds and the incentive to increase its own R&D spending. Figure 20.4 illustrates typical research intensity reaction functions.

However, determining whether the reaction functions slope downward or upward—whether the two firms' R&D efforts are strategic substitutes or complements—does not tell us what the equilibrium level of R&D spending is. In particular, there can be no presumption that the presence of large R&D spillovers and hence the case of strategic complements will

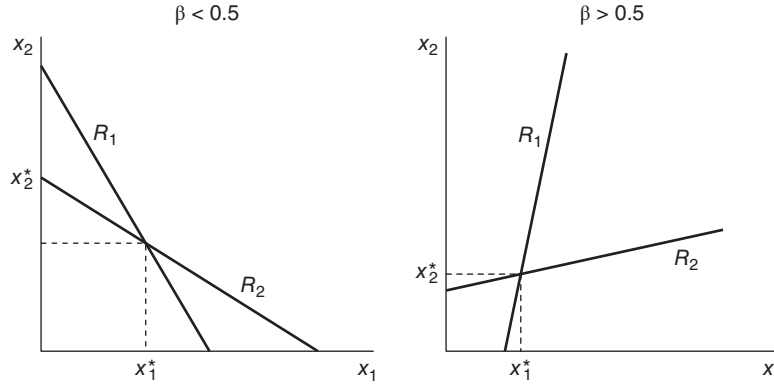


Figure 20.4 Best response functions for research intensity in the noncooperative R&D game

result in a higher equilibrium level of R&D spending than the case in which such spillovers are low. The Nash equilibrium occurs at the intersection of the two response functions, and the case in which this point is farthest from the origin is far from obvious.

In order to illustrate this last point, we focus for the remainder of our discussion on a numeric example. The Appendix gives a more general mathematical analysis. Let demand for the good be $P = 100 - 2Q$, and each firm's marginal production cost currently be \$60. The firms can choose two levels of research intensity: $x_i = 10$ or $x_i = 7.5$. Further, we assume that the degree of research spillover (which is outside the control of the two firms) takes one of two values: a low value of $\beta = 1/4$ or a high value of $\beta = 3/4$.

Consider first the low-spillover case and assume that firm 2 chooses the high research intensity of $x_2 = 10$. If firm 1 also chooses high research intensity, its output and profits will be, from equations (20.13) and (20.14)

$$q_1^C = \frac{(40 + 17.5 - 5)}{6} = 8.75; \pi_1^C = \frac{(40 + 17.5 - 5)^2}{18} - \frac{100}{2} = \$103.13$$

By contrast, if firm 1 chooses the low research intensity, its output and profits will be

$$q_1^C = \frac{(40 + 13.125 - 5)}{6} = 8.02; \pi_1^C = \frac{(40 + 13.125 - 5)^2}{18} - \frac{56.25}{2} = \$100.54$$

Now assume that firm 2 chooses the low research intensity of $x_2 = 7.5$. If firm 1 chooses the high research intensity, its output and profits will be

$$q_1^C = \frac{(40 + 17.5 - 3.75)}{6} = 8.96; \pi_1^C = \frac{(40 + 17.5 - 3.75)^2}{18} - \frac{100}{2} = \$110.50$$

By contrast, if firm 1 chooses the low research intensity, its output and profit will be

$$q_1^C = \frac{(40 + 13.125 - 3.75)}{6} = 8.23; \pi_1^C = \frac{(40 + 13.125 - 3.75)^2}{18} - \frac{56.25}{2} = \$107.31$$

Table 20.3(a) Payoff matrix with low R&D spillovers, $\beta = 0.25$

		<i>Firm 2</i>	
		<i>Low Research Intensity</i>	<i>High Research Intensity</i>
<i>Firm 1</i>	<i>Low Research Intensity</i>	\$107.31; \$107.31	\$100.54; \$110.50
	<i>High Research Intensity</i>	\$110.50; \$100.54	\$103.13; \$103.13

Table 20.3(b) Payoff matrix with low R&D spillovers, $\beta = 0.75$

		<i>Firm 2</i>	
		<i>Low Research Intensity</i>	<i>High Research Intensity</i>
<i>Firm 1</i>	<i>Low Research Intensity</i>	\$128.67; \$128.67	\$136.13; \$125.78
	<i>High Research Intensity</i>	\$125.78; \$136.13	\$133.68; \$133.68

The same calculations apply to firm 2. We then have the payoff matrix of Table 20.3(a). *The Nash equilibrium in this case of low spillovers is for both firms to adopt high research intensities.*

When the degree of R&D spillover is high, with $\beta = 0.75$, the same calculations lead to the payoff matrix of Table 20.3(b). *The Nash equilibrium in this case is for both firms to adopt low research intensities.*

An increase in the degree of R&D spillover causes the two firms to reduce their research intensities. Why? Consider first the case in which R&D spillovers are weak. In this case the more firm 1 spends on R&D, the less firm 2 will spend because the primary effect of such spending is to strengthen the competitive position of firm 1. Yet somewhat paradoxically, this gives each firm an incentive to spend aggressively on R&D so as to avoid being the loser in this competition. If firm 1 spends a lot on R&D and firm 2 spends nothing, virtually all the benefits of firm 1's spending stay with firm 1. Firm 2 would find itself losing significant market share and profit to a much lower-cost competitor. When each firm tries to avoid falling behind in this manner, the net result can be a substantial amount of R&D effort, both at each firm and in total.

Just the opposite holds in the case of large spillovers. Yes, the more firm 1 spends on R&D, the more firm 2 is induced to spend by virtue of the strategic-complements setting, but this relation is a two-edged sword. Even if firm 1 spends only a little on R&D it knows that this will still induce firm 2 to do a fair bit of research activity. Moreover, the research activity at firm 2 will bring substantial benefits to firm 1 by virtue of the large spillovers. In this setting, the incentive for either firm to spend much on R&D can be quite small as each firm seeks to free ride on the other's efforts.

Where graphs fail to give clear results, algebra can often save the day. That this is true here is shown in the Appendix. The Nash equilibrium level of research intensity is:

$$x_1^C = x_2^C = \frac{2(A - c)(2 - \beta)}{9B - 2(2 - \beta)(1 + \beta)} \quad (20.16)$$

The amount of research done by each firm *decreases* as β , the degree of R&D spillover, increases—the free-riding effect to which we have just referred.

20.5.2 Technology Cooperation

We now consider two other possible arrangements between the duopoly firms that can alter the outcome from that described above. The first possibility is that the two firms agree that while each will continue to do its own R&D, they will coordinate the extent of such research effort. Thus, the two firms now choose x_1 and x_2 to maximize their joint profit. They continue to recognize that they will compete as Cournot firms in the product market. The other possibility we consider is that the firms explicitly share their R&D activities by setting up a Research Joint Venture (RJV). One way this scenario might work in practice would be for the two firms to jointly set up a laboratory for experimentation and analysis with all the discoveries made at that laboratory to be made fully available to both firms.

We introduce this RJV arrangement into the model by letting the two firms pick x_1 and x_2 cooperatively but by adding the further assumption that the degree of spillover is complete, that is, $\beta = 1$. Whatever is learned in the research lab—whether discovered by a firm 1 scientist or a firm 2 scientist—reduces the cost of both firms by the same amount.

We start with the simple coordination case. What we want to do is to pick the values of x_1 and x_2 that maximize the sum of the individual profit expressions shown in equation (20.14). The mathematical solution, as in the previous section, is developed in the Appendix to this chapter. The equilibrium research intensity is:

$$x_1^{RC} = x_2^{RC} = \frac{2(A - c)(1 + \beta)}{9B - 2(1 + \beta)^2} \quad (20.17)$$

We shall concentrate once again on our simplified example as given by the payoff matrices of Tables 20.3(a) and 20.3(b). When the firms coordinate their research efforts they choose the combination of R&D intensities that maximizes the sum of the profits in the cells of the relevant payoff matrix. *When R&D spillovers are low, coordination leads each of the firms to choose the low R&D intensity.* Cooperation then increases each firm's profits from \$103.13 to \$107.31. By contrast, *when the R&D spillover is high, coordination leads each firm to choose the high R&D intensity*, increasing their profits from \$128.67 to \$133.68.

