

On the division of profit in sequential innovation

Jerry R. Green*

and

Suzanne Scotchmer**

In markets with sequential innovation, inventors of derivative improvements might undermine the profit of initial innovators through competition. Profit erosion can be mitigated by broadening the first innovator's patent protection and/or by permitting cooperative agreements between initial innovators and later innovators. We investigate the policy that is most effective at ensuring the first innovator earns a large share of profit from the second-generation products it facilitates. In general, not all the profit can be transferred to the first innovator, and therefore patents should last longer when a sequence of innovations is undertaken by different firms rather than being concentrated in one firm.

1. Introduction

■ Knowledge and technical progress are cumulative in the sense that products are often the result of several steps of invention, modification, and improvement. Indeed, the "development" aspect of "research and development" can be as commercially important as the "research." But when innovation occurs in two stages, the first innovator may have insufficient incentive to invest. First, competition from improved products could undermine the original innovator's profit. Profit erosion could be so severe that the original research could be unprofitable, and that could stymie the entire line of technology. But even when derivative products are not direct competitors to the first product—e.g., when they are "applications" of the first product that serve unrelated markets—the first innovator might have deficient incentive to invest. This is because the social value of an early innovation includes the net social value of the applications it facilitates. If the first innovator does not collect that value as profit, he might not invest even if the combined profit of the innovations exceeds the combined costs.

* Harvard University.

** University of California, Berkeley.

A previous version of this article circulated under the title "Antitrust Policy, the Breadth of Patent Protection, and the Incentive to Develop New Products." We thank the National Science Foundation, grant nos. SES-88-09107 and SES-89-09503, for financial support. We thank Nancy Gallini for very useful comments that much improved this article. We also thank Joseph Farrell, Neil Gandal, two anonymous referees, conference participants at the C.E.P.R. conference on Industrial Organization held in May 1992 in Barcelona, and seminar participants in Toulouse.

Incentives to innovate are protected overall by granting a sufficient patent life. Nordhaus (1969) argued that patent lives should be finite, even though some R&D might be deterred, in order to reduce monopoly distortions on average. In this article we are concerned not only with the *total* profit earned by the innovators, but also with the *division* of profit. For example, it is useless to ensure that a sequence of two innovations is jointly profitable if all the profit goes to the second innovator and the first innovator cannot cover his costs. In that case neither the first nor second innovation would be invented.

In Section 2 we specify a model that we use to investigate the division of profit. For simplicity we assume that after the first innovation is made, the idea for each derivative improvement occurs to only one firm, which is uniquely capable of developing it at a cost. The first innovator can profit by selling his product in the market, by licensing his patent to second-generation innovators after their innovations are achieved, or by sharing profit in *ex ante* agreements to develop second-generation products. At the end of the article we comment on whether our conclusions are robust to these assumptions.

Using this model we investigate the complementary roles of patent length and breadth. The breadth determines how profit is divided in each period of the patent, and the length determines the total profit that is collected by the firms jointly. The division of profit in each period depends on whether the second product infringes the first patent. If a second product infringes, the second innovator must license, which transfers profit from the second innovator to the first. Because the breadth of the first patent determines whether a product infringes, it thus determines the division of profit. If the division of profit is too unfavorable to one innovator or the other, the patent must last a long time to ensure that the less-favored innovator covers his costs. Our object is to investigate how the patent breadth should be chosen so as to keep the patent life as short as possible.

In Section 3 we make a general point about patent life that is independent of the breadth: the patent should last longer when R&D takes place in several firms than if it is concentrated in one firm. If R&D is concentrated in one firm, the patent should last just long enough to cover total costs. But with two firms, such a short patent life would place a severe demand on the division of profit: each researcher would have to earn exactly enough revenue to cover his own costs. We show that this will typically not happen, and in fact the second innovator will make positive profit whatever the patent breadth. Therefore, the incentive to undertake basic research will inevitably be too weak if the patent life provides only zero total profit.

Nevertheless, our objective is to make the patent life as short as possible. This can be done by finding ways to transfer profit from the second innovator to the first. In Section 4 we ask how the patent breadth should be chosen in order to do this. We give circumstances in which the first patent should be very broad so that all second-generation products infringe, and other circumstances in which this is not the best policy.

Our conclusions about optimal patent policy rely on certain assumptions about what types of licensing agreements the firms can make. Licensing agreements can occur at two stages: *ex ante*, which is before the second innovator invests in the improvement or application, or *ex post*, which is after the improvement or application is achieved. The difference is in whether the second innovator has sunk his costs at the time of the negotiation. We stress in this article that *ex ante* licensing has an important advantage over *ex post* licensing: it can ensure that the innovators invest in the improvement or application if and only if it would increase their joint profit.¹ *Ex ante* agreements may have a second advantage for the firms as well: without an *ex ante* agreement, the firms might become *ex post* competitors.

¹ Thus we study the mirror image of the problem studied by Gallini (1984), who showed how *ex ante* licensing can prevent a competitor from inefficiently duplicating an invention. In both cases *ex ante* licensing can remedy an investment decision that would decrease the joint profit of the two firms. See also Spence (1984).

The question of what kinds of licensing agreements should be allowed is only slightly murkier than the question of what types are in fact allowed under current antitrust law. Although we think our maintained hypotheses in Sections 2 to 4 are reasonable, we point out in Section 5 that they are not the only ones imaginable. In Section 5 we turn our attention to optimal design of antitrust policy, again with a view toward trying to understand what policies divide profit in such a way that the patent life can be relatively short. In Section 6 we present our conclusions.

2. The model

■ We call the quality of the first product x , and the quality of the second-generation product $x + y$, so that y represents the size of the improvement. These qualities are related to consumers' willingness to pay. We assume that if the first product alone is marketed as a monopoly, the monopolist earns revenue $\pi_x(T)$ when the patent lasts a period T , and if the improved product is marketed by a monopolist, the revenue is $\pi_{x+y}(T)$. If the two products compete, the revenues are respectively $\pi_x^c(T)$ and $\pi_y^c(T)$, and we assume (reasonably) that $\pi_x^c(T) \leq \pi_x(T)$ and $\pi_{x+y}(T) \geq \pi_x^c(T) + \pi_y^c(T)$. These profits increase with T . To reduce notation, we often omit the argument T .

A special case used in Propositions 3 and 5 below is when

$$\pi_{x+y} = \pi_x + \pi_y^c = \pi_x^c + \pi_y^c.$$

This circumstance can arise in at least two ways. In the first, the entire commercial value of the basic research resides in second-generation products, as when each second-generation product is an "application." Then it is reasonable to assume that $\pi_{x+y} = \pi_y^c$ and $\pi_x = \pi_x^c = 0$. For example, the first innovation might be laser technology, which has no direct value to consumers, and an application might be laser surgery. The two innovations are linked only through the fact that the first one facilitated invention of the second one. In the second circumstance, the two innovations are not close substitutes because they serve different markets. For example, a surgical device for humans might generate, as a spin-off, a surgical device for pets.

After the first product is patented, a firm called firm 2 gets an "idea" for an improved product, and this idea is described by (y, c_2) drawn from distributions (G, H) known to everyone prior to the first investment. The investment in the first product (with quality x) must be made before the "idea" for the second product (with quality $x + y$) or the identity of the second firm is known. Although both y and c_2 are unknown at the time of the first investment, some of the uncertainty will typically be resolved prior to the second investment. Below we consider two possibilities: (i) that all the uncertainty is resolved, so that the prospective second innovator knows both the value and cost of the improvement when he makes his investment decision, and (ii) that only the cost is resolved, but the value is still uncertain.

For simplicity we will assume that little time elapses between the innovations, so that both patents begin and end at the same time. Relaxing this assumption would complicate the model without yielding substantially different conclusions. The firms will decide whether to invest in the idea according to a game described below. First, however, we discuss what would be efficient.

□ **Efficiency.** The social value of entry by firm 1 has two parts. First, there is the social value of the first innovation of quality x , which might be marketed with no further improvements, and second, there is the option value of a potential application or improvement y by firm 2. Even if the social value of the first product alone does not exceed its costs, the possibility of a valuable second-generation product(s) might justify the investment. However, the social value that accrues from the two products depends on the market

structure in which they are marketed. The social surplus (profit plus consumers' surplus) is different if the products are sold by one monopolist than if the differentiated products compete in the market. Below we permit the two innovators to join their interests in an *ex ante* agreement, and they will typically find this profitable. Therefore we can assume that *ex ante* agreements always occur in equilibrium and that the two goods will be sold jointly by a monopolist for the duration of the patent. They will jointly earn $\pi_{x+y}(T)$ if both x and y are invented. The improvement y will be invented if $\pi_{x+y}(T) - \pi_x(T) - c_2 > 0$.

We assume that if the first patent is profitable for some T , investing in it is efficient; investment is efficient if $\pi_x(T) - c_1 + E_{G,H} \max \{[\pi_{x+y}(T) - \pi_x(T) - c_2], 0\} > 0$. This is because the monopolist's profit is no greater than the social surplus. Consumers collect some surplus even during the patent life, and their surplus continues beyond the patent life.

□ **Patent breadth.** The patent breadth is a value y^* with the interpretation that if the subsequent innovator discovers a product of quality $x + y$, with $y \geq y^*$, then this product is deemed not to infringe the patent. If $y < y^*$, then this product will infringe. In U.S. patent law, patent claims will be upheld by the examiners and the courts only if the applicant invented the claimed technologies, and if they satisfy "novelty" and "nonobviousness." Although these tests refer to scientific and technical considerations, and not to economic values, our model uses the latter as a proxy.²

□ **Antitrust rules.** There are two stages in our model at which an agreement between the sequential innovators could be reached. The first is when the second innovator gets his idea for the improved product but before he has sunk costs. We call this an *ex ante* license or *ex ante* agreement. A second opportunity arises after the costs of the improved product have been sunk and the resulting product infringes. We call this an *ex post* license. We do not permit agreements between firms prior to the first innovation. Although such agreements could achieve first-best incentives for research in our model, they would be difficult to negotiate; prior to invention of the first technology, it is difficult for the first innovator to identify the firms that will think of second-generation products.

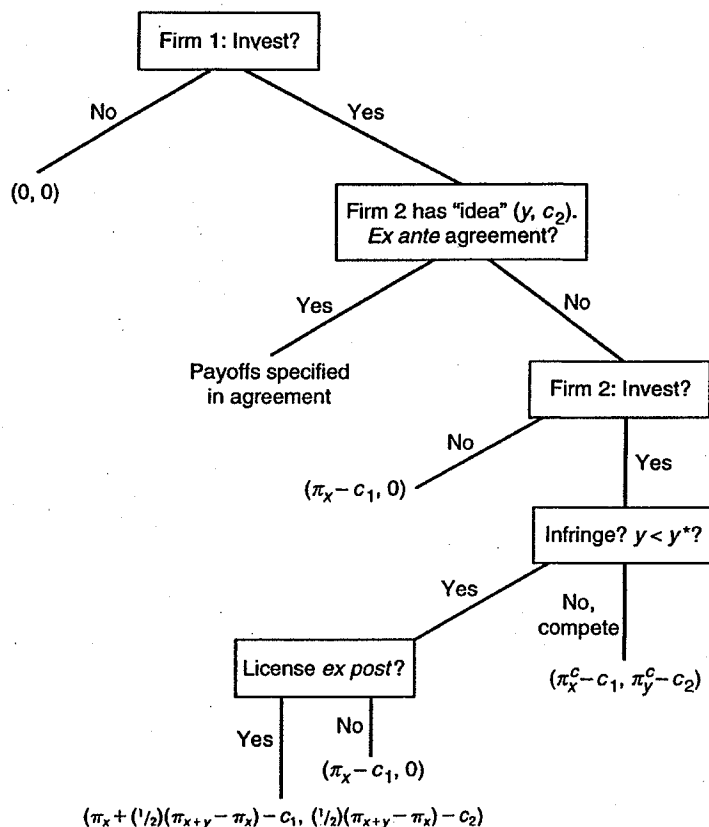
3. The division of profit

■ Figure 1 shows the order of decisions and payoffs. If firm 1 does not invest, nothing further happens and both firms get zero profit. If the first product is developed, firm 2 has its "idea" (y, c_2) for an improved product. The two firms could then make an *ex ante* agreement in which they share both the cost c_2 and the incremental revenue $\pi_{x+y} - \pi_x$ of the improved product. If they do not form this agreement, firm 2 must decide whether to invest c_2 and develop the product alone.