What our example and the more detailed analysis in the Appendix show is the following. First, it is now the case that the higher is the level of R&D spillover—the larger is β —the more each firm spends on research: this follows directly from equation (20.17). This is because the agreement between the two firms to set their R&D efforts jointly explicitly forces each firm to internalize the external benefits that such spending has upon its rival. In turn, this eliminates the free-riding problem that characterizes R&D competition when there are high spillovers. The ability to avoid this problem also means that the two firms will each enjoy a profit at least as great as that which they would have earned in the absence of such cooperation.¹⁶

The second point to note is that the outcome under the simple coordination plan may not necessarily be good for the consumer. In particular, consumers are hurt by the technology cooperation when $\beta < 0.5$ and the extent of spillover is small. The reason is straightforward. When β is small, then without cooperation each firm tends to do a fair bit of research. It does so because a low value of β means that most of the benefit from its innovative efforts will accrue to it alone and because it knows that its rival is proceeding along the same line of

¹⁶ For $\beta = 0.5$ each firm earns exactly the same profit under uncoordinated R&D spending as each does with an R&D cartel. For all other values of β each firm's profit is higher with the R&D cartel.

attack. From the viewpoint of consumers, this is great because there has been considerable cost reduction and therefore a sharp decline in the price they pay. If, in this setting, we now introduce a cooperative R&D agreement, the two firms realize that their best bet is to reduce R&D intensity, which otherwise simply makes competition tougher in the product market. By decreasing R&D intensity, the firms increase their profits. Unfortunately, the lower rate of innovation also implies a higher price to consumers.

By contrast, when the degree of R&D spillover is high ($\beta > 0.5$), both firms and consumers benefit from a cooperative R&D agreement. This happens because now the primary effect of the R&D cooperation is to correct a market failure. In the absence of cooperation, a large degree of spillover leads each firm to free ride on the R&D efforts of its rivals and to take no account of the beneficial effects its own R&D expenditures have on the costs and profits of other firms. R&D cooperation internalizes these effects because it forces the cooperating firms to look at the impact their R&D expenditures have on aggregate profit rather than merely on their individual profit.

What about a research joint venture? As noted, an RJV can be best thought of as a case in which the firms take actions not only to coordinate their research expenditures but also to ensure that the spillover from one firm's research to the other's is complete, that is, so that $\beta = 1$. A little thought should convince you that an RJV will likely yield the maximum benefits to both firms and consumers. As we just saw, coordination of R&D levels benefits both producers and consumers whenever $\beta > 0.5$. Moreover, the profit outcomes in Tables 20.3 indicate that if the firms could find some way of increasing the technology spillover from $\beta = 0.25$ to $\beta = 0.75$, they would both benefit at *any* research intensity. We can make this discussion even more general. In the Appendix we show that an increase in the spillover parameter β increases the research intensity and the profits of each firm *and* increases the output that each firm brings to the market. In other words, *both firms and consumers benefit* from an increase in β . The RJV takes this to its logical conclusion by ensuring that $\beta = 1$, its highest possible value. The equilibrium research intensity with $\beta = 1$ is:

$$x_1^{RJV} = x_2^{RJV} = \frac{4(A - c)}{9B - 8} \quad (20.18)$$

The RJV gives higher profits *and* lower prices than any other arrangement.

In our example, the benefits of an RJV are easily confirmed. Consider the case in which both firms choose the high degree of research intensity. Thus, with perfect R&D spillover, the profits to each firm are $(40 + 10 + 10)^2/18 - 50 = \150 , while if each chooses the low research intensity, the profits to each firm will be $(40 + 7.5 + 7.5)^2/18 - 56.25/2 = \139.93 . Clearly, the RJV will go for the high research intensity leading to the lowest costs. In turn, this will translate into the lowest consumer prices that these Cournot firms will offer.

The intuition behind the foregoing analysis is as follows. First, by maximizing the extent of spillovers, the RJV also maximizes the benefits of R&D. Every discovery is spread instantly to all firms in the industry. Second, despite this perfect spillover, the free-riding problem is now avoided. Because the two firms have agreed to coordinate their research efforts they fully internalize the otherwise external effects of research. Thus, firms will pursue extensive research, which, partly because of the perfect spillover effect of sharing, will lead to a sizable reduction in costs for every firm. This substantial cost reduction translates into an equally impressive reduction in the price to consumers.¹⁷ The policy

¹⁷ While we have derived this result for a duopoly, Kamien et al. (1992) show that it extends to an n -firm oligopoly.

implication of this is obvious and important. Research joint ventures should be encouraged because they benefit both consumers and producers so long as the antitrust authorities can ensure that such cooperation on research effort does not also extend to cooperation in production and prices, that is, to a price-fixing cartel.

The potentially large benefit from technology cooperation is undoubtedly the reason that research joint ventures—unlike price-fixing agreements—are not treated as *per se* violations by the antitrust authorities. Instead, they are evaluated on a rule of reason basis. Indeed, the US Congress passed legislation in 1984 to require explicitly the application of a reasonability standard in the specific case of RJVs.

20.6 EMPIRICAL APPLICATION: R&D SPILLOVERS IN PRACTICE

R&D spillovers suggest a diffusion-like process. The greater the spillover, the more rapid or the more complete is the diffusion of technological advances in one firm to the productivity of other firms. We might also suspect that a similar process is at work at a national and even international level. In particular, it seems likely that the R&D efforts of one country could spill over to enhance the productivity of its neighbors. Here again, the extent of such spillover is of interest. If technical advances in one country spread quickly to others, β will be high. In a world in which the international transfer of technical knowledge is weak, β will be low.

Wolfgang Keller (2002) explores the extent of international technical spillover by looking at data covering twelve broadly defined manufacturing industries from fourteen countries over the years, 1970–95. To understand his basic approach, consider a simple Cobb-Douglas production function (see Section 4.5) for industry i in country c .

$$Q_{ci} = K_{ci}^{1-\sigma} L_{ci}^{\sigma} \quad (20.19)$$

Here Q_{ci} is output (value added) and K_{ci} and L_{ci} are capital and labor inputs, respectively, in industry i in country c and σ is the share of costs accounted for by labor. Taking logarithms then yields:

$$\ln Q_{ci} = (1 - \sigma) \ln K_{ci} + \sigma \ln L_{ci} \quad (20.20)$$

For industry i , we define $\ln \bar{Q}_{ci}$ to be the average log of output across all countries. Similarly, for industry i , let $\ln \bar{K}_{ci}$ and $\ln \bar{L}_{ci}$ be the average amount of capital and labor inputs (again in logs), respectively, across all countries. If we define total factor productivity, TFP_{ci} , in industry i and country c as the difference between the log of output and the weighted average level of inputs, i.e., $TFP_{ci} = \ln Q_{ci} - (1 - \sigma) \ln K_{ci} - \sigma \ln L_{ci}$, then the *relative* (to the mean) factor productivity F_{ci} of industry i in country c at a point in time is:

$$F_{ci} = (\ln Q_{ci} - \ln \bar{Q}_{ci}) - (1 - \sigma_{ci})(\ln K_{ci} - \ln \bar{K}_{ci}) - \sigma_{ci}(\ln L_{ci} - \ln \bar{L}_{ci}) \quad (20.21)$$

where we now let the cost share of labor σ_{ci} vary across countries and industries. Equation (20.21) is a measure of the extent to which output in industry i in country c remains above average even after correcting for any above average use of inputs. It is thus a measure of the productivity advantage (or disadvantage) in industry i in country c at any point in time. This is why it is called *relative* productivity. Of course, this will probably

change over time due to R&D and other factors. For this reason, Keller (2002) measures relative productivity for each year from 1970 to 1995. This means that for each of the twelve industries in each of the fourteen countries, Keller (2002) has a measure of relative productivity in each year, 1970 to 1995. Because we are now measuring this term over time as well as over industries and across countries, relative factor productivity now has an additional time subscript, i.e., F_{cit} . It is this series of relative productivity measures that Keller seeks to explain.

Because, by construction, variations in the relative productivity measure F_{cit} reflect variations in factors other than capital and labor inputs, it is natural to identify these remaining differences as those due to differences in technology. In turn, these technical differences ought to reflect differences in R&D. Keller's (2002) approach in this respect is to distinguish between R&D done domestically in industry i and that done abroad. The first research question is whether foreign R&D spills over to domestic productivity. The second is whether these spillovers are greater for countries that are closer to each other.