Figure 1 shows the defect of using only *ex post* licensing as a way to encourage research. Suppose we ignore the opportunity for an *ex ante* agreement. If the second product infringes, but the second innovator invests anyway, *ex post* he will bargain with the first patentholder over the incremental profit of the improvement, $\pi_{x+y} - \pi_x$ but not over

² See Merges and Nelson (1990, 1992), Merges (1992a, 1992b), Matutes, Regibeau, and Rockett (1990), and Chang (1995) for discussions of how to interpret these clauses. A more fully elaborated model would have the scientific "novelty" (n) of the improvement distributed jointly with its economic value (y). A product would be noninfringing if it were sufficiently novel, $n > n^*$, irrespective of y . If we reinterpret y in our model as $E(y | n)$ and if y and n were positively correlated, then a higher n^* cutoff translates into a higher $E(y | n)$ and hence a higher y^* . However, Chang (1995) argues that whether or not patent law insists on a cutoff policy, in the special case of unit demand and Bertrand competition, incentives are better under a patent policy in which either very trivial or very important innovations are deemed to infringe.

FIGURE 1 ON THE DIVISION OF PROFIT BETWEEN SEQUENTIAL INNOVATORS



the sunk research costs c_2 . We assume each firm earns half the bargaining surplus. Because firm 2 earns only $\frac{1}{2}(\pi_{x+y} - \pi_x) - c_2$, investment in second-generation products may be deterred even if such investments add to the firms' joint profits and are efficient. We could bolster the second firm's incentive to invest by ensuring that the second innovation never infringes (e.g., $y^* = 0$), but then another problem arises. The first innovator's profit is eroded through competition, and, anticipating this, the first innovator might not invest.

Ex ante agreements (before investing c_2) can help solve these problems. Neither firm would agree to less profit in the *ex ante* agreement than it would get by forgoing the agreement and doing whatever would happen instead. What would happen without an *ex ante* agreement determines the "threat points" for the *ex ante* agreement. We assume that in the *ex ante* agreement each firm will receive its threat-point profit plus one-half the bargaining surplus that the *ex ante* agreement makes available. *Ex ante* agreements can ensure that the second investment is undertaken whenever it adds to joint profit (and is therefore efficient). However, we show in the following proposition that if the patent life provides zero joint profit ($\pi_{x+y} - (c_1 + c_2) = 0$), then the first innovator will have deficient incentive to invest. Therefore the patent life must be longer when the two stages of R&D are undertaken in different firms than when concentrated in one firm.

Proposition 1. Given T , firm 1's profit is never more than $\pi_{x+y}(T) - (c_1 + c_2)$, and for some second-generation products it is less.

Proof. First suppose that firm 2 would enter even if there were no *ex ante* agreement. There are two cases: $y > y^*$ and $y < y^*$. If $y > y^*$, so that the second innovator could compete *ex post* with the first innovator, the threat points for the *ex ante* agreement are $(\pi_x^c - c_1, \pi_y^c - c_2)$. Our assumption that firm 2 would enter means that $\pi_y^c - c_2 > 0$.

The extra profit the firms earn by making an *ex ante* agreement, hence the bargaining surplus, is the profit available by avoiding *ex post* competition, namely $\pi_{x+y} - (\pi_x^c + \pi_y^c) > 0$. Splitting the surplus, their profits in the *ex ante* agreement are $(\pi_x^c - c_1 + 1/2(\pi_{x+y} - (\pi_x^c + \pi_y^c)))$, $\pi_y^c - c_2 + 1/2(\pi_{x+y} - (\pi_x^c + \pi_y^c))$. Because $\pi_x^c < \pi_{x+y} - \pi_y^c$, the profit earned by firm 1 is

$$1/2(\pi_{x+y} + \pi_x^c - \pi_y^c) - c_1 < 1/2(2\pi_{x+y} - 2\pi_y^c) - c_1 = \pi_{x+y} - c_1 - \pi_y^c < \pi_{x+y} - (c_1 + c_2).$$

That is, firm 1 earns less than the full joint profit.

Suppose now that firm 2 would enter without an *ex ante* agreement and $y < y^*$. Firm 2 could not compete with firm 1 *ex post*. An *ex post* license would enable them to bring the improvement to market and sell it at joint profit π_{x+y} . The *ex ante* agreement provides the same joint profit as entry with *ex post* licensing, and because there is no surplus to split, the division of profit will be the same as with *ex post* licensing. The bargaining surplus for the *ex post* license is $\pi_{x+y} - \pi_x$. Splitting this surplus equally, the profits of the firms in both an *ex post* license and *ex ante* agreement are

$$(\pi_x - c_1 + 1/2(\pi_{x+y} - \pi_x), 1/2(\pi_{x+y} - \pi_x) - c_2).$$

The circumstance only arises if $1/2(\pi_{x+y} - \pi_x) - c_2 \geq 0$, or $1/2\pi_x \leq 1/2\pi_{x+y} - c_2$. Thus the profit earned by firm 1 is