The fourteen countries in Keller's (2002) sample are: Australia, Canada, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Spain, Sweden, the United Kingdom, and the United States. Five of these countries—France, Germany, Japan, the United Kingdom, and the United States—account for over 92 percent of all the R&D in the sample. Hence, Keller (2002) treats these G5 countries as the potential engines of technical change and examines how their R&D affects productivity in the remaining nine. Specifically, he estimates the parameters of the following equation:

$$F_{cit} = \alpha_{ci} + \alpha_t + \lambda \ln \left[S_{cit} + \gamma \left(\sum_{g \in G5} S_{git} e^{-\delta D_{cg}} \right) \right] + \varepsilon_{cit} \quad (20.22)$$

Here, F_{cit} is the relative productivity measure derived above for industry i in country c in year t , measured for each of the nine countries examined. The first term is a country and industry specific constant that permits for a time-independent productivity advantage or disadvantage for that sector in that country. The second is meant to pick up productivity increases over time that affect all firms in all countries in common. The key parameters are embedded in the next term. S_{cit} is a measure of the R&D done in industry i in country c up to time t . In contrast, S_{git} is a measure of R&D in that same industry but in one of the G5 countries and D_{cg} is the distance of that country from the domestic country in question. ($D = 1$ implies a distance of 235 kilometers.) Together, S_{cit} and the summation term for the G5 countries are meant to capture the R&D relevant to productivity in the domestic industry i at time t .

The effect of that combined industry-based R&D on productivity in that same industry in the domestic country is captured by the parameter λ . However, two adjustments are included to distinguish the impact of foreign R&D from that of domestic R&D. To understand the first of these adjustments suppose that all G5 countries were right next to the domestic country in question ($D_{cg} = 0$). Then the contribution of their R&D on domestic productivity in industry i to total industry-relevant R&D is adjusted by the parameter γ (taken to be same for all G5 countries). That is, if $\gamma < 1$, foreign R&D contributes less than the full effect of domestic R&D in adding to the knowledge relevant to a particular domestic industry's productivity. In many ways then, γ is comparable to the β of our industry analysis above. However, Keller (2002) also introduces a second source of distinction between domestic and foreign research lies in the distance term, D_{cg} . As this distance grows, the contribution

of that G5 country's R&D to domestic productivity diminishes further if the parameter δ is positive. In short, the specification permits both for the possibility that simply by being foreign R&D may contribute less to domestic technology than home-grown research, and also for the more complicated fact that spillovers from foreign R&D grow smaller as the source of that R&D is farther away. Of course, the error term ε_{cit} picks up any remaining random factors that affect productivity.

The specification in equation (20.22) assumes that the decay parameter δ is the same throughout the time period. Keller (2002) recognizes, however, that increased globalization over the twenty-five years of his sample suggests that δ will decline over this period. He therefore estimates an alternative specification given by:

$$F_{cit} = \alpha_{ci} + \alpha_t + \lambda \ln \left[S_{cit} + \sum_{g \in G5} \gamma_G (1 + \psi_F I_t) S_{git} e^{-\delta(1 + \psi_D I_t) D_{cg}} \right] + \varepsilon_{cit} \quad (20.23)$$

In this equation, I_t is a 1,0 dummy variable equal to zero over the first half of the sample to 1982 and then 1 in the thirteen years thereafter. The coefficient ψ_F permits the effect of G5 R&D to have a different effect on domestic productivity in the second half of the sample than it does in the first, holding the distance between the domestic and G5 countries constant. Similarly, the coefficient ψ_D permits the extent to which spillovers decline with distance to change from the first half of the sample to the second half. Note too that this specification allows the effect of G5 R&D to differ across each G5 country by permitting a different coefficient γ_G for each one. This is reasonable as the different languages in these countries may affect the ease with which a technology can be transferred.

Because the contribution of foreign R&D to the total relevant measure of R&D depends on the parameters γ (or γ_G) and δ that are also to be estimated, equations (20.22) and (20.23) cannot be estimated by ordinary least squares. Instead, a nonlinear least squares estimation is required. In this procedure, we begin with a starting value for the nonlinear parameters and then estimate the regression with OLS. We may then use these estimates to reiterate the process until the coefficient estimates stop changing and converge to stable values. Table 20.4 shows the key parameter estimates and their standard errors that Keller (2002) obtains from this maximum likelihood process for both specifications.

Table 20.4 Regression estimates of international R&D spillovers

Parameter	Specification 1		Specification 2	
	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
λ	0.078	(0.013)	0.096	(0.008)
δ	1.005	(0.239)	0.384	(0.047)
γ	0.843	(0.059)	—	—
γ_J	—	—	1.000 (set)	set
γ_{US}	—	—	1.031	(0.059)
γ_{UK}	—	—	0.863	(0.060)
γ_{GER}	—	—	1.157	(0.060)
γ_F	—	—	1.011	(0.060)
ψ_D	—	—	−0.784	(0.068)
ψ_F	—	—	−0.061	(0.108)

The estimates in Specification 1 suggest that technical spillovers are strongly localized. The cumulative productivity effect of overall R&D is to raise productivity by 7.8 percent. However, foreign (G5) R&D contributes only 84 percent to the technical base that domestic research does and that is only if the domestic country is right next to the G5 source nation so that $D = 0$. The estimate of $\delta = 1.05$ though, indicates that that contribution dies out rapidly. Half of it is gone when $D = 0.69$ or at a distance of 162 kilometers (100 miles), and the rest is virtually eliminated once the source country of the foreign R&D is more than 400 miles away.

However, the results from Specification 2 qualify the foregoing findings. It is useful to first note that the estimate of ψ_F is insignificantly different from zero. Hence, correcting for distance and country of origin, the contribution of a G5 country's R&D on domestic technical know-how is pretty much the same throughout the sample years and, on average, not too different from the 84 percent found in Specification 1 when the distance to the G5 country is $D = 0$. The real change comes in the extent to which the impact of G5 R&D declines with distance. Now the estimate of δ is a much smaller 0.384, indicating that the effect declines much more slowly with distance, even in the first half of the sample when $I_t = 0$. Over the latter half, though, when I_t is 1, the estimate of $\psi_D = -0.784$ indicates that this small rate of decline is even smaller from 1983 to 1995 than it was previously. Together, these estimates indicate that at least half of the effect that a G5 nation's R&D would have had if the domestic country had been right next to it ($D = 0$) is still there as far out as 424 kilometers (263 miles) from 1970 to 1982, and is felt as far out as 1,963 kilometers (1,217 miles) after 1983. This implies a larger and growing degree of technical spillovers between industries in different nations.

Because Keller's spillover estimates apply to whole sectors separated by national boundaries there may well be a lower bound for the extent of such spillovers between firms within the same domestic industry. If this is so, then these empirical estimates when taken together with our analysis of the noncooperative outcome with high spillovers suggest that the market will likely be characterized by inefficiently low R&D. If that is the case, then the argument for permitting R&D cooperation and/or joint ventures becomes noticeably more compelling.

Summary

Research and development is the wellspring of technical advancement. Such advancement is the key source of the gain in per capita income and living standards that has characterized the developed economies for almost all of the past 200 years. Firms often compete by investing in R&D projects in the hope of uncovering cost-saving innovations and new consumer products. Such competition in R&D raises the question as to whether the market will yield efficient R&D outcomes. It also raises a potential tension between allocative efficiency at a point in time and dynamic efficiency over time.

The tension between the gains from competition and the gains from innovation, i.e., the tension between the replacement effect

and the efficiency effect is unavoidable. It has led economists to consider which market environment—competitive or monopolistic—will foster greater research and development. The Schumpeterian hypothesis is, broadly speaking, that oligopolistic market structures are best in this regard.

Both theory and empirical data give ambiguous evidence as to the market structure most conducive to R&D effort. Competitive markets can sometimes fail to be as innovative as their less-competitive counterparts but a surprising number of key inventions have come from small firms. Policy has a role to play here, too. One role for policy is to encourage cooperation in research efforts. Empirical evidence suggests that we live

in an increasingly interconnected world in which the benefits from one firm's R&D spill over to other firms including, in particular, its rivals. In such a world, the noncooperative outcome is likely to be one with too little R&D effort. Policy that fosters research cooperation among firms can be helpful in this setting. Yet caution is also necessary. The trick is somehow to foster cooperation on R&D without simultaneously inducing collaboration on prices and product design.

A similar tension arises in the role of patent policy. Patents can enhance the incentives for firms to pursue technological innovations. Yet, by temporarily granting monopoly power, patents can also weaken competitive forces and reduce consumers' access to those breakthroughs. We consider patents and related policy issues in the next chapter.

Problems

1. Assume that inverse demand is given by the linear function $P = A - BQ$ and that current marginal costs of production are c .
 - a. By how much would an innovation have to reduce marginal cost for it to be a drastic innovation?
 - b. Use your answer to derive a condition on the parameters A , B , and C that determines whether a drastic innovation is feasible. (Hint: costs cannot be negative.)
- For Problems 2 through 5 assume the following: Inverse demand is given by $P = 240 - Q$. The discount factor is 0.9. Marginal production costs are initially \$120:
2. Calculate the market equilibrium price, output, and profits (if any) on the assumption that the market is currently
 - a. monopolized;
 - b. a Bertrand duopoly;
 - c. a Cournot duopoly.
 3. Suppose that a research institute develops a new technology that reduces marginal costs to \$60.
 - a. Confirm that this is not a drastic innovation in either the Bertrand or Cournot cases.
 - b. Calculate the new market equilibrium price, output, and profits for the monopolist and each duopolist, given that in the duopoly case the innovation is made available to only one firm.
 - c. How much will the monopolist and duopolist each be willing to pay for the innovation?
 4. Now assume that there is a potential entrant in the monopolized case and that the research institute is considering offering the innovation to this firm as well as to the monopolist. How does this affect the amount that the incumbent monopolist will be willing to pay for the innovation?
 5. Now return to the duopoly case but assume that the research institute is considering whether it should actually sell the innovation to both firms. Will it wish to do so
 - a. in the Bertrand duopoly?
 - b. in the Cournot duopoly?
 6. Assume that annual inverse demand for a particular product is $P = 150 - Q$. The product is offered by a pair of Bertrand competitors, each with marginal costs of \$75. The discount factor is 0.9. What is the current equilibrium price and total surplus?
 7. Return to Problem 6. Assume now though that if R&D is conducted at rate x , it incurs one-off costs of $r(x) = 10x^2$ and reduces marginal costs to $(75 - x)$. Suppose that one firm decides to conduct R&D at rate $x = 10$. This research will be protected by a patent of T years.
 - a. What profit (ignoring the one-off costs of R&D) does the innovating firm make each year during the period of patent protection?
 - b. What is the new equilibrium price and total surplus once patent protection expires?
 8. Use your answers to 7(a) and 7(b) to write the total net surplus from the innovation as

- a function of the period of patent protection. Use a spreadsheet program to derive an approximation to the socially optimal period of patent protection.
9. How are your answers to 7(a) and (b) affected if the innovating firm conducts research at rate $x = 15$?
 10. Why might industries facing less elastic demand do proportionately more R&D [Dasgupta and Stiglitz (1980)]?

References

- Arrow, K., 1962. "Economic Welfare and the Allocation of Resources for Inventions," in R. Nelson, ed., *The Rate and Direction of Inventive Activity: Economic and Social Factors*. National Bureau of Economic Research. Princeton: Princeton University Press.
- Barro, R., and X. Sala-i-Martin. 1995. *Economic Growth*. New York: McGraw-Hill.
- Blundell, R., Griffith, R., and Van Reenen J., 1995. "Dynamic Count Data Models of Technological Innovation," *Economic Journal* 105 (March): 333–344.
- Cohen, W., and Levin, R. 1989. "Empirical Studies of Innovation and Market Structure," in R. Schmalensee and R. Willig, eds., *Handbook of Industrial Organization*, Vol. 2. Amsterdam: North-Holland, 1059–98.
- Cohen, W., and S. Klepper. 1996. "A Reprise of Size and R and D," *Economic Journal* 106 (July): 925–951.
- Dasgupta, P., and J. Stiglitz. 1980. "Industrial Structure and the Nature of Innovative Activity," *Economic Journal* 90 (January): 266–93.
- d'Aspremont, C., and A. Jacquemin. 1988. "Cooperative and Noncooperative R&D in Duopoly with Spillovers," *American Economic Review* 78 (September): 1133–7.
- Gayle, P., 2002. "Market Structure and Product Innovation," Working Paper, Department of Economics, Kansas State University.
- Geroski P., 1990. "Innovation, Technology Opportunity and Market Structure," *Oxford Economic Papers* 42 (July): 586–602.
- Gilbert, R. J. 2006. "Competition and Innovation," *Journal of Industrial Organization Education*: Vol. 1: Issue 1, Article 8. Available at: <http://www.bepress.com/jioe/vol1/iss1/8>.
- Gilbert, R. J., and D. M. G. Newbery. 1982. "Preemptive Patenting and the Persistence of Monopoly," *American Economic Review* 72 (June): 514–27.
- Kamien, M. I., E. Muller, and I. Zang. 1992. "Research Joint Ventures and R&D Cartels," *American Economic Review* 82 (December): 1293–306.
- Keller, W. 2002. "Geographic Localization of International Technology Transfer," *American Economic Review*, 92 (March): 120–42.
- Klepper, S. 2002. "Firm Survival and the Evolution of Oligopoly," *Rand Journal of Economics* 33 (Spring): 37–61.
- Levin, R., and P. Reiss. 1984. "Tests of a Schumpeterian Model of R and D and Market Structure," in Z. Griliches, ed., *R and D, Patents and Productivity*, Chicago: NBER University of Chicago Press.
- Levin, R., W. Cohen, and D. C. Mowery. 1985. "R and D Appropriability, Opportunity, and Market Structure: New Evidence on Some Schumpeterian Hypotheses," *American Economic Review, Papers and Proceedings* 75 (May): 20–4.
- . 1987. "Firm Size and R&D Intensity: A Reexamination," *Journal of Industrial Economics* 35 (June): 543–65.
- Levin, R., A. Klevorick, R. Nelson, and S. Winter. 1987. "Appropriating the Returns from Industrial Research and Development," *Brookings Papers on Economic Activity: Microeconomics* 2: 783–822.
- Lunn, J. 1986. "An Empirical Analysis of Process and Product Patenting: A Simultaneous Equation Framework," *Journal of Industrial Economics* 34 (February): 319–30.
- Markides, C., and P. Geroski. 2005. *Fast Second: How Smart Companies Bypass Radical Innovations to Enter and Dominate New Markets*, San Francisco: Jossey-Bass.
- Peters, T., and R. Waterman. 1982. *In Search of Excellence*. New York: Harper & Row.
- Porter, M. 1990. *The Competitive Advantage of Nations*, New York: The Free Press.

- Reinganum, J. 1983. "Uncertain Innovation and the Persistence of Monopoly." *American Economic Review*, 73 (September): 741–48.
- Romer, D. 2006. *Advanced Macroeconomics*, New York: McGraw-Hill/Irwin, 3rd ed.
- Scherer, F. M. 1965. "Firm Size, Market Structure, Opportunity and the Output of Patented Innovations," *American Economic Review* 55 (September): 1097–125.
- Scherer, F. M. 1967. "Market Structure and the Employment of Scientists and Engineers," *American Economic Review* 57 (June): 524–31.
- Schumpeter, J. A. 1942. *Capitalism, Socialism, and Democracy*, New York: Harper.
- Scott, J. T. 1990. "Purposeful Diversification of R&D and Technological Advancement," in A. Link, ed., *Advances in Applied Microeconomics*, Vol. 5. Greenwich, Conn., and London: JAI Press.
- Solow, R., 1956. "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics* 70 (February): 65–94.
- von Hippel, E., 1988. *The Sources of Innovation*, New York: Oxford University Press.

Appendix

EQUILIBRIUM NONCOOPERATIVE R&D EFFORT IN THE PRESENCE OF SPILLOVERS

Differentiation of the profit equation (20.14) with respect to research effort of firm i and Solving for x_i implies best response curves R_1 and R_2 for firm 1 and firm 2 of:

$$R_1 : x_1 = \frac{2(2 - \beta)[A - c + x_2(2\beta - 1)]}{[9B - 2(2 - \beta)^2]}. \quad (20.A1)$$

and

$$R_2 : x_2 = \frac{2(2 - \beta)[A - c + x_1(2\beta - 1)]}{[9B - 2(2 - \beta)^2]}. \quad (20.A2)$$

These slope upward if $\beta > 0.5$ and downward if $\beta < 0.5$. In any equilibrium $x_1 = x_2$ by symmetry. Substitution then yields the Nash equilibrium research intensity, output, and profit levels:

$$x_1^C = x_2^C = \frac{2(A - c)(2 - \beta)}{9B - 2(2 - \beta)(1 + \beta)} \quad (20.A3)$$

$$q_1^C = q_2^C = \frac{3(A - c)}{9B - 2(2 - \beta)(1 + \beta)} \quad (20.A4)$$

$$\pi_1^C = \pi_2^C = \frac{(A - c)^2[9B - 2(2 - \beta)^2]}{[9B - 2(2 - \beta)(1 + \beta)]^2} \quad (20.A5)$$

Equation (20.A3) is decreasing in β . Higher research spillovers decrease research intensity.