$$1/2\pi_{x+y} + 1/2\pi_x - c_1 \leq \pi_{x+y} - (c_1 + c_2).$$

Contrary to what we have assumed so far, firm 2 might not enter without an *ex ante* agreement; e.g., if the product would infringe ($y < y^*$), and profit would seriously be eroded by *ex post* licensing. Firm 2's profit with *ex post* licensing, $1/2(\pi_{x+y} - \pi_x) - c_2$, might be negative. If firm 2 would not enter absent an *ex ante* agreement, the bargaining surplus is the incremental profit $\pi_{x+y} - \pi_x - c_2$ provided by the improvement if it is positive, so the profits in the *ex ante* agreement are

$$(\pi_x - c_1 + 1/2(\pi_{x+y} - \pi_x - c_2), 1/2(\pi_{x+y} - \pi_x - c_2)).$$

Because the two firms' profits sum to $\pi_{x+y} - (c_1 + c_2)$, and because firm 2's profit is greater than or equal to zero (and for some products will be positive), firm 1's profit is no greater than $\pi_{x+y} - (c_1 + c_2)$.

This completes the proof, as we have exhausted all the cases. *Q.E.D.*

The three cases considered in the proof show why the second innovator has bargaining power, so that the first innovator cannot collect all the profit. First, if the second product will not infringe, the first innovator is threatened with profit-eroding *ex post* competition, as well as the inability to collect royalties on the new product. Second, if the product infringes but its incremental benefits are large relative to its costs, the second innovator may find it profitable to invest even though he must negotiate licensing fees *ex post* after his costs are sunk. He will not give away more profit in the *ex ante* agreement than is available by declining the *ex ante* agreement in favor of *ex post* licensing. And third, if the second innovator has an exclusive ability to develop the product, he might have a credible threat not to do so unless he gets a positive fraction of the incremental profit.

4. Optimal patent breadth

■ In the case of isolated inventions, it would be optimal, if possible, for the patent life to be just long enough to cover the costs of R&D.³ But Proposition 1 shows that this

³ This is true so long as R&D costs are "lump sum" and there is no issue of distorting flow rates of investment. See our discussion in Section 6.

principle does not apply when two stages of innovation are undertaken by different firms, because not all the profit can be transferred to the first innovator. If the patent life is just long enough to cover total R&D costs ($c_1 + c_2$), the first innovator will in general not enter because he could anticipate negative profit. Nevertheless, the more profit we can transfer to the first innovator, the shorter is the patent life required to stimulate research. It is therefore of interest to investigate the policies regarding patent breadth and the antitrust policy toward licensing that contribute to the goal of transferring profit from the second innovator to the first.

Turning first to the optimal breadth of the first patent, we ask whether it is best to implement the natural and straightforward policy of setting $y^* = \infty$ so that improved products and applications automatically infringe the original patent. So far we have assumed that firm 2's "idea" is a realization (y, c_2) , which means that all the uncertainty about the benefits and costs are resolved before investing. In this circumstance it turns out that $y^* = \infty$ is the best policy because it minimizes firm 2's profit in the *ex ante* agreement. To see this, first consider ideas (y, c_2) such that $\frac{1}{2}(\pi_{x+y} - \pi_x - c_2) < 0$ (so that firm 2 would not enter without an *ex ante* agreement if the product infringed), and $\pi_y^c - c_2 \geq 0$ (firm 2 would enter without an *ex ante* agreement if the second product did not infringe). Firm 2 gets $\frac{1}{2}(\pi_{x+y} - \pi_x - c_2)$ in the *ex ante* agreement if the product infringes, and $\frac{1}{2}(\pi_{x+y} - (\pi_x^c + \pi_y^c)) + (\pi_y^c - c_2)$ if not. Firm 2 is worse off if the product infringes, because

$$\begin{aligned} \frac{1}{2}(\pi_{x+y} - \pi_x - c_2) &< \frac{1}{2}(\pi_{x+y} - \pi_x^c - c_2) + \frac{1}{2}(\pi_y^c - c_2) \\ &= \frac{1}{2}(\pi_{x+y} - (\pi_x^c + \pi_y^c)) + (\pi_y^c - c_2). \end{aligned}$$

Now consider ideas (y, c_2) such that firm 2 would not enter without an *ex ante* agreement, even if its product did not infringe. Then firm 2 gets the same profit in the *ex ante* agreement whether or not the product infringes, because the firms split the same surplus *ex ante* in both cases. Finally, consider ideas (y, c_2) such that firm 2 would enter whether or not the second product infringed. The profit the firms would split in the case of infringement would be $\pi_{x+y} - \pi_x$, and in the case of noninfringement it would be $\pi_{x+y} - (\pi_x^c + \pi_y^c)$. Therefore, in the *ex ante* agreement, firm 2 gets $\frac{1}{2}(\pi_{x+y} - \pi_x) - c_2$ in the case of infringement, and $\frac{1}{2}(\pi_{x+y} - (\pi_x^c + \pi_y^c)) + (\pi_y^c - c_2)$ in the case of noninfringement. Because

$$\frac{1}{2}(\pi_{x+y} - \pi_x) - c_2 < \frac{1}{2}(\pi_{x+y} - \pi_x^c + \pi_y^c) - c_2 = \frac{1}{2}(\pi_{x+y} - (\pi_x^c + \pi_y^c)) + (\pi_y^c - c_2),$$

firm 2 gets more profit if its product does not infringe. Thus, for all ideas (y, c_2) , firm 1 is better off in the *ex ante* agreement if firm 2's product infringes. Therefore:

Proposition 2. If all the uncertainty on y and c_2 is resolved prior to investment in the second product, the best patent breadth is $y^* = \infty$.