EQUILIBRIUM R&D EFFORT WITH R&D COOPERATION

With cooperation, each firm's optimal research intensity maximizes the sum of the two firms' profits, given noncooperative output determination in stage two. From equation (20.14) aggregate profit is:

$$\begin{aligned} \pi_1^C + \pi_2^C = & \frac{[A - c + x_1(2 - \beta) + x_2(2\beta - 1)]^2}{9B} - \frac{x_1^2}{2} \\ & + \frac{[A - c + x_2(2 - \beta) + x_1(2\beta - 1)]^2}{9B} - \frac{x_2^2}{2} \end{aligned} \quad (20.A6)$$

Differentiating and imposing symmetry, yields the first order condition:

$$\frac{2(1 + \beta)[A - c + x^{RC}(1 + \beta)] - 9Bx^{RC}}{9B} = 0 \quad (20.A7)$$

(where the superscript RC denotes "R&D cooperation"). The equilibrium R&D intensity then is:

$$x_1^{RC} = x_2^{RC} = \frac{2(A - c)(1 + \beta)}{9B - 2(1 + \beta)^2} \quad (20.A8)$$

This is increasing in β . Output and profit levels may be obtained by substitution into equations (22.13) and (22.14) and simplifying to yield:

$$q_1^{RC} = q_2^{RC} = \frac{3(A - c)}{9B - 2(1 + \beta)^2} \quad (20.A9)$$

$$\pi_1^{RC} = \pi_2^{RC} = \frac{9B(A - c)^2}{[9B - 2(1 + \beta)^2]^2} \quad (20.A10)$$

EQUILIBRIUM R&D EFFORT WITH A RESEARCH JOINT VENTURE (RJV)

An RJV agreement amounts to cooperation with 100 percent R&D spillovers, i.e., $\beta = 1$ in equations (20.A8)–(20.A10) above. Thus:

$$x_1^{RJV} = x_2^{RJV} = \frac{4(A - c)}{9B - 8} \quad (20.A11)$$

$$q_1^{RJV} = q_2^{RJV} = \frac{3(A - c)}{9B - 8} \quad (20.A12)$$

$$\pi_1^{RJV} = \pi_2^{RJV} = \frac{9(A - c)^2}{(9B - 8)^2} \quad (20.A13)$$

An RJV gives the highest per-firm profits and the lowest consumer prices of any arrangement.

Patents and Patent Policy

In 1769, an English inventor, Richard Arkwright, patented a spinning frame that would revolutionize the production of cotton cloth. Two years later, in 1771, Englishman James Hargreaves introduced another invention, the spinning jenny. With these inventions, Britain entered the Industrial Revolution. Equally important, the inventions allowed Arkwright and Hargreaves to establish a commanding position in the production of cloths and, more generally, textile products. This allowed the inventors to reap large profits and to sell at a high price in the American colonies even after these became independent states.

The British energetically protected their monopoly position. Westbound ships out of London were searched thoroughly to make sure that no passenger was a former Arkwright or Hargreaves employee or had a copy of the design plans for the Arkwright–Hargreaves machines that firms outside of Britain might copy. Such restrictions along with the high textile price for British textiles vexed many Americans. Consumers did not like paying the monopoly prices and firms were eager to get some version of the machines that would permit them to compete with the British producers. Some firms offered “bounties” for English apprentices who would be able to obtain the necessary information. Finally, in 1789, an enterprising young Englishman and former Arkwright partner, Samuel Slater, responded to just such a bounty offer. After completely memorizing the engineering details of the Arkwright–Hargreaves machines, he disguised himself as a common laborer and set sail for America. Shortly thereafter, Slater arrived in Pawtucket, Massachusetts and established the first of many New England textile mills consolidating the region’s manufacturing base and finally breaking the British monopoly.

The issues raised by Slater’s entrepreneurship (what some might call theft) lie at the heart of this chapter. The central question is how strongly a firm’s innovation should be protected from imitative competition. On the one hand, information about an innovation is a public good so that once the information is produced, efficiency requires that access to this information, i.e., new production techniques and new products, should be unrestricted to prevent the rise of monopoly. On the other hand, if the government does not protect innovators against imitation, there may be little incentive to do the hard work that led to the invention in the first place.

The patent system was designed to create incentives for innovative activity. Patents and copyrights confer ownership to new inventions, new designs, and new creative works. In turn, those property rights permit innovators to restrict the use of their ideas just as the British restricted the flow of information on their textile technology. The patent holder can act as

a monopolist regarding its discovery and earn a monopoly profit as a result. Yet while that profit may create an incentive to undertake R&D efforts, the monopoly that generates the profit reduces the total surplus below what it could be given that the invention has occurred.

Getting this balance right is not easy. We can imagine just how much less productive the economy would be if the science behind electric lighting, the aerodynamics of airplanes, and semiconductors had never been developed. Or how much less healthy we would be without drugs to lower cholesterol and combat polio and malaria. However, society would also be less productive and less healthy were those same technologies not now widely available to all firms and consumers. At some point, policy must shift from a stance of protecting innovators from imitation to one of permitting the use of the innovation on as wide a basis as possible. The sixty-four-million-dollar question is, exactly where does that point arise? When has protection of the innovator extended sufficiently far that we ought to start thinking about protection of consumers?

The issue as to how far patent rights should extend has two dimensions. First, what is the length of time for which any patent rights ought to last? Second, to what range of products should the patent apply? Should the developer of a new AIDS treatment based on a special combination of protease inhibitors be protected against a rival's later development of an alternative AIDS treatment based on a different combination of protease inhibitors? What about a new AIDS treatment that is not based on protease inhibitors? Or what if a protease inhibitor treatment originally created as a treatment for AIDS is now applied as a treatment for multiple sclerosis? These issues—typically referred to as patent length and patent breadth—are the central questions in patent policy.

21.1 OPTIMAL PATENT LENGTH

Current patent law establishes a patent duration that varies from country to country. In the United Kingdom and the United States, patent law grants protection for twenty years from the date of filing the application. In both countries it is up to the patent holder to ensure that the patent is renewed during its life and to ensure that the patent is not infringed.

Economic theory can provide some insights as to whether that duration makes sense. The key is to find a balance between the innovator's ability to earn a return on its R&D investment and the benefits that accrue to consumers once the patent expires and competition emerges. The basic model, which is due to Nordhaus (1969), is presented below.

Imagine a competitive industry in which each firm is pursuing a nondrastic innovation. Innovative efforts incur costs. Each firm's unit operating cost is currently c . If a firm invests in R&D at some intensity x , it expects to reduce its unit operating costs from c to $c - x$. The cost of undertaking R&D at intensity x is $r(x)$. We assume that such costs rise as the level of research intensity increases and that they do so at an increasing rate. Formally, this means that $dr(x)/dx > 0$ and $d^2r(x)/dx^2 > 0$. Thus, R&D is expensive to do and exhibits decreasing returns in that a doubling of research intensity will give less than double the reduction in operating costs.

Our assumption of a competitive market implies that price equals marginal cost, which means that the initial market price is c and that the output level is Q_0^C . This is shown in Figure 21.1. A successful innovator will be able either to produce at the lower unit cost of $c - x$ and drive out all its rivals by setting a price just one penny less than the current price, or to license its discovery to its competitors for a fee of $c - x$ per unit produced. Either way, the current market price and volume of output remain unchanged. The innovator,

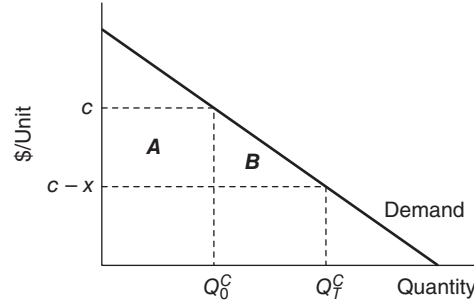


Figure 21.1 Innovation gains during period of patent protection (T years) and after patent protection

The innovator receives profit of area A for the T years that the patent is in effect. When the patent expires, competition lowers the price to $c - x$. Consumers gain the former profits A and also area B as consumer surplus. However, only the area B is a net increase in welfare.

however, earns a profit equal to area A in Figure 21.1. Assuming that the life of the innovator's patent is T years, this profit will last for T years as well.