It is intuitive that the best way to protect firm 1's profit is to give firm 1 a broad patent. But curiously, this is not necessarily true when only the uncertainty on c_2 has been resolved. To take the extreme case, suppose that an "idea" is (G, c_2) ; firm 2 knows how much it must invest, but it does not know at the time of investment how valuable the outcome will be. In this case firm 2 cannot make a different entry decision for every y , conditional on c_2 . However, the optimal breadth may now be finite, as shown in Proposition 3. The condition that $E_G \frac{1}{2}(\pi_{x+y} - \pi_x) - c_2 \leq 0$ for all c_2 in the support of H means that if the firm must license *ex post* for sure, it would never enter without an *ex ante* agreement. We have already interpreted the assumption that $\pi_{x+y} = \pi_x + \pi_y^c = \pi_x^c + \pi_y^c$; it could be either the special case where the second-generation

product is an application, or the special case where the second-generation product is a spin-off product that serves an unrelated market.

The intuition behind Proposition 3 is that a very broad patent can give firm 2 a credible threat not to enter, because it knows that with high probability its profit will be eroded through *ex post* licensing. This credible threat enables firm 2 to bargain *ex ante* for half the incremental value of the new product. If the first patent is narrower, there is some possibility that firm 2's product will turn out to be noninfringing, which raises the expected profit of entering without an *ex ante* license. A narrower patent can overcome firm 2's credible threat not to enter, while keeping firm 2's expected profit less than half the incremental value of the new product.

Proposition 3. Suppose firm 2's "idea" is (G, c_2) . Suppose further that $E_G \frac{1}{2}(\pi_{x+y} - \pi_x) - c_2 \leq 0$ for all c_2 in the support of H , and that $\pi_{x+y} = \pi_x + \pi_y^c = \pi_x^c + \pi_y^c$. Then the optimal patent breadth is finite, $y^* < \infty$.

Proof. If firm 2 enters in the absence of an *ex ante* agreement, its expected profit is

$$\int_{y^*}^{\infty} \pi_y^c dG + \int_0^{y^*} \frac{1}{2}(\pi_{x+y} - \pi_x) dG - c_2 = \int_{y^*}^{\infty} \pi_y^c dG + \int_0^{y^*} \frac{1}{2}\pi_y^c dG - c_2.$$

We can assume that $E_G \pi_y^c - c_{2L} > 0$, where c_{2L} is the smallest c_2 in the support of H ; otherwise no policy will induce entry of firm 2, and there is no problem to solve. Because $E_G \frac{1}{2}\pi_y^c - c_{2L} < 0 < E_G \pi_y^c - c_{2L}$, there is $y^{**} < \infty$ that satisfies

$$\int_{y^{**}}^{\infty} \pi_y^c dG + \int_0^{y^{**}} \frac{1}{2}\pi_y^c dG - c_{2L} = 0.$$

For all projects with $c_2 > c_{2L}$, the patent breadth y^{**} provides the same profit to firm 1 as $y^* = \infty$, namely $\max \{ \frac{1}{2}(E_G \pi_y^c - c_2), 0 \}$. (Without an *ex ante* agreement, firm 2 would not enter, and therefore, in the *ex ante* agreement, the two firms split $E_G \pi_y^c - c_2$.) For the project with cost c_{2L} , firm 1 earns more profit with y^{**} than with $y^* = \infty$. With the breadth $y^* = \infty$, the bargaining surplus is $E_G \pi_y^c - c_{2L}$. With the breadth y^{**} , there is no longer a bargaining surplus to be split, because firm 2 would invest even without an *ex ante* agreement and earn zero profit, which is less than $\frac{1}{2}(E_G \pi_y^c - c_{2L})$. Thus the policy y^{**} transfers more profit to firm 1 than the policy $y^* = \infty$. *Q.E.D.*

5. Antitrust policy

■ Our main objective has been to study patent policy, and to show the important role of patent breadth in setting threat points for negotiating licensing agreements between sequential innovators. By setting the threat points, the patent breadth determines the division of profit in each period and affects the minimum patent length required to stimulate R&D.

A subsidiary issue, which cannot be studied in depth without more structure than we have imposed on our model, is what types of licensing agreements the antitrust authorities should allow. So far we have taken the antitrust policy as fixed. We have assumed that *ex post* licensing is possible only if the second product infringes. This excludes *ex post* collusive licensing between innovators who would otherwise compete in the market. We have also assumed that *ex ante* agreements are always legal. There is a slight inconsistency here in that the firms are allowed to collude by forming agreements *ex ante* even when such collusion would be prohibited *ex post*. Our first task in this section is to explain these assumptions, and our second task is to shed some light on the optimal design of antitrust policy.

The antitrust authorities would presumably allow *ex post* collusion only if it was necessary *ex ante* to encourage R&D; otherwise they would prohibit it in order to protect

consumers. To justify collusion *ex post*, the antitrust authorities would have to reason that without assurance of *ex post* permission to collude, the firms would not have invested. The problem of verifying this circumstance *ex post* seems daunting, especially because firms have incentive to misrepresent. Once R&D is done, the antitrust authorities might easily want to protect consumers rather than compensate firms for their sunk costs of R&D.

Contrast this with the problem of giving *ex ante* permission to collude: the firms might argue that they will not invest without an *ex ante* agreement because *ex post* competition between them will erode their profit. Such an argument might be more credible if it is made before costs are sunk rather than after. In addition, it is unknown *ex ante* whether the second product will infringe; if it does, the firms will end up merging their interests *ex post* anyway, and there is no consumer loss due to the *ex ante* agreement. These considerations might make the antitrust authorities more lenient. In addition, the firms' opportunity to merge their interests *ex ante* will reinforce the antitrust authorities in their skepticism about collusive licensing *ex post*. If petitioned by opportunistic firms *ex post*, the antitrust authorities could reasonably reply that if the firms had truly been worried about profit-eroding competition, they should have petitioned to merge their interests *ex ante*.

Although we believe that our assumptions about contracting opportunities are reasonable and might reflect current policy, the arguments we have just made are not unassailable, and it is therefore of interest to consider the optimal design of antitrust policy. What types of licensing *should* be legal? *Ex ante* and not *ex post*? *Ex post* and not *ex ante*? *Ex ante* only when the second product would infringe? It is obvious that at least one type of licensing should be permitted. Otherwise no firm would ever invest in an infringing second-generation product (because he could not bring it to market), and this would impede progress if firms have specialized abilities.

We first consider *ex post* licensing, assuming that *ex ante* licensing is legal. It is hard to imagine that the authorities would prohibit such licensing, because they would reason that an infringing product could not come to market without a license. However, this argument is flawed if firms can license *ex ante* instead of *ex post*. When *ex ante* licensing is available, the only role of *ex post* licensing is to shift the threat points for the *ex ante* agreement. The possibility of *ex post* licensing does not affect the total profit that is available to the firms, but only the division of it.⁴ It turns out that the possibility of *ex post* licensing favors the first innovator because it may take away firm 2's credible threat not to enter, which reduces his bargaining power in an *ex ante* agreement. (The argument is similar to the argument above in Proposition 3.) To show this, we return to the assumption that firm 2's "idea" is a realization (y, c_2) , although a similar argument applies when an "idea" is (G, c_2) .

Proposition 4. Assume *ex ante* licensing is legal. For any patent breadth, the first innovator earns greater profit if *ex post* licensing is permitted than if not.

Proof. We need to show that for a given T and y^* , the possibility of *ex post* licensing increases firm 1's profit in the *ex ante* agreement and decreases firm 2's profit. We divide our analysis into three cases. The first case is when the second product does not infringe ($y > y^*$). Then *ex post* licensing is irrelevant, and the firms' profits in an *ex ante* agreement are the same whether or not *ex post* licensing is legal. The second case is when the second product infringes ($y < y^*$), but firm 2 would not enter without an *ex ante* agreement even if licensing were legal (and *a fortiori* would not enter if licensing were not legal). Then the firms' profits with an *ex ante* agreement are respectively the same whether or

⁴ Horbulyk (1991) shows how compulsory licensing, which affects the terms of the *ex post* licensing agreement, can have a further salutary effect on the threat points. In his model the compulsory fee is fixed in advance by the authorities.

not licensing is legal. The only case where *ex post* licensing matters for dividing the profit *ex ante* is when the second product infringes ($y < y^*$) and firm 2 would enter if *ex post* licensing were legal, but not otherwise: i.e., when $\frac{1}{2}(\pi_{x+y} - \pi_x) - c_2 > 0$. When *ex post* licensing is prohibited, the firms' profits with no agreement would be $(\pi_x - c_1, 0)$, and with an *ex ante* agreement they would be

$$(\pi_x - c_1 + \frac{1}{2}(\pi_{x+y} - \pi_x - c_2), \frac{1}{2}(\pi_{x+y} - \pi_x - c_2)).$$

When licensing is legal, the firms' profits with *ex post* licensing would be $(\pi_x - c_1 + \frac{1}{2}(\pi_{x+y} - \pi_x), \frac{1}{2}(\pi_{x+y} - \pi_x) - c_2)$. Hence, firm 1's profit is greater when licensing is legal than when it is not. *Q.E.D.*

A more troublesome question is whether *ex ante* agreements should be permitted, given that *ex post* licensing is permitted. The consequence of permitting *ex ante* agreements is that for the duration of the patents the products are sold by a single monopolist who earns profit $\pi_{x+y}(T)$. If *ex ante* licensing were prohibited, and if the second product infringed, then the two products would compete in the market and earn $\pi_x^c(T)$ and $\pi_y^c(T)$ instead of the larger $\pi_{x+y}(T)$. Although the per-period profit of firms is then smaller, the per-period consumers' surplus is larger. To compensate firms for smaller per-period profit, the patents must be longer, but lengthening the patent hurts consumers.

With no more structure on the model than we have so far imposed, it is impossible in general to say whether it is better to force competition for a longer period or permit collusion for a shorter period. Nevertheless, we can identify at least one important circumstance in which *ex ante* agreements should clearly be allowed. The next proposition says that when the second-generation product is an application of a basic technology (rather than an improvement of it), allowing *ex ante* agreements permits a shorter patent life while not affecting per-period monopoly distortions. Therefore, *ex ante* agreements improve social welfare.

Proposition 5. Suppose that second-generation products are applications of the first technology, which has no value as a stand-alone product, that is, $\pi_{x+y} = \pi_y^c$. Then whatever the patent breadth, *ex ante* licensing improves social welfare.

Proof. The per-period monopoly distortions are the same whether the license is *ex ante* or *ex post*, or whether the second product is noninfringing and marketed with no license. In all these regimes, the second product is marketed by a monopolist, and the first product does not compete with it. *Ex ante* agreements increase social welfare because they can ensure that the second product is developed (which improves utility even though there are monopoly distortions), and they can increase the profit of the first innovator. If the patent life is such that firm 2 would enter without an *ex ante* agreement (whether or not the second product infringed), then the *ex ante* agreement will not change either the monopoly distortions or the firms' profits, and is irrelevant. But if the second innovator would not enter (either because the patent life is so short that $\frac{1}{2}(\pi_y^c(T) - c_2) < 0$ or because *ex post* licensing is illegal), then if $\frac{1}{2}(\pi_y^c(T) - c_2) > 0$, *ex ante* agreements will ensure that the investment is made, and will also increase the first innovator's profit. Thus, permitting *ex ante* agreements never decreases social welfare, and for some T increases social welfare. *Q.E.D.*

The unresolved question is for the case that the two products compete, e.g., the second is an improved version of the first, or they are horizontally differentiated. To address this case we would need more structure in our model. By assuming the same structure of preferences and the same type of competition as Klemperer (1990) assumes, we might be able to infer whether *ex ante* agreements should be allowed. However, we do not wish to adopt additional structure, as our other conclusions do not require it.