When the patent expires, all firms will have access to the technology for free. Competition will reduce the price to $c - x$, and output will expand to Q_T^C . The profit that the innovator used to earn becomes consumer surplus. This is simply a transfer from a producer to consumers and so does not reflect a net gain. However, the expansion of industry output to the higher level Q_T^C does bring such a net benefit by virtue of the additional consumer surplus this generates. This additional surplus is shown in Figure 21.1 as area B .

The longer the duration of the patent (the higher is T), the longer is the time over which the innovator earns the profit A and the greater is the innovator's incentive to do costly R&D. Denote the per period profit flow to the innovator (area A in Figure 21.1) as $\pi^m(x; T)$ and the discount factor as R . The present value of the innovator's profit from R&D is¹

$$V_i(x; T) = \sum_{t=0}^{T-1} R^t \pi^m(x; T) = \frac{1 - R^T}{1 - R} \pi^m(x; T) \quad (21.1)$$

Therefore, the R&D has a net value to the innovator of

$$V_i(x; T) - r(x) \quad (21.2)$$

For a given value of T chosen by the patent office, the innovator selects a level of R&D activity, $x^*(T)$, that maximizes this expression. This choice just balances the marginal gain of additional discounted profit against the marginal cost of doing more R&D work.

Of course, a rational patent office recognizes that its choice of patent life T affects the firm's choice of R&D effort. We suppose that the patent office can work out this relationship precisely. In other words, the patent office can determine the innovator's profit-maximizing research intensity, $x^*(T)$, as a function of T . To choose T optimally, the patent office

¹ This result uses the following equation in calculating discounted value. Assume that a sum A is to be received each period for T periods, and recall from Chapter 2 that $R = (1 + r)^{-1}$ where r is the interest rate. Then the discounted value of these cash flows is $S = A + RA + R^2A + R^3A + \dots + R^{T-1}A = A(1 + R + R^2 + \dots + R^{T-1}) = A(1 - R^T)/(1 - R)$.

wishes to pick the patent duration that maximizes the net social gain to both consumers and producers given how firms choose their research intensities. Let us denote by $cs(x; T)$, the per-period increase in social surplus that the innovation generates once it becomes freely available, which as we have seen is the area $A + B$ in Figure 21.1. The present value of this increase in surplus is then

$$CS(x; T) = \sum_{t=T}^{\infty} R^t cs(x; T) = \frac{R^T}{1-R} cs(x; T) \quad (21.3)$$

The total net social surplus from the innovation is

$$NS(x^*(T); T) = V_i(x^*(T); T) + CS(x^*(T); T) - r(x^*(T)) \quad (21.4)$$

and the objective of the patent office is to choose the patent duration T^* that maximizes this net surplus. This is a complicated expression but we can develop an intuitive argument to support a very important proposition, namely, that *the optimal patent duration is finite*.

To see why, note that as the patent office initially increases patent duration it induces greater R&D effort and, at first, a greater discounted net surplus to producers and consumers. If patent duration is zero, the returns to an innovator are also zero because the results of the innovation will be imitated immediately. Accordingly, there is no R&D and no change in the social surplus. If we now increase the patent length to a value $T > 0$, we induce some innovation and, thereby, some increase in the total surplus. Beyond some point, however, continued increases in T reduce net social surplus even though they lead to more R&D and therefore greater reductions in production cost. Two forces work to limit the optimal value of T . The first is our assumption of diminishing returns to R&D activity. Because it becomes progressively more expensive to lower production costs, it takes progressively greater increases in T to achieve a given additional cost saving. The second force limiting optimal patent duration is the fact of discounting. The consumer benefits shown as area B in Figure 21.1 are not realized until after the patent expires. If the patent office chooses a very long duration time T the present value of those benefits is very small indeed.

This is particularly important because it has sometimes been argued that innovation should be granted patent protection forever. Such a long patent duration puts far too heavy a value on the monopoly profits that patent protection generates and gives too little consideration to the additional consumer surplus that emerges only after the patent protection has expired.²

Let the inverse demand function for a particular product be $P = 100 - Q$, and let it be provided by a group of competitive firms, each with an identical marginal (and average) cost of \$70 per unit.

- Show that the current market output and price are, respectively, $Q = 30$ and $P = \$70$.
- Imagine that one firm can conduct R&D at a pace x , at a cost of $r(x) = 15x^2$. Let the interest rate, r , be 10 percent so that the discount factor, R , is 0.9091. Show that a patent length of 25 years will induce the firm to pursue R&D at a level of approximately

² Author Mark Helprin has similarly argued for an infinite copyright for creative works: "A Great Idea Lives Forever, Shouldn't Its Copyright?" *New York Times* 20 May 2007. Note, the argument for an infinite patent life is moot if there is continual innovation that effectively limits the economic life of any one patent.

$x = 10$. Note that if $x = 10$, the firm's research activity will reduce the unit cost from \$70 to \$60.

- c. Would the firm's R&D effort increase or decrease if patent duration was reduced to 20 years?
- d. Would total social welfare increase or decrease if patent duration was reduced to 20 years?

21.2 OPTIMAL PATENT BREADTH

The question of the optimal patent breadth is trickier than that of patent length, mainly because there is no universally accepted measure of breadth comparable to time as a measure of duration. Conceptually, the idea is to set a minimum amount by which a new innovation must differ from an existing process (or good) in order for the new one either to avoid infringement on an existing patent or to be itself patentable. The larger this required minimal degree of difference, the more difficult it is for other firms to "invent around" the patent and to cut into the inventor's profit. We could in principle work out the optimal patent breadth just as we worked out the optimal patent length. But the lack of a clear method for measuring breadth makes implementing this plan very difficult.

This lack of precision is reflected in the language of the patent office. Each application for a patent is required to specify all the "related" existing patents and to indicate not only how the patent being applied for is a discovery distinct from those already patented, but also to show that the discovery is "novel, nonobvious, useful." Such language leaves the patent office a lot of discretion regarding how it rules in any particular case.

What makes the question of the optimal patent breadth even more difficult is that it cannot be divorced from the question of optimal duration. Patent policy must set both dimensions of patent protection. Typically, this amounts to choosing between a system in which patents should have a short duration but a broad coverage, the "short and fat" approach, or a long duration combined with a very narrow coverage, the "long and thin" solution. As always, these choices involve balancing the need to maintain the incentive to innovate against the need to distribute the benefits of innovation as widely as possible.

The complications that arise when a second dimension of breadth is introduced into patent policy can be seen in a number of formal models. Consider for instance the analysis of Gilbert and Shapiro (1990). They define patent breadth in terms of the extent to which the patent-holder can charge a price above marginal cost. Broader patents decrease consumer substitute options and permit a higher price-cost margin. This margin of course is the source of the patent-holder's profit. Suppose that we know the desired innovative effort level x , and hence the cost $r(x)$ necessary to achieve that effort. The trick then is to do this with a patent design that produces the necessary (discounted) profit at the lowest possible social cost. That is, we may frame the objective as choosing patent breadth and length so as to minimize the deadweight loss per unit of innovator profit subject to that profit level being sufficient to undertake the desired inventive activity.

Given the social objective and their definition of patent breadth, Gilbert and Shapiro (1990) then demonstrate that the optimal patent is to have very narrow but infinitely long patents. Why? The underlying intuition is as follows. If we think of time as a sequence of equally long intervals, then each interval may be thought of as a separate market. A standard condition for welfare maximization is that it should not be possible to raise welfare by shifting production from one market to another, i.e., the net marginal value of an extra

unit should be the same in each market or, in our case, in each period. A finite patent does not typically satisfy this condition. During the patent life, the price is high due to the patent-holder's monopoly power. Once the patent has expired, however, the price falls to marginal cost. The only way to avoid this discontinuity is to keep the price above marginal cost in all markets, i.e., for all time periods into the infinite future, while limiting the accompanying distortion that this brings by restricting the patent's breadth so that price is just enough above cost permanently that the necessary profit level is achieved. In other words, optimal patent policy induces small but equal price distortion for many periods rather than a few large distortions in some periods and none in others.

However, the Gilbert and Shapiro (1990) approach is not the only way to model patent breadth. Klemperer (1990) for example offers an alternative approach that relates patent breadth more directly to product differentiation. If we think of a Hotelling line segment of finite length, Klemperer's view is that a useful definition of patent breadth is the fraction of the line segment that is covered by the patent. Again assuming that the goal is to minimize the ratio of social cost to innovator profit subject to covering desired innovation costs, Klemperer (1990) then shows that there are cases in which optimal patent design is just the opposite of that implied by the Gilbert and Shapiro analysis. That is, it is often the case that the best patent design is one of broad patents that are short-lived.

To understand Klemperer's (1990) argument, consider a simple example. Suppose that the new good costs nothing to produce (all the costs are sunk design costs) and that there are ten potential customers for the product. Assume further that each customer values the good at \$10. If there were no other substitutes available, the monopolist firm would then simply set a price of \$10, sell one unit to each of the ten consumers, and claim all of the \$100 surplus from the market. There would be no consumer surplus but also no welfare loss. Total output would also have been ten had the monopolist priced at marginal cost ($= 0$).

Now suppose that the transport cost of buying an alternative legal substitute is different for each of the ten consumers. Specifically, let one consumer incur a transport cost of \$1 per unit of distance the substitute is from the product; a second incur a transport cost of \$2 per unit; and so on. In this setting, patent breadth w is interpreted as how far consumers have to travel to obtain a legal alternative brand. A very wide breadth or high value for w effectively puts one back in the setting of no alternatives. Hence, if w is very broad, the market outcome will again be a price of \$10, and there will be no deadweight loss. Now, however, consider what happens if we limit the patent width to $w = 1$ (or just a bit less). In this case, at any price $p \geq \$1$, the patent holder loses some customers. At a price of \$1, the patent holder loses one client. At a price of \$2, a second is lost, so on. The best option (assuming whole-dollar prices) is then to set a price of \$5, in which case the patent holder sells five units and earns profit of \$25. Now there is a deadweight loss. Real resources are being used to produce the less desired substitutes into which consumers are shifting. Consumers are incurring transport costs, as well. In other words, the narrower is the patent the greater is the deadweight loss, implying the need for broad patents. Patent length should then be set at the minimum necessary to achieve the desired innovative expenditure.

As noted, Klemperer's (1990) analysis also yields conditions under which a narrower but longer-lived patent is preferred. This occurs for example when, unlike the case above, it is the transport cost that is the same for all and the valuation that varies across consumers. However, the crucial result is that when consumer variation primarily reflects differences in transport cost or strength of preference for the brand of the new good and not in the basic valuation of that good, broad patents of relatively short derivation are preferred.

Gallini (1992) provides a further reason why short-lived broad patents may be best. She makes the point that imitators can often get around patent protection if they spend enough money to imitate the product without infringement. They will be particularly encouraged to do so when patents are long because otherwise, entry into the market will be greatly delayed. When patents are short, imitation is less attractive because firms now find it cheaper simply to wait for the patent to expire than to engage in costly efforts to imitate legally now. In other words, Gallini (1992, 2002) makes the important point that costly imitation efforts also need to be accounted for in considering any welfare effects. If these imitation costs are sizeable, then broad but short-lived patents are preferable.

Denicolò (1996) synthesizes many of these features in a framework that also incorporates the extent of market competition. He finds that “Loosely speaking, the less efficient is the type of competition prevailing in the product market, the more likely it is that broad and short patents are socially optimal” (264). By “efficient,” Denicolò means roughly the extent to which competition drives firms close to the competitive ideal. Denicolò’s statement implies that markets in which firms have a greater degree of monopoly power will do best with the “short and fat” approach, while markets characterized by a good bit of competition will do best with patents that are “long and thin.”

As a policy recommendation, a major drawback of Denicolò’s proposal is that it seems to suggest applying different standards to different innovators depending on the structure of the innovator’s basic industry. In reality, the rule of law cannot be applied so selectively without risking serious inconsistency. Even apart from that consideration, there is a further difficulty in implementing any of the proposed standards. The problem again is that it is not always easy to make the concept of patent breadth operational. We do not have an easy way to translate real markets into a spatial representation and no obvious measure of distance. Indeed, as Scotchmer (2004) has noted, Klemperer’s (1990) horizontal concept of breadth is itself too limiting. There is also a vertical component reflecting how much better (or how much worse) a rival’s product has to be before it does not infringe on the patented good. Recognizing this second dimension of patent breadth makes its measurement all the more difficult from a practical perspective. While it is probably fair to say that we will not go too far wrong if we adopt a one-size-fits-all policy of granting patents with “reasonable” breadth but constrained length, precisely what this means in practice is a lot less clear.

21.3 PATENT RACES

Our discussion of market structure and innovative activity was largely motivated by Schumpeter’s observation that innovation is a crucial and different kind of competition. The Schumpeter vision is one in which firms vie with each other by racing to develop new technologies or new goods and in which this sort of rivalry is potentially deadly for those who come up short. This is particularly true when innovations are eligible for patent protection. With patents, coming in first is all that matters whether one wins by several lengths or by just a nose. The first firm to discover a cure for male baldness or to engineer a successful system for producing “talking” pictures leaps far ahead of its rivals and stays there for some time by virtue of patent or copyright protection. Patent awards have a “winner-take-all” feature so that finishing second is no better than finishing third or fourth or, for that matter, tenth.

Innovative competition can be regarded as a race in which one player’s success is the other player’s serious defeat. The loser of a patent race may see years of investment and

hard work wiped out overnight when the rival announces its breakthrough. We now turn to some of the issues that arise when we consider the implications of a patent system that generates races in which finishing first is all that matters. What are the consequences of such races? Do they lead to inefficient investment in R&D? Does the innovative activity generated by the race influence market structure?

Consider a patent race between two firms that can choose to invest in research with a view to developing a new product. The first to make the breakthrough wins the race and files a patent giving that firm exclusive rights to its invention. This is what gives the race its winner-take-all aspect. The loser walks away less than empty-handed, having expended resources on R&D with no return.

Let there be two firms, BMI and ECN, who both are considering doing the R&D necessary to create a new product. They each estimate that if the innovation is successful they can produce this new product at a marginal cost of c and that demand for the new good is $P = A - BQ$. They are also confident that the new product is a sufficiently radical departure that it will have a negligible impact on their existing businesses and so will not affect their existing profits—that is, there is no replacement effect.

The R&D effort by each firm requires establishment of a research division that costs a fixed sum, K . This sum covers both the costs of research and of development if the research is successful and, once sunk, can never be recaptured. Given that such a division is established, the probability of a successful innovation is ρ . If only one firm is successful in its R&D efforts, we assume that the innovation is protected from imitation, perhaps by a patent or by some other means so that the successful innovator earns the monopoly profit. If both are successful simultaneously, both firms file a patent application and we assume that each firm has a 50 percent probability of its patent application being successful, in which case this firm earns the monopoly profit. To keep matters reasonably simple, we assume that both firms discount the future heavily—that is, the interest rate r is so large that the discount factor $R \approx 0$.

In order to identify the incentives each firm has to establish the research division, we need to identify their profits with and without a successful innovation. If neither firm attempts to develop the new product, neither firm enters this new market. As a result, each earns zero profit in this new market.

If both firms undertake R&D but only one firm, say BMI, is successful in its R&D efforts while ECN is not, then BMI will be a monopolist in the new product market earning the monopoly profits. ECN will earn nothing from the new market. The profits of the two firms in this case, again ignoring the costs of establishing the R&D division, become

$$\pi_b = \frac{(A - c)^2}{4B} ; \pi_e = 0 \quad (21.5)$$

Of course, if ECN is successful but BMI is not, these profits are reversed.

Conversely, if both firms undertake R&D and both are successful in making the innovation, each has a 50 percent probability of being awarded a patent, so that expected profit, ignoring the costs of establishing the R&D division, is

$$\pi_b = \pi_e = \frac{(A - c)^2}{8B} \quad (21.6)$$

We can now calculate the expected profit for each firm depending on whether or not it establishes a research division. If neither firm sets up such a division, neither innovates and each earns zero profit in this new market. Now consider the expected profit if only one

firm, say BMI, establishes an R&D division. For BMI expected profit is made up of two components:

1. profit if the R&D division is unsuccessful, which is zero and occurs with probability $(1 - \rho)$;
2. profit if the R&D division is successful, which is the monopoly profit $(A - c)^2/4B$ and occurs with probability ρ .

As a result, the expected *net* profit of BMI if it is the only firm to establish an R&D division is

$$\pi_b = \rho \frac{(A - c)^2}{4B} - K \quad (21.7)$$

Of course, the expected profit of ECN, given that only BMI has established an R&D division, is zero. By symmetry, we reverse these payoffs to get expected profits if ECN is the only firm to establish a research division.

If both firms establish R&D divisions, the expected profit to either firm is given by

1. profit if the firm's R&D division is successful and the rival's is not, which is $(A - c)^2/4B$, and occurs with probability $\rho(1 - \rho)$;
2. profit if both R&D divisions are successful and the firm wins the patent application, which is $(A - c)^2/8B$, and occurs with probability ρ^2 .

Of course if neither firm is successful in R&D they earn nothing from the new market. This means that the expected *net* profit of each firm, given that they both operate R&D divisions, is

$$\pi_b = \pi_e = \rho(1 - \rho) \frac{(A - c)^2}{4B} + \rho^2 \frac{(A - c)^2}{8B} - K = \frac{(A - c)^2}{8B} \rho(2 - \rho) - K \quad (21.8)$$

Before we put these payoffs into a payoff matrix, we can do a bit of simplifying. The profit equations share a common expression, the monopoly profit, which we denote as $M = (A - c)^2/4B$. We can use this to define a parameter $S = K/M$, which is the share of the monopoly profits that are needed to establish the R&D division. With the substitution of S and M , the expected profits are summarized in the payoff matrix of Table 21.1 (below). This matrix allows us to identify the possible Nash equilibria for this R&D game. As we shall see, these are dependent upon the relative magnitudes of the two parameters, S and ρ .

There are three possibilities that have to be considered:

1. *Neither Firm Wishes to Establish an R&D Division.* For this to be a Nash equilibrium, the payoff to BMI, for example, from not having an R&D division, given that ECN also has no R&D division, must be greater than the expected profit from investing in R&D,

Table 21.1 Payoff matrix for the duopoly patent race

		<i>BMI</i>	
		<i>No R&D Division</i>	<i>R&D Division</i>
<i>ECN</i>	<i>No R&D Division</i>	0, 0	0, $M(\rho - S)$
	<i>R&D Division</i>	$M(\rho - S)$, 0	$M\left(\frac{\rho(2 - \rho)}{2} - S\right)$, $M\left(\frac{\rho(2 - \rho)}{2} - S\right)$

again given that ECN does not. In other words, BMI expects to make more profit from the strategy combination (No R&D, No R&D) than from the combination (No R&D, R&D). This requires that $M(\rho - S) < 0$, which implies that $S > \rho$, the probability of success is less than the fraction of monopoly profit required to fund the R&D. This expression is illustrated by the line OA in Figure 21.2. All parameter combinations above OA give the Nash equilibrium (No R&D, No R&D).

2. *Only One Firm Wishes to Establish an R&D Division.* Assume that the firm that establishes the R&D division is BMI. Then for the strategy (No R&D, R&D) to be a Nash equilibrium, two conditions must be satisfied:
 - a. BMI expects its expenditure on R&D to be profitable, given that ECN is not investing in R&D—that is, BMI expects to make more profit from the strategy combination (No R&D, R&D) than from the strategy combination (No R&D, No R&D). This is just the opposite of the expression derived in part 1. It requires that $S < \rho$.
 - b. ECN does not expect its expenditure on R&D to be profitable, given that BMI is investing in R&D—that is, ECN prefers the strategy combination (No R&D, R&D) to (R&D, R&D). For this to be the case, the following must be true: $M\left(\frac{\rho(2-\rho)}{2} - S\right) < 0$ which requires that $S > \frac{\rho(2-\rho)}{2}$. This relationship is illustrated by the curve OB in Figure 21.2. All parameter combinations that lie between OA and OB are such that only one of the firms will establish an R&D division.
3. *Both Firms Wish to Establish an R&D Division.* For this to be a Nash equilibrium, the payoff to, for example, BMI from having an R&D division, given that ECN also has an R&D division, must be greater than the expected profit from not investing in R&D, again given that ECN does. In other words, BMI expects to make more profit from the strategy combination (R&D, R&D) than from the strategy combination (R&D, No R&D). For this to be the case we must have that

$$M\left(\frac{\rho(2-\rho)}{2} - S\right) > 0 \text{ which requires that } S < \frac{\rho(2-\rho)}{2}$$

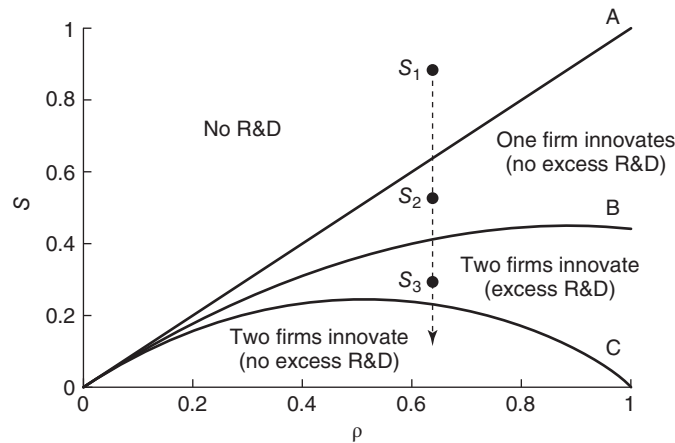


Figure 21.2 A patent race with a duopoly

Of course, exactly the same condition guarantees that ECN prefers the strategy combination (R&D, R&D) to the strategy combination (No R&D, R&D). Thus, all parameter combinations below 0B are such that both firms will establish an R&D division.

One question about patent races is whether the potential profit from successful innovation can lead the two firms to overinvest in R&D. Neither of the firms establishes an R&D division unless this division is expected to be profitable. For the strategies (R&D, No R&D), (No R&D, R&D), and (R&D, R&D) to be equilibria, they must each give positive expected profits to the two firms. This tells us that no equilibrium in which only one firm invests in R&D is characterized by “excessive” R&D in the sense that the firms would be better off without the R&D. The question that is left is whether there is “too much” R&D when both firms establish R&D divisions. Are there situations in which the strategy combination (R&D, R&D) is a Nash equilibrium but generates less aggregate profit than the strategy combinations (R&D, No R&D) or (No R&D, R&D)? For this to be the case, it must be that

$$2M\left(\frac{\rho(2-\rho)}{2} - S\right) < M(\rho - S) \text{ which requires that } S > \rho(1 - \rho)$$

This is illustrated by the curve 0C in Figure 21.2. All parameter combinations between 0C and 0B lead to excessive R&D as the two firms race to be first to discover and introduce the new product.

Our simple model delineates three distinct possibilities. First, neither firm will invest in R&D unless it is expected to be profitable. Hence R&D must have a reasonably low cost relative to the monopoly profits that it might generate (low S), or a reasonably high probability of success. Second, for any given probability of success, a larger number of firms will establish R&D facilities when there is a lower cost of R&D relative to the profit the innovation is expected to generate. Thus, for any given probability of success ρ , the equilibrium number of firms with R&D divisions increases from zero to one and finally to two as S is reduced. Third, there is an intermediate range of values for the cost of R&D in which there is excessive R&D in that both firms establish R&D divisions although this reduces their aggregate profits. In this range, the lure of profit from innovation involves the firms in a competitive R&D race that they would be better to avoid.

So far, we have only considered the gain that research brings in terms of the expected profit of the two firms. From a public policy perspective, however, increased profit is not the only potential benefit of innovation. We should also consider the gain in consumer surplus that development of this new product will generate. While we have just shown that the level of R&D activity can be excessive from the viewpoint of the firms’ combined profits, we have not demonstrated that this is the case when viewed with the objective of maximizing the total gain of profit plus consumer surplus. The R&D which seems excessive to the firms may still be worthwhile to society overall if the additional consumer surplus more than offsets the reduction in aggregate profit. However, R&D can be excessive even when evaluated with this broader criterion, (see Practice Problem 21.2). The patent race can lead both firms to establish research divisions even when the total cost of such divisions is not justified by the sum of expected producer and consumer surplus.

Even more interesting is that we can easily show that the possibility of too little R&D—as judged from a social welfare perspective—is quite real. Suppose that $S > \rho$, in which case neither firm undertakes R&D. Suppose further that S is so close to ρ that one firm could almost expect to break even if it pursued the innovation and its rival did not. If the new product generates any significant consumer surplus at all, then it is socially desirable that the research takes place. The value of the expected consumer surplus more than provides