An intermediate policy between always allowing or always disallowing *ex ante* licensing is to allow a complete *ex ante* merger of interests when the second product would infringe the first patent, but not otherwise. If the second product would likely not infringe, assume that the firms can agree *ex ante* to share costs c_2 , but must compete *ex post* in the market.⁵ Of course such a rule would be hard to interpret unless all the uncertainty about the second product ("y") were resolved before investing in it, because it would not be known whether the product would infringe. For purposes of discussion we will assume that the uncertainty is largely resolved. If it were known that the second product would infringe, then the division of profit in the *ex ante* license would be as discussed above. If it were known that the second product would not infringe, there would be no *ex ante* agreement to share costs. If $\pi_y^c(T) - c_2 > 0$, the second firm would enter without an agreement, and then there is no reason for the agreement, because it cannot increase both firms' profits. If $\pi_y^c(T) - c_2 < 0$ so that the second firm would not enter without some cost sharing, then, because $\pi_x^c(T) - \pi_x(T) < 0$, it follows that $(\pi_x^c(T) + \pi_y^c(T)) - \pi_x(T) - c_2 < 0$. The latter is the increment to profit that the firms would jointly earn investing in the second product, and they will therefore not invest. If *ex post* competition were not required, then the incremental joint profit with an *ex ante* licensing agreement would be $\pi_{x+y}(T) - \pi_x(T) - c_2$, which might be positive even when $\pi_y^c(T) - c_2 < 0$.

Thus, forcing firms to compete except in cases of infringement undermines the joint profit the two firms can collect. In order to ensure a fixed profit for the first innovator, the patent must last longer when *ex post* competition is required in cases of noninfringement than when the firms can always merge their interests through *ex ante* licensing. The welfare considerations are similar to those discussed above: it might be better to prevent collusion and let the patent last longer rather than to permit collusion for a shorter period. Our model does not have enough structure to address this issue.

We conclude this section with a remark on the interaction between patent policy and antitrust policy. In general, optimal patent policy should depend on what contracts the firms are permitted to make or are likely to make. A patent policy that assumes firms will license *ex ante* could be counterproductive if there were obstacles to such agreements, such as asymmetries of information, that would make contracts difficult to negotiate. For example, for the case that all uncertainty on y is resolved prior to the second investment, we concluded that if *ex ante* licensing is available, the first patent should be so broad that all second products infringe it. This is not the optimal policy if *ex ante* agreements are prohibited. Suppose that $\frac{1}{2}(\pi_{x+y}(T) - \pi_x(T)) - c_2 < 0$ for all products. Then prohibiting *ex ante* agreements would ensure there would be no investment in second-generation products. A better patent policy would be to make some second-generation products noninfringing.⁶

6. Conclusion

■ We have studied the interaction between patent breadth and antitrust policy in a simple model designed to focus attention on the division of profit between sequential innovators, and to see how the division of profit depends on patent breadth and on the opportunities for cost-sharing through licensing. Despite the simplicity of the model, we believe that some of our propositions are very robust.

⁵ Such a rule would be in the spirit of the literature following d'Aspremont and Jacquemin (1990), in which firms are allowed to cooperate in R&D but required to compete *ex post* in the market. In that literature, R&D reduces unit costs of production, and there are positive spillovers between firms in reducing costs. Firms will cooperate to exploit those spillovers even if they must compete *ex post*.

⁶ If the high-value second products would be invented but not the low-value products, then one might want to make only the low-value products noninfringing. Chang (1995) discusses such a policy, although it might be hard to reconcile with patent law. We have assumed that patent breadth must be a "cutoff" policy y^* such that only products with $y > y^*$ are noninfringing.

Ex ante agreements have the potential to increase the expected profits of both parties without inhibiting later research, but even with *ex ante* agreements, the first innovator cannot collect all the profit from second-generation improvements. This conclusion, which derives primarily from the fact that the second innovator has bargaining power, seems quite general, and it has a clear consequence for patent length: in order to give sufficient incentive for basic research, patents must last longer when cumulative research is undertaken by different firms than when both generations of research are concentrated in the same firm.⁷ In the latter case, there is no problem of dividing profit.

In the model presented here, the sequential innovators will always form *ex ante* agreements, and the roles of *ex post* licensing and patent breadth are to set "threat points" for such agreements. We concluded that the possibility of *ex post* licensing has a salutary effect on the first innovator's profit in *ex ante* agreements. We would have been quite disturbed to reach the opposite conclusion, because *ex ante* agreements might not occur for a variety of reasons not discussed here, and then *ex post* licensing is necessary to bring improved products to market. We were less conclusive about the circumstances in which *ex ante* agreements should be legal, except in the (not uncommon) case where second-generation products are applications of a basic technology. In that case, *ex ante* agreement unambiguously serves the public interest. As to the breadth of patent protection, we have shown circumstances in which the first patent should be broad, but other circumstances in which a broad patent reduces the first innovator's profit because it increases the later innovator's bargaining power. It does this by giving the later innovator a credible threat not to undertake a profitable investment unless the first innovator is willing to share the costs. Counterintuitively, the first innovator can be better off with a narrower patent when the patent's life is fixed.

Except for the qualifications in Section 5, the premise behind all these conclusions is that it is desirable to minimize the patent life in order to minimize the duration of monopoly distortions. We should notice, however, that minimizing the patent life could have a distortionary effect not yet discussed: it could reduce the rates of R&D investment, as in the extensive patent race literature based on the Poisson process. Our model avoids this issue because R&D investments are lump sum. However, if rates of investment were endogenous, our model would face the same tradeoff as any other patent race model: by reducing the patent value (patent life), one reduces the rate at which firms find it optimal to invest, and this could be inefficient. (This issue is discussed in Loury, 1979; Nordhaus, 1969; Dasgupta and Stiglitz, 1980; and Gandall and Scotchmer, 1993.) In this richer model, the optimal patent life must consider the effect both on rates of investment and on monopoly distortions.

For completeness, it is worth noting that our main result on patent life would follow if we were concerned with rates of R&D investment as well as with minimizing monopoly distortions. To take the extreme case, let T^* be the patent life that elicits efficient rates of investment in the first stage of innovation with a fixed number of firms, possibly one, when each first-stage firm is capable also of developing the second product. Compare this with what happens if the two stages take place in different firms. By our argument in Section 2, the second-stage firms will collect some of the joint profit, and therefore each firm at the first stage races for a smaller reward. The firms will therefore reduce their rates of investment. To restore the efficient rates of investment, the patent life (patent value) must be increased.

⁷ If the sequential patents could have different lengths, and if the lengths could depend on the costs of R&D, then all the profit could be transferred to the first innovator. However, patent law has no provision for patent length to reflect R&D costs, and in any case, for such a scheme to work, the length of the patent on the initial innovation would have to be different for each second-generation product. See Scotchmer (1991a, 1991b, forthcoming) for further discussion.

One might worry that the conclusions of this article depend critically on the fact that a unique firm can develop the improvement or application. Although it is easy to imagine that a single firm has the idea or capability for the improvement or application, it is also easy to imagine that the idea leaks out, and that several firms might compete for an *ex ante* license for the same improvement or application. Such competition should aid the first patentholder in capturing profit. However, Scotchmer (1991b) argues that even in that case the first innovator cannot capture all the profit. The firm that receives the *ex ante* license might nevertheless find itself in a race. If another firm achieves the improvement first, the license becomes less valuable because two firms now hold blocking patents on the improved product. We are therefore back to bilateral bargaining, and this possibility will reduce each firm's willingness to pay for an *ex ante* license.

Finally, we reiterate that patent length and breadth play a different role when innovation is cumulative and *ex ante* agreements are available than when they are not. In the models of Gilbert and Shapiro (1990) and Klemperer (1990) there is only one innovator, and the authors study the question of whether it is better to provide a fixed amount of profit by granting a long, narrow patent or a short, broad one. Broadening the patent increases the per-period profit earned by the innovator. In the Gilbert and Shapiro model this is because a broader patent permits the innovating monopolist to charge a greater markup. In the Klemperer model it is because a broader patent prevents competitors from selling close substitutes. In our model the per-period joint profit of the firms is fixed because the sequence of innovators will act like a joint monopolist. The role of patent breadth is not to determine the *level* of per-period profit, but rather its *division* between sequential innovators. It does this by establishing the bargaining positions of sequential innovators who might end up holding blocking patents if they did not agree *ex ante* to license. If one looks only at the first innovator, then it follows from Proposition 3 that for a fixed patent life, narrowing his patent could increase his profit, contrary to what happens in the models of Klemperer and Gilbert and Shapiro.

References

- CHANG, H.F. "Patent Scope, Antitrust Policy, and Cumulative Innovation." *RAND Journal of Economics*, Vol. 26 (1995), pp. 34-57.
- DASGUPTA, P. AND STIGLITZ, J. "Uncertainty, Industrial Structure, and the Speed of R&D." *Bell Journal of Economics*, Vol. 11 (1980) pp. 1-28.
- d'ASPREMONT, C. AND JACQUEMIN, A. "Cooperative and Noncooperative R&D in Duopoly with Spillover." *American Economic Review*, Vol. 80 (1990), pp. 641-642.
- GALLINI, N.T. "Deterrence by Market Sharing: A Strategic Incentive for Licensing." *American Economic Review*, Vol. 74 (1984), pp. 931-941.
- GANDAL, N. AND SCOTCHMER, S. "Coordinating Research through Research Joint Ventures." *Journal of Public Economics*, Vol. 51 (1993), pp. 173-193.
- GILBERT, R. AND SHAPIRO, C. "Optimal Patent Length and Breadth." *RAND Journal of Economics*, Vol. 21 (1990), pp. 106-112.
- GREEN, J. AND SCOTCHMER, S. "Antitrust Policy, the Breadth of Patent Protection and the Incentive to Develop New Products." H.I.E.R. Working Paper, Harvard University, December 1989.
- HORBULYK, T.M. "Compulsory Licensing of Innovation and the Incentives for Cumulative Research." Mimeo, University of Calgary, 1991.
- KLEMPERER, P. "How Broad Should the Scope of Patent Protection Be?" *RAND Journal of Economics*, Vol. 21 (1990), pp. 113-130.
- LOURY, G.C. "Market Structure and Innovation." *Quarterly Journal of Economics*, Vol. 93 (1979), pp. 395-410.
- MATUTES, C., REGIBEAU, P., AND ROCKETT, K.E. "Optimal Patent Protection and the Diffusion of Innovation." Mimeo, Department of Economics, Northwestern University, December 1990.
- MERGES, R.P. "Uncertainty and the Standard of Patentability." *High Technology Law Journal*, Vol. 7 (1992a), pp. 1-70.
- . "On Limiting or Encouraging Rivalry in Technical Progress: The Effect of Patent Scope Decisions." Center for Law and Economic Studies Working Paper no. 72, Columbia University, 1992b.

MERGES, R.P. AND NELSON, R.R. "On the Complex Economics of Patent Scope." *Columbia Law Review*, Vol. 90 (1990), pp. 839-916.

——— AND ———. "On Limiting or Encouraging Rivalry in Technical Progress: The Effect of Patent Scope Decisions." Center for Research in Management, Consortium on Competitiveness and Cooperation, Working Paper no. 92-8, University of California, Berkeley, 1992.

NORDHAUS, W.D. *Invention, Growth and Welfare*. Cambridge, Mass.: MIT Press, 1969.

SCOTCHMER, S. "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law." *Journal of Economic Perspectives*, Vol. 5 (1991a), pp. 29-41.

———. "Protecting Early Innovators: Should Second Generation Products be Patentable?" Graduate School of Public Policy Working Paper no. 183, University of California, Berkeley, 1991b.

———. "Patents as an Incentive System." *Proceedings of the 1992 World Congress of the International Economics Association*, Paris, forthcoming.

SPENCE, A.M. "Cost Reduction, Competition, and Industry Performance." *Econometrica*, Vol. 52 (1984), pp. 101-121